Sensitivity analysis for unmeasured confounding in principal stratification settings with binary variables

Scott Schwartz[1*] Fan Li[2], and Jerome P. Reiter[2]

[1]Department of Statistics, Texas A&M University, College Station, TX 77843-3143
[*]scott@stat.tamu.edu
[2]Department of Statistical Science, Duke University, Durham, NC 27708-0251

October 17, 2011

Abstract

Within causal inference, principal stratification (PS) is a popular approach for dealing with intermediate variables, i.e., variables affected by treatment that also potentially affect the response. However, when there exists unmeasured confounding in the treatment arms—as can happen in observational studies—causal estimands resulting from PS analyses can be biased. We identify the various pathways of confounding present in PS contexts and their effects for PS inference. We present model-based approaches for assessing the sensitivity of complier average causal effect estimates to unmeasured confounding in the setting of binary treatments, binary intermediate variables, and binary outcomes. These same approaches can be used to assess sensitivity to unknown direct effects of treatments on outcomes since, as we show, direct effects are operationally equivalent to one of the pathways of unmeasured confounding. We illustrate the methodology using a randomized study with artificially introduced confounding and a sensitivity analysis for an observational study of the effects of physical activity and body mass index on cardiovascular disease.

KEYWORDS: Causal; Intermediate; Principal stratification; Confounding; Observational; Sensitivity Analysis

1 Introduction

Intermediate variables are post-treatment variables potentially affected by treatment and affecting response. Regardless of whether the study design is randomized or observational, intermediate variables are frequently present, e.g., in settings involving non-compliance, missing data, and surrogate endpoints. Under such circumstances, standard intention-to-treat analyses may not be sufficient to estimate treatment efficacy, so that intermediate variables must be dealt with for causal inference. However, it is well documented that applying standard methods of pre-treatment variable adjustment to intermediate variables, such as regression or per-protocol analysis, can result in post-treatment selection bias, see e.g., [1]. To illustrate, let $Y_i(Z_i)$ and $D_i(Z_i)$ be respectively the potential outcomes [2] of the response of interest and the intermediate variable for unit i under an assigned binary treatment, $Z_i = 0, 1$. In general, the comparison between $\{Y_i(0)\colon D_i(0) = d\}$ and $\{Y_i(1)\colon D_i(1) = d\}$ for all $i = 1, \ldots, n$ units in the study is not a causal effect when Z_i affects D_i, because $\{i\colon D_i(0) = d\} \neq \{i\colon D_i(1) = d\}$.

A principled approach to handling intermediate variables in causal inference is principal stratification (PS), in which one compares $\{Y_i(1) : S_i = s\}$ and $\{Y_i(0) : S_i = s\}$ [3]. Here $S_i = (D_i(1), D_i(0))$ is called a principal stratum. The key insight is that S_i is invariant under treatment assignment, so that the principal strata may be used as pre-treatment variables. That is, comparisons within $S_i = s$, known as principal effects (PE), are well-defined causal effects. Specifically, PEs in strata $\{S_i\colon D_i(0) = D_i(1)\}$ can be interpreted as direct effects of treatment on response, while PEs in strata $\{S_i\colon D_i(0) \neq D_i(1)\}$ can be interpreted as the effects of treatment mediated through the intermediate variable response plus any direct effects of treatment on response [4].

Since S are not fully observed, the identifiability of PEs usually relies on a set of structural assumptions, e.g., no unmeasured confounding and an exclusion restriction (ER) (see Section 2). These assumptions may not be true, particularly in observational studies. For example, consider the Swedish March National Cohort (NMC) analyzed by Sjölander et al. [5] and by us in Section 5. Here, Z is one's physical activity (PA) level, D is one's body mass index (BMI), and Y is the event of cardiovascular disease (CVD). Interest lies in the roles of PA and BMI in influencing CVD risk. It is known that BMI is directly affected by PA and obesity is highly correlated with CVD risk, making BMI a possible intermediate variable on the causal pathway between PA and CVD. However, the treatment variable PA is self-selected by participants, and this selection could be confounded with other healthy lifestyle practices that are not observed in the data. Furthermore, PA could improve CVD outcomes directly, which would violate the ER.

When identifying assumptions are suspect, it is prudent to examine the sensitivity of results to violations of them [6]. For example, analysts can remove identifying assumptions to derive bounds for PS estimands [7, 8]. Alternatively, analysts can encode identifying assumptions as sensitivity parameters that are included in models for causal effects [5, 9]. To do so, expert opinion can be used to specify plausible values of the sensitivity parameters, and examine how PS estimates change over those plausible values. We note that limits on the values of plausible sensitivity parameters imply bounds for PS estimands as well.

Following [5] and [9], we present a sensitivity analysis framework that encodes unmeasured confounding as continuous sensitivity parameters. We identify two types of (simultaneous) unmeasured confounding in PS: (1) S-confounding, which affects the estimation of principal strata, and (2) Y-confounding, which affects the estimation of effects on the response within principal strata. For the setting of binary treatment, binary intermediate variable, and binary outcome, we develop a nonparametric expression for the bias in standard PS estimators without covariate adjustment. For the same setting, we also present an approach to sensitivity analysis using model-based inference with covariate adjustment. Both strategies can be used to assess sensitivity to unknown direct effects since, as we show, direct effects are operationally indistinguishable from Y-confounding in this PS context.

Although the general approach applies to any PS estimand, we focus on the complier average causal effect (CACE) [10], also known as the local average treatment effect (LATE) [11]. CACE is the PE in the principal stratum $\{S_i\colon D_i(0) \neq D_i(1)\}$. In randomized trials with noncompliance, CACE represents the efficacy of the treatment [see, e.g., 12]. In mediation studies, CACE represents the average causal effect for the units in the latent subpopulation whose intermediate variable would change due to the treatment; this includes the effects that the treatment has on the response both mediated and not mediated via the intermediate variable of interest. The CACE has been studied in observational settings by [13] and [14], and alternative PEs have been considered in the observational settings by [5] and [15]. All of these analyses were based on the assumption of no unmeasured confounding. Sensitivity to unmeasured confounding has been examined in instrumental variables settings, which are closely related to PS CACE analysis, by [16]. They use a permutation distribution to determine the level of unmeasured confounding that discounts a significant treatment effect. The key differences between the methods of [16] and our framework include (1) we explicitly identify and parameterize confounding pathways via a model-based approach, and (2) we examine settings predicated on the possibility of direct effects.

The remainder of the article is organized as follows. Section 2 reviews the standard PS assumptions, clarifies the role of the assumption of no unmeasured confounding, and demonstrates the effects of confounding on the CACE estimate using the nonparametric method of moments. Section 3 presents a general parametric approach to sensitivity analysis for CACE estimation that addresses unmeasured confounding (and direct effects). Inferences from both frequentist and Bayesian paradigms are provided. Section 4 illustrates the ability of the parametric approach to recover the CACE in the presence of confounding in a constructed observational study. Section 5 demonstrates how one can apply sensitivity analyses using the NMC study. Finally, Section 6 concludes with a discussion.

2 Confounding in Principal Stratification

When Z_i and D_i are binary, the principal strata are $S_i \in \{(0,0),(1,0),(1,1),(0,1)\}$. In non-compliance contexts, the S_i are often called, in the order shown, never-takers ($S_i = n$), compliers ($S_i = c$), always-takers ($S_i = a$), and defiers ($S_i = d$), as in [17]. Principal strata can be defined in settings other than non-compliance; for instance, PA as a treatment, obesity

(BMI$>$ 30) as a binary intermediate variable, and the event of CVD as a response. We use the familiar nomenclature of non-compliance to generically refer to S_i.

In order to identify the principal effects, the following assumptions are often made.

A1. *Stable unit treatment value assumption* (SUTVA) [18]. There are no different versions of any single treatment arm and no interference between units.

A2. *Monotonicity.* $D_i(1) \geq D_i(0)$ for all i, ruling out the principal stratum of defiers.

A3. *Exclusion restriction* (ER). If $D_i(1) = D_i(0)$, then $Y_i(1) = Y_i(0)$ for all i, implying that compliers, always-takers, and never-takers experience no direct effect of treatment on response.

A4. *No unmeasured confounding.* $(Y_i(0), Y_i(1), S_i) \perp\!\!\!\perp Z_i | X_i$ for all i and observed covariates X. This is referred to as strong ignorability of assignment [19].

In randomized experiments, analysts can assume A4 by design, but A3 may not hold due to direct effects of treatment on response; in observational studies, neither A3 nor A4 are guaranteed. We assume A1 and A2 for the remainder of the article. However, we depart from the classical PS set-up by not assuming A3 and A4. As we show in Section 2.2, the effects of violations of the ER are not distinguishable from the effects of Y-confounding, so that we examine the consequences for inference when both A3 and A4 are incorrect but applied regardless. We begin by characterizing unmeasured confounding.

2.1 Characterizing Unmeasured Confounding

The no unmeasured confounding assumption A4 can be expressed as

$$\Pr(Y_i(0), Y_i(1), S_i | Z_i = 1, X_i) \quad = \quad \Pr(Y_i(0), Y_i(1), S_i | Z_i = 0, X_i), \tag{1}$$

for all i. We do not condition on D in (1) since it is completely determined given S and Z. Under (1), the PS setting may be represented graphically by Figure 1a. Unmeasured confounding arises, and (1) fails, when some possibly multidimensional variable U that effects $(Y_i(0), Y_i(1))$ and S—after adjustment for X—also affects Z, as represented in Figure 1b.

When such a U exists, we can rewrite (1) as

$$\Pr(Y_i(0), Y_i(1), S_i | Z_i, X_i, U_i) = \Pr(Y_i(0), Y_i(1) | Z_i, S_i, X_i, U_i^{Y|S}) \Pr(S_i | Z_i, X_i, U_i^S).$$

Here, we partition the unmeasured confounders into $U^{Y|S}$ and U^S, which are the possibly overlapping subsets of U that affect each component of the likelihood. This factorization suggests that unmeasured confounding can arise via two pathways.

1. *S-confounding*: the distribution of S varies with Z because of U^S (see Figure 1c), i.e.,

$$\Pr(S_i | Z_i = 1, X_i) \quad \neq \quad \Pr(S_i | Z_i = 0, X_i);$$

2. *Y-confounding*: within S, the distribution of $(Y(0), Y(1))$ varies with Z because of $U^{Y|S}$ (see Figure 1d), i.e.,

$$\Pr(Y_i(0), Y_i(1)|Z_i = 1, S_i = s, X_i) \neq \Pr(Y_i(0), Y_i(1)|Z_i = 0, S_i = s, X_i).$$

When S-confounding or Y-confounding exists, (1) no longer holds, and inferences predicated on this assumption can be biased.

[Figure 1 about here.]

2.2 Implications of Unmeasured Confounding and a False Exclusion Restriction

The role of A3 and A4 can be illustrated in the simple setting of no covariates, binary outcomes, and an additive treatment effect. Here, the CACE is

$$\theta_c^* = \Pr(Y_i(1)|S_i = c) - \Pr(Y_i(0)|S_i = c).$$

Under A1–A4, the CACE is identifiable, and equal to

$$\begin{aligned} \hat{\theta}_c^{obs} &= \Pr(Y_i(1)|S_i = c, Z_i = 1) - \Pr(Y_i(0)|S_i = c, Z_i = 0) \\ &= \frac{(p_{11}\pi_{11} - p_{10}\pi_{10}) + (p_{01}\pi_{01} - p_{00}\pi_{00})}{1 - \pi_{01} - \pi_{10}}. \end{aligned} \quad (2)$$

Here, $p_{dz} = \Pr(Y_i^{obs} = 1|D_i = d, Z_i = z)$ and $\pi_{dz} = \Pr(D_i = d|Z_i = z)$ for $d = 0, 1$ and $z = 0, 1$, where $Y_i^{obs} = Z_i Y_i(1) + (1 - Z_i)Y_i(0)$ [17]. All quantities in (2) are estimable from the observed proportions. The value of p_{11} results from a mixture of compliers and always-takers, and the value of p_{00} results from a mixture of compliers and never-takers. A1-A4 identifies the complier contribution in each mixture.

The ER implies that, for always-takers and never-takers, there is no direct effect of treatment on response, i.e., $\Pr(Y_i(1) = 1|S_i = s, Z_i = z) = \Pr(Y_i(0) = 1|S_i = s, Z_i = z)$, where $s \in \{a, n\}$. If, instead, there is an unknown direct effect of treatment on response (as in Figure 1e) for the always-takers or never-takers, then for some $s \in \{a, n\}$ we have

$$\tau_s^* = \Pr(Y_i(1) = 1|S_i = s, Z_i = z) - \Pr(Y_i(0) = 1|S_i = s, Z_i = z) \neq 0, \quad \text{for } z = 0, 1.$$

The direct effect τ_s^* is constant across treatment arms z for $s \in \{a, n\}$ for coherency. We do not define an analogous τ_c^* as it equals θ_c^*; hence, the CACE includes both direct effect of treatment on response and indirect effects carried through the intermediate variable.

No unmeasured confounding implies that the principal strata distributions are the same across treatments, so that $\Pr(S_i = a) = \Pr(S_i = a|Z_i = 0) = \pi_{10}$ and $\Pr(S_i = n) = \Pr(S_i = n|Z_i = 1) = \pi_{01}$. This is false if there is S-confounding, where for some $s \in \{a, n\}$ (c stratum is automatically determined by a and n), we have

$$\xi_s^* = \Pr(S_i = s|Z_i = 1) - \Pr(S_i = s|Z_i = 0) \neq 0.$$

With the ER, no unmeasured confounding further implies that the distribution of outcomes for the always-takers and never-takers are the same across treatments, so that

$$\begin{aligned} p_{10} &\overset{A2}{=} \Pr(Y_i^{obs}=1|S_i=a,Z_i=0) \overset{A4}{=} \Pr(Y_i(0)=1|S_i=a) \overset{A3}{=} \Pr(Y_i(1)=1|S_i=a), \\ p_{01} &\overset{A2}{=} \Pr(Y_i^{obs}=1|S_i=n,Z_i=1) \overset{A4}{=} \Pr(Y_i(1)=1|S_i=n) \overset{A3}{=} \Pr(Y_i(0)=1|S_i=n). \end{aligned}$$

It also implies that $\Pr(Y_i(0),Y_i(1)|S_i=c,Z_i=1)=\Pr(Y_i(0),Y_i(1)|S_i=c,Z_i=0)$. These fail in the presence of Y-confounding, since even with the ER, for some $s\in\{a,n,c\}$ we have

$$\eta_s^* = \Pr(Y_i(z)=1|S_i=s,Z_i=1)-\Pr(Y_i(z)=1|S_i=s,Z_i=0)\neq 0, \quad \text{for } z=0,1.$$

For coherency, the η_s^* is constant across potential outcomes $Y(z)$ for each $s\in\{a,n,c\}$.

When A3 and A4 fail, i.e., $\tau^*_{\{a,n\}}\neq 0, \eta^*_{\{a,n,c\}}\neq 0$, and $\xi^*_{\{a,n\}}\neq 0$, (2) is a biased estimator of θ_c^* since, letting $p_{sz}=\Pr(Y_i(z)=1|S_i=s,Z_i=z)$ and $\pi_{sz}=\Pr(S_i=s|Z_i=z)$, we have

1. $p_{10}=p_{a1}-\tau_a^*-\eta_a^*$ rather than p_{a1} when $\tau_a^*,\eta_a^*\neq 0$,
2. $p_{01}=p_{n0}+\tau_n^*+\eta_n^*$ rather than p_{n0} when $\tau_n^*,\eta_n^*\neq 0$,
3. $\pi_{10}=\pi_{a1}-\xi_a^*$ rather than π_{a1} when $\xi_a^*\neq 0$,
4. $\pi_{01}=\pi_{n0}+\xi_n^*$ rather than π_{n0} when $\xi_n^*\neq 0$, and
5. $\Pr(Y_i(1)|S_i=c,Z_i=1)-\Pr(Y_i(0)|S_i=c,Z_i=0)=\theta_c^*+\eta_c^*$, when $\eta_c^*\neq 0$.

We can use these facts to define a new estimator of θ_c^* when A3 and A4 do not hold, namely

$$\hat{\theta}_c^{adj} = -\eta_c^* + \frac{p_{11}\pi_{11}-(p_{10}+\tau_a^*+\eta_a^*)(\pi_{10}+\xi_a^*)}{1-\pi_{01}-(\pi_{10}+\xi_a^*)} - \frac{p_{00}\pi_{00}-(p_{01}-\tau_n^*-\eta_n^*)(\pi_{01}-\xi_n^*)}{1-\pi_{10}-(\pi_{01}-\xi_n^*)}. \tag{3}$$

A key observation from this formulation is that for $s\in\{a,n\}$, observable direct effects τ_s^* and Y-confounding effects η_s^* are not distinguishable since both always appear together as a sum. This is apparent in Figures 1d and 1e, which are indistinguishable as data generating mechanisms. Thus, operationally, for $s\in\{a,n\}$, τ_s^* and η_s^* can be treated as a single parameter. For example, in (3), we can define $\delta_s^*=\tau_s^*+\eta_s^*$ to represent the direct effect plus Y-confounding.

2.3 Illustration of Confounding

Applying (2) in the presence of S-confounding and Y-confounding can result in invalid conclusions. However, adjusting for confounding using (3) can correct these problems. To illustrate this, we use the data in Table 1, which mimics the observed π_{dz} and p_{dz} data from a flu vaccine trial described in [20], ignoring covariates. For now, the response and treatment are left context-free to emphasize the generality of these issues.

[Table 1 about here.]

Since we only see D and not S, many population proportions consistent with the data in Table 1 exist. For example, the topmost example in Table 2 has no confounding: the proportions of always-takers and never-takers, and proportions of $Y^{obs} = 1$ within these two strata do not change with Z. The true $\theta_c^* = .001 - .117 = -0.116$, and $\hat{\theta}_c^{obs}$ correctly estimates θ_c^*. Alternatively, the middle of Table 2 shows one S-confounding example where the proportions of principal strata differ across Z ($\xi_a^* = 0.13$ and $\xi_n^* = -0.09$). Finally, the bottommost example in Table 2 is a Y-confounding example, where the outcome proportions differ across Z within the always-taker and never-taker strata ($\delta_a^* = -0.019$ and $\delta_n^* = -0.020$).

[Table 2 about here.]

Regardless of which population proportions from Table 2 are true, $\hat{\theta}_c^{obs} = -0.116$. However, the true θ_c^* for the various settings from Table 2 are approximately -0.116, -0.053, and 0.023. Clearly, $\hat{\theta}_c^{obs}$ is biased for θ_c^* in cases of S- and Y-confounding. The bias is striking when interpreted as proportion change from baseline: The true θ_c^* values represent a 99% reduction, a 54% reduction and a 230% increase in rates. Using (3) with correctly specified values of $\delta_{\{a,n\}}^*$ and $\xi_{\{a,n\}}^*$ (and $\eta_c^* = 0$) implied by Table 2 appropriately adjusts $\hat{\theta}_c^{adj}$ so that it is consistent for θ_c^*. This holds for simultaneous S-confounding and Y-confounding as well.

In practice, δ_s^* and ξ_s^* are not known, so analysts should examine the sensitivity of conclusions to a range of their plausible values. For example, $\xi_a^* = 0.13$ implies an additional 13% more always-takers for units with $z = 1$ than with $z = 0$; and, $\eta_a^* = -0.019$ implies that, among always-takers, the probability of $Y_i(z) = 1$ is about 2% smaller for units with $z = 1$ than with $z = 0$. Analysts can determine the values of δ_s^* and ξ_s^* that alter conclusions based on $\hat{\theta}_c^{obs}$, and judge the plausibility of those values. We do not present examples of these approaches here, preferring instead to illustrate sensitivity analysis for model-based application of PS.

3 Sensitivity Analysis using Parametric Models

3.1 Parametric Models

The non-parametric estimator $\hat{\theta}_c^{adj}$ is useful when observed covariates X are deemed not to cause bias, e.g., when they are finely balanced across treatment groups and are not strongly predictive of S. In settings where this is not the case, it may be possible to stratify samples by X (or perhaps by percentiles of propensity scores) and use $\hat{\theta}_c^{adj}$ separately in each stratum. When this is not possible, analysts can control for X using parametric models for PS. This has several potential advantages over the nonparametric approach, including (1) parametric modeling readily adjusts for multiple observed covariates, which can reduce bias and improve precision, and (2) parametric modeling offers conceptually straightforward ways to incorporate complexities like multilevel structure, multiple outcomes, and latent variables. A disadvantage of parametric approaches is the risk of model misspecification, making it

essential for analysts to employ model checking procedures, e.g., posterior predictive checks in the Bayesian paradigm, and examine alternate model specifications.

Typically, two models are specified in PS analysis: one for the marginal distribution of S_i given (Z_i, X_i) and one for the conditional distribution of $Y_i(z)$ given (S_i, Z_i, X_i). When both Z_i and D_i are binary, a natural and common choice for the S_i model is the multinomial logit regression model [12]. Using compliers as the reference group, we have

$$\log \frac{\Pr(S_i = s|Z_i, X_i)}{\Pr(S_i = c|Z_i, X_i)} = X_i\beta_s + Z_i\xi_s, \quad s \in \{a, n\}, \tag{4}$$

where X includes an intercept term and $\Pr(S_i = c|Z_i, X_i) = 1 - \sum_{s\in\{a,n\}} \Pr(S_i = s|Z_i, X_i)$. As with ξ_s^* in Section 2.2, ξ_s in (4) represents S-confounding; however, ξ_s is a different parameter than ξ_s^* defined on a multiplicative scale. Specifically, for $s \in \{a, n\}$ we have

$$\exp(\xi_s) = \frac{\Pr(S_i = s|Z_i = 1, X_i = x)/\Pr(S_i = c|Z_i = 1, X_i = x)}{\Pr(S_i = s|Z_i = 0, X_i = x)/\Pr(S_i = c|Z_i = 0, X_i = x)}. \tag{5}$$

Each ξ_s is assumed to be constant across x. Thus, ξ_s is the conditional odds ratio for being in stratum $s \in \{a, n\}$ versus being a complier when going from $Z = 0$ to $Z = 1$, given X. For instance, the middle example of Table 2 was created using $\exp(\xi_a) = 1/1.5$ and $\exp(\xi_n) = 1.5$, so that the ratio of never-takers to compliers within each level of X increases by a factor of 1.5 when going from $Z = 0$ to $Z = 1$. Since there are no covariates, this corresponds to $\xi_a^* = 0.13$ and $\xi_n^* = -0.09$ in the notation of Section 2.2.

For sensitivity analysis in practice, it is convenient to select the range of ξ_a and ξ_n to be examined on the basis of the observed principal strata probabilities, π_{dz}, resulting from aggregating the data across the levels of X. For example, in Table 1, $\pi_{01} = \pi_{n1} = .69$ and $\pi_{01} = \pi_{a0} = .12$. To examine a level of confounding that could potentially result in, say, $\pi_{n0} = .69 \pm .1$ and $\pi_{n0} = .12 \pm .05$, we could first create a grid of ξ_a^* and ξ_n^* sensitivity specifications (as defined in Section 2.2) that produced the values of π_{n0} and π_{a1} under consideration, and convert that grid to ξ_a and ξ_n values to be subsequently used in model-based sensitivity analysis. A complementary approach is to specify bounds for each ξ_s using observed covariate magnitudes. For example, researchers may hypothesize that the magnitude of ξ_n could be up to twice the magnitude of the largest (standardized) estimated β_n. Finally, one could set each ξ_s via interpretations of the odds ratios, e.g., set $\exp(\xi_a) = 2$ so that the odds of never-takers to compliers within each level of X doubles when going from $Z = 0$ to $Z = 1$.

Binary potential outcomes $Y_i(z)$ can be modeled using a logistic regression,

$$\operatorname{logit} \Pr(Y_i(z) = 1|Z_i = z, S_i, X_i) = X_i\alpha_x + I_{S_i=c}Z_i(\theta_c + \eta_c) + \sum_{s'\in\{a,n\}} I_{S_i=s'}(\alpha_{s'} + Z_i\delta_{s'}) \tag{6}$$

where $I_{S_i=s'}$ is an indicator function that equals one if $S_i = s'$ and equals zero otherwise. We take the CACE to be

$$\theta_c = \operatorname{logit} \Pr(Y_i(1) = 1|Z_i = z, S_i = c, X_i) - \operatorname{logit} \Pr(Y_i(0) = 1|Z_i = z, S_i = c, X_i) \tag{7}$$

which is fixed to be the same value for all compliers regardless of z. We also define

$$\eta_c = \text{logit}\Pr(Y_i(z) = 1|Z_i = 1, S_i = c, X_i) - \text{logit}\Pr(Y_i(z) = 1|Z_i = 0, S_i = c, X_i). \quad (8)$$

for all compliers for any z. Similarly to θ_c, for $s \in \{a, n\}$, we have

$$\exp(\delta_s) = \text{logit}\Pr(Y_i(1) = 1|Z_i = z, S_i = s, X_i) - \text{logit}\Pr(Y_i(0) = 1|Z_i = z, S_i = s, X_i). \quad (9)$$

For computational convenience, we assume that $(\theta_c, \eta_c, \delta_a, \delta_n)$ are constant across x. As with $(\theta_c^*, \eta_c^*, \delta_a^*, \delta_n^*)$, $(\theta_c, \eta_c, \delta_a, \delta_n)$ represent, respectively, a CACE, Y-confounding for compliers, and Y-confounding plus direct effect for always-takers and never-takers. However, as a result of the non-collapsibility of logistic regression models [21], $(\theta_c, \eta_c, \delta_a, \delta_n)$ in (6) must be interpreted conditional on Z and X in contrast to the interpretations of $(\theta_c^*, \eta_c^*, \delta_a^*, \delta_n^*)$. This issue relates to another potential disadvantage of the parametric approach compared to the nonparametric one: one must consider estimands and corresponding sensitivity parameters that are natural to scale imposed by the parametric model (e.g., log odds ratio in logistic regression), whereas with the nonparametric approach there is no such constraint and one can consider a range of estimands such as difference or relative risk.

The sensitivity parameters can be interpreted via odds ratios. For instance, the bottommost example of Table 2 was created using $\exp(\delta_a) = \exp(\delta_n) = 1/1.25$, so that the odds of $Y_i(z) = 1$ are 1.25 times greater when $z = 0$ than when $z = 1$ for both always-takers and never-takers. These correspond to $\delta_a^* = -0.019$ and $\delta_n^* = -0.020$ (with difference due to rounding). The CACE, θ_c, is indistinguishable from the Y-confounding effect in the compliers strata; however, it is identifiable after specification of η_c (and the other sensitivity parameters). For instance, in the bottommost example in Table 2, $\exp(\theta_c + \eta_c) = (.033/.967)/(.01/.99) = \exp(1.21)$, so that θ_c is not identified until η_c is specified.

Specification of $\delta_{\{a,n\}}$ and η_c can be based on subject-matter knowledge. For η_c, this involves the extent to which the odds for $Y_i(z) = 1$ within the complier strata could change across z as a result of Y-confounding only. For $\delta_{\{a,n\}}$, interpretation involves the extent to which the odds for $Y_i(z) = 1$ within the always-taker and never-taker strata could change across z as a result of direct effects or Y-confounding. If helpful, analysts can decompose δ_s into its components from Y-confounding and direct effects. If the scientific experts do not suspect direct effects, δ_s could reflect only the effects of confounding. Absent or as a complement to subject-matter knowledge, analysts can specify η_c and $\delta_{\{a,n\}}$ via methods similar to those for setting ξ_s. We demonstrate such methods in Section 5.

The models in (4) and (6) use additive effects of Z and sensitivity parameters that do not vary with X. It is possible to include interactions of Z and X in the predictor functions and use the methodology as indicated. However, this presumes that the confounding effects are constant (on the log odds scale) across all levels of X, which may not be sensible if one presumes that treatment effects differ with X. Arguably, in scenarios with interactions of Z and X, analysts should specify sensitivity parameters for each level of X that is interacted with Z. This can create a large number of sensitivity parameters, making computation and interpretation cumbersome. Hence, the sensitivity analysis approach here is most appropriate for settings with additive treatment effects. Similarly, because the sensitivity parameters

do not depend on X, they represent overall effects of confounding (averaged across covariates) across treatment groups. When confounding does not vary within the levels of X, this specification is completely adequate. When there are differences in confounding, the sensitivity analyses may be too coarse, resulting in inaccurate results. The nature of this inaccuracy and its dependence on X is uncertain and a subject for further research.

3.2 Estimation of CACE with Sensitivity Parameters

Given Z_i and X_i, the analyst can model Y_i^{obs} and D_i with

$$\Pr(Y_i^{obs}, D_i = d_i | Z_i = z_i, X_i) \;\; = \sum_{s \in \mathcal{S}(z_i, d_i)} \Pr(Y_i^{obs} | S_i = s, z_i, X_i) \Pr(S_i = s | z_i, X_i), \quad (10)$$

where $\mathcal{S}(z_i, d_i)$ denotes the set of all possible principal strata that are consistent with the observed z_i and d_i. The two distributions on the right side of (10) are specified by (6) and (4).

As illustrated by an example in Section 4, without any further constraints, the sensitivity parameters are usually not identifiable from the data since there is no observed information in the data about the confounding structure underlying (10). We thus recommend the following multi-step sensitivity analysis procedure. First, specify ξ_s and δ_s for $S \in \{a, n\}$ and estimate $\theta_c + \eta_c$. Second, specify η_c to identify θ_c. The estimation process is repeated for the range of plausible values of ξ_s, δ_s, and η_c.

For any fixed set of sensitivity parameters, the estimation of $\theta_c + \eta_c$ can proceed using an Expectation Maximization (EM) algorithm [22] or Bayesian data augmentation [10]. EM finds posterior modes comparatively quickly, whereas full Bayesian inference automatically provides measures of inferential uncertainty (given the values of the sensitivity parameters). The EM algorithm alternately replaces the unobserved S_i with their expected values given current draws of the parameters, and maximizes the parameters given the expected values of all S_i. For the Bayesian analysis, after first specifying prior distributions—we use the added data conjugate prior distribution of Hirano et al. (2000)—analysts can sample from the posterior distributions of $\theta = (\beta_a, \beta_n, \alpha_x, \alpha_a, \alpha_n, \theta_c + \eta_c)$ using Metropolis proposals within a Gibbs sampler. To accomplish this, the posterior distributions of each S_i must be sampled, each instance of which results in new covariate matrices and response vectors in (6) and (4), respectively. Since a given imputation of S_i may result in a likelihood that is maximized on the boundary of the parameter space (e.g., $\theta_c + \eta_c = -\infty$), the prior distributions play a key role in stabilizing the sampling. Mixing can be improved by parameterizing (6) without an intercept [12] and by using a Metropolis subchain rather than a single proposal for θ to better follow the fast mixing principal strata. We recommend that analysts obtain MLEs from EM, and initialize the Gibbs sampler with the MLEs to obtain point and interval estimates.

4 Demonstration using Introduced Confounding in a Randomized Study

In this section we show that unadjusted model-based PS estimation of θ_c is biased in the presence of S-confounding and Y-confounding, but that analysts can recover the truth using the sensitivity methodology. To do so, we manipulate data from a randomized experiment to induce unmeasured confounding in known ways, and examine EM point estimates using the known correct sensitivity parameter specifications. To streamline the presentation, we postpone full Bayesian analysis to Section 5. We assume A1 and A2. SUTVA can be tenuous in infectious disease contexts, but we do not deal with the complication in this article. Monotonicity is plausible in this setting.

We use data from the second year (1979-1980) of the study done by [20], which is a randomized encouragement design in which $Z_i = 1$ if person i is encouraged to take an influenza vaccine by his/her physician and $Z_i = 0$ otherwise. The intermediate variable is actual receipt of the vaccine, with $D_i = 1$ if person i indeed takes the vaccine and $D_i = 0$ otherwise. The response, flu-related hospitalization, is $Y_i^{obs} = 1$ if person i gets the flu and $Y_i^{obs} = 0$ otherwise. The randomization is done at the level of physician rather than patient, so that the data are actually clustered. We ignore this feature of the data for illustrations.

The available covariates comprise age in years, sex, race (white/non-white), chronic obstructive pulmonary disease (COPD), heart disease (HD), diabetes, renal disease, and liver disease for 2901 participants. The covariates are closely balanced across encouragement arms. Regression analyses indicate that higher age and COPD are predictive of taking the vaccine, whereas HD and COPD are predictive of getting the flu. The predictive role of these three variables closely mirrors that of the first three eigenvectors of a principal components analysis of the covariates, which capture age, an approximate COPD/sex/race relationship, and an approximate heart disease/diabetes relationship. The variation captured in the remaining components does not provide any further predictive benefit. We transform age to a four level factor (< 40; $(40, 60]$; $(60, 80]$; ≥ 80) based on the observed relationship between age and taking the vaccine. Alternative covariate specifications would not entail a different model fitting approach. Restriction to complete cases using age, COPD, and HD yields 2893 participants.

Adopting a naive interpretation, the observed relationships suggest that (1) older populations generally have more always-takers and compliers than comparable younger populations, i.e., the distribution of S_i varies with age; and, (2) HD pervasive populations generally have greater flu prevalence relative to comparable heart healthy populations, i.e., the distribution of $(Y_i(0), Y_i(1))$ varies with heart disease prevalence. Thus, in a similar but hypothetical observational study, if elderly people are more likely to receive encouragement but age was not controlled for, there would be a higher proportion of compliers and always-takers in the treatment arm, resulting in S-confounding. Likewise, if HD patients are more likely to receive encouragement but HD was not controlled for, there would be a greater proportion of individuals at risk for flu in the same arm, resulting in Y-confounding. The operational distinction between S-confounding and Y-confounding is not clear-cut since age and HD have

some association. Indeed, the example of COPD directly suggests that S-confounding and Y-confounding may be intimately connected. Nonetheless, separating S-confounding and Y-confounding offers both operational and conceptual convenience in sensitivity checks, as we shall discuss. In fact, as there is often overlap between S-confounding and Y-confounding, comparing to a combined analysis with a single set of parameters characterizing the joint confounding, this separation may lead to more conservative conclusions, which is usually not a concern in sensitivity analysis.

For our demonstration, we use discretized age and COPD in the sub-model for S_i, and COPD and HD in the sub-model for $Y_i(Z_i)$. We use the data with complete cases, except we discard two more observations (with $Z = 0$, $D = 1$, $Y = 1$, age= 1, COPD= 1, and HD = 0/1) so that a no S-confounding and no Y-confounding specification is consistent with the observed data. Without this adjustment, a specification of no S-confounding and no Y-confounding results in an MLE that lies on the boundary of the parameter space, i.e., $\Pr(Y_i(1) = 1|S_i = c, Z_i = 1, X_i = x_i) = 0$. We refer to this reduced data set as the test data. In practice, if a given sensitivity parameter specification results in extreme estimates of $\Pr(Y_i(z) = 1|S_i = s, Z_i = z, X_i = x_i)$, e.g., 0 or 1, it is likely not consistent with the observed data.

We take the truth to be the estimated coefficients in (10) for the test data without any sensitivity adjustments, i.e., all sensitivity parameters equal zero. The MLE for the CACE in the test data is $\hat{\theta}_c = -1.87$.

We introduce S-confounding by removing half of the observed never-takers in the $Z_i = 1$ arm and half of the observed always-takers in the $Z_i = 0$ arm. Removal was done randomly but ensuring that $\Pr(Y_i(0) = 1|S_i = a, Z_i = 0)$ and $\Pr(Y_i(1) = 1|S_i = n, Z_i = 1)$ were not changed from the observed probabilities in the test data. This guards against inadvertently inducing Y-confounding, and ensures that the covariate and flu outcome relationships are not changed. Observed covariate balance remains good for age, COPD, and HD after this manipulation. Since half of the never-takers in the $Z_i = 1$ arm and half of the always-takers in the $Z_i = 0$ arm have been removed, and $\Pr(Y_i^{obs} = 1|S_i, Z_i, X_i)$ does not change, the S-confounding sensitivity parameters for this manipulation are $\exp(\xi_a) \approx 2$ and $\exp(\xi_n) \approx 1/2$.

Y-confounding is introduced by keeping only HD = 1 individuals in the $T_i = 1$ arm, and not using HD as a covariate so that HD is an unmeasured confounder. After the manipulation, 56.2% of individuals have HD = 1 in the $T_i = 0$ arm, and 100% of individuals have HD = 1 in the $T_i = 1$ arm. The distributions of age and COPD remain balanced in the treatment arms after the manipulation. This was applied on top of the S-confounding manipulation, but since HD does not strongly associate with D, we suspect that it will not drastically alter the previously induced S-confounding of $\exp(\xi_n) \approx 1/2$. However, if HD is more prevalent among compliers than always-takers, or vice-versa, then $\exp(\xi_n) \approx 1/2$ will no longer hold. The log odds ratios of the outcomes without covariate adjustment before and after the manipulations were -0.163 and 0.086, respectively, which corresponds to approximately correct sensitivity parameters for the Y-confounded data of $\delta_a = \eta_a = \delta_n = \eta_n = \eta_c \approx 0.25$.

We fit the model implied by (10) to the confounded data with a variety of possible values

for the sensitivity parameters. Figure 2 displays the results in two panels. The top panel shows contour plots for $\hat{\theta}_c$ across a variety of combinations of $\exp(\xi_a) \in [1/3, \cdots, 3]$ and $\exp(\xi_n) \in [1/3, \cdots, 3]$ with $\eta_c = \delta_a = \delta_n = 0$. The bottom panel shows the contours for the same range for ξ_a and ξ_n with $\eta_c = \delta_a = \delta_n = 0.25$.

[Figure 2 about here.]

As seen in the top panel of Figure 2, fitting PS in the confounded data ignoring unmeasured confounding results in a biased estimate of the CACE. Correctly specifying $\exp(\xi_a) = 2$ and $\exp(\xi_n) = 1/2$, but wrongly setting $\eta_c = \delta_a = \delta_n = 0$ also results in a biased estimate. As evident in the bottom panel of Figure 2, using the approximately correct sensitivity specifications for S-confounding and Y-confounding nearly recovers the CACE estimate. Examinations of the α and β parameters show similar results. Allowing sensitivity parameters to be estimated by the data rather than be pre-specified results in the EM algorithm finding $\hat{\eta}_a = \hat{\eta}_n = .027$, $\hat{\xi}_a = 1.42$ and $\hat{\xi}_n = .80$. These result in $\hat{\theta}_c = .128$ (with $\eta_c = 0$). Estimation when allowing $\hat{\eta}_a \neq \hat{\eta}_n$ resulted in $\hat{\eta}_a = -.32, \hat{\eta}_n = .011, \hat{\xi}_a = 1.77, \hat{\xi}_n = .88$, and $\hat{\theta}_c = .42$. Hence, in all cases, using MLE with free sensitivity parameters results in biased estimates.

Figure 2 provides a visualization of the topographical nature of potential confounding. The primary benefit of these plots, however, is to provide a diagnostic alternative to careful bound specification for $\xi_{\{a,n\}}$, η_c, and $\delta_{\{a,n\}}$. Analysts instead can consider large spaces of potential values for the parameters, construct plots like Figure 2, and identify the levels of confounding that would alter study conclusions. These values can be interpreted using (5) and (9), so that scientific experts can decide if the identified levels are plausible enough to cast doubt on conclusions. This approach is related to the sensitivity checks done by [23] in observational study contexts that do not involve PS.

5 Application to the Swedish NMC data

We now apply the sensitivity methodology to the observational Swedish NMC study. The NMC was conducted in year 1997, when 300,000 Swedes participated in a national fundraising event organized by the Swedish Cancer Society. Each participant was asked to complete a questionnaire that included items on known or suspected risk factors for cancer and cardiovascular disease (CVD). These individuals were followed from year 1997 to 2004 using the Swedish patient registry, and each cancer and CVD event was recorded. We seek to investigate the causal effect of physical activity (PA) on CVD mediated through body mass index (BMI). Following Sjölander et al. [5], our analysis assumes PA drives BMI and does not examine possible reverse causality. Further details on the NMC can be found in [24].

For each subject i, $Z_i = 1$ if he/she reported having low PA and $Z = 0$ otherwise; $D_i = 1$ if he/she had BMI greater than 30 in the baseline year and $D_i = 0$ otherwise; and, $Y_i = 1$ if he/she had at least one recorded CVD event during follow-up and $Y_i = 0$ otherwise. PS analysis of the same dataset with BMI being treated as a continuous immediate variable can be founded in Schwartz et al. [25]. Among all subjects, 38,349 reported high PA and 2,956

reported low PA. The former included 2,262 cases of CVD, and the latter included 172 cases. Adapting the non-compliance language, in this setting the always-takers and never-takers are the subjects who would be obese and not obese, respectively, regardless of their PA level; the compliers are the subjects who would be obese if they did not exercise and not obese if they exercised. We believe that SUTVA and monotonicity are plausible in this setting.

In the data we analyzed, the only available covariate is age recorded in days. It is well known that age is highly predictive of CVD and PA, so that we should control for age in the analysis. We first balance the covariate distribution of age in the treatment groups by conducting one-to-one nearest neighbor matching without replacement on age. This results in 2,956 pairs of high-exercisers and low-exercisers, with 111 and 172 CVD cases, respectively.

We fit the models in (4) and (6) on the matched dataset, including age as a covariate, to estimate θ_c. We examined models including an interaction between treatment and age, but the interaction coefficient was insignificant and the estimated treatment effect did not change substantially, so we chose not to include the interaction. This model check, while not conclusive, suggests that the treatment effect does not vary across the levels of X. We use the Bayesian analysis described in Section 3 to estimate a posterior mode for θ_c of 4.9; the 95% credible interval does not include zero, indicating a higher risk of CVD among low exercisers whose weight would be impacted by exercise. However, even for the matched dataset, the assumption of no unmeasured confounding is questionable, because people with high PA differ from those with low PA in ways that are related to CVD risk, e.g., diet and life style. In addition, it is widely accepted in the medical community [e.g. 26] that PA has direct effects on CVD, implying that the ER is not applicable. Because of these potential violations, we perform sensitivity analysis for θ_c.

To propose a range for potential S-confounding sensitivity, we consider the observed principal strata probabilities aggregated over age, i.e., $\Pr(D = 1|Z = 0) = \pi_{a0} = .073$ and $\Pr(D = 0|Z = 1) = \pi_{n1} = .869$. We posit that S-confounding could result in $\pi_{a1} = .073 \pm .02$ and $\pi_{n0} = .869 \pm .02$, which implies values of $(\pi_{c0}, \pi_{c1}) \in [.04, .08]$. Thus, we allow for up to a two-fold difference (in either direction) in the percentage of compliers in the treatment arms, which seems a reasonably strong amount of S-confounding. We proceed by making a grid of values $(\pi_{a1}, \pi_{n0}) \in [-.02, .02] \times [-.02, .02]$, each entry of which, when combined with the observed π_{a0} and π_{a1}, identifies π_{c0} and π_{c1}. Each point in this grid is converted to specifications for ξ_a and ξ_n using (5) without covariates resulting in $(\xi_a, \xi_n) \in [-0.66, 0.66] \times [-.90, 1.00]$. These bounds are crude because the principal strata probabilities actually vary with age, but they enable sensible assessments of the effects of S-confounding on estimation of θ_c.

To propose a range for Y-confounding and direct effects sensitivity, we make the simplifying assumption that $\delta_a = \delta_n$, i.e., always-takers and never-takers share the same Y-confounding plus direct effect, so that it can be represented by one coefficient in (6). In language of the NMC, setting $\delta_a = \delta_n$ implies that the difference between those who exercise frequently and those who do not are the same for people whose BMI is always low and for people whose BMI is always high (regardless of exercise). A similar assumption is used by [5] in their estimation of direct effects of PA on CVD. Setting $\delta_a = \delta_n$ is largely motivated

by parsimony: it enables us to explore a three-dimensional sensitivity space as opposed to a four-dimensional one. In principle, when the equality is far from plausible, analysts could conduct the four-dimensional sensitivity analysis, although this can be unwieldy. Alternatively, analysts could identify the maximum δ-effect for the always-takers and never-takers, and set both δ_a and δ_n equal to that maximum as a "worse-than-expected" scenario.

We expect any direct effects to increase CVD incidence, because the treatment is low PA. Furthermore, potential Y-confounding most likely would increase CVD incidence, since the high-exercisers may maintain other-CVD protective habits beyond regular PA. Therefore, we examine the sensitivity parameters $\delta_a = \delta_n \in [0, 1.1]$ and $\eta_c \in [0, .5]$. The former corresponds to a maximum three-fold ($e^{1.1} = 3.0$) increase in the odds of getting CVD due to direct effect and Y-confounding for individuals whose BMI is not affected by PA, and the later corresponds to a maximum 1.7-fold ($e^{0.5} = 1.7$) increase in the odds of getting CVD due to direct effect and Y-confounding for individuals whose BMI is affected by PA. We note that δ_a and δ_n are interpreted as the effects of unmeasured confounding only when the ER is assumed to hold.

Figure (3) shows the sensitivity of CACE estimates from the logistic regression to these levels of potential S-confounding and Y-confounding plus direct effect, including uncertainty in the CACE estimates via 'maximal' point-wise 95% credible intervals. As is seen in Figure (3), there is little sensitivity to ξ_n, larger ξ_a implies slightly larger θ_c, and the sign of the estimated θ_c is sensitive to $\delta_a = \delta_n$. Thus, for $\delta_a = \delta_n < .4$, i.e., 1.5 fold increase in odds, $(\xi_a, \xi_n) \in [-0.66, 0.66] \times [-.90, 1.00]$, and $\eta_c \in [0, .5]$, there is a significant and large protective of benefit PA, mediated through the effect on BMI, on CVD. Evidence for a negative significant effect requires $\delta_a = \delta_n$ approximately larger than .9, indicating a $e^{.9} = 2.5$ fold increase in the odds of getting CVD due to Y-confounding plus direct effect for individuals whose BMI is not affected by PA.

[Figure 3 about here.]

Using the same data, [5] found evidence of a significant direct effect of PA, with a point estimate of 0.26 and standard error of .085. They suggested that this estimate was conservative and the true effect could be larger. Our analysis suggests that evidence for a protective indirect effect of reduced BMI on CVD as a result of PA depends primarily on the strength of δ_a (and δ_n). Scientific experts who believe that the direct protective benefits of PA and the effects of unmeasured confounders produce greater than a 1.5-fold decrease in the odds of CVD should be skeptical of the beneficial effect on CVD of reducing BMI via PA, whereas those who believe that such ratios are unlikely can feel confident in the unadjusted PS conclusions.

6 Concluding Remarks

While facilitating rich investigations of the robustness of results to unmeasured confounding, conducting a full sensitivity analysis involves specification of many parameters. Analysts may choose to reduce the number of free parameters for rougher but faster checks. For example,

setting $\delta_a = \delta_n = \eta_c$ will collapse Y-confounding to one parameter. This implies that the unobserved confounders are distributed uniformly across the strata, i.e., that Y-confounding does not vary by S, which may be a useful simplification even if not strictly true.

While the interpretation of each sensitivity parameter does not depend on the settings of the other sensitivity parameters, the parameters are not variation independent. For example, the effects of setting $\delta_n = 1$ differ when $\xi_n = 1$ than when $\xi_n = -1$, and some combinations actually cannot be possible given the data. This motivates why we recommend that analysts examine plots like Figure 2 for a wide array of combinations of parameters to identify scenarios in which S- and Y-confounding alter study conclusions. Scientific experts then can evaluate the plausibility of those sensitivity regions taking the subject matter into account. Impossible combinations are indicated by non-sensible results, such as parameter estimates going off to infinity or negative estimates of various probabilities. While not pursued in this article, it would be possible to produce an algorithm to find sensitivity parameter bounds on the basis of consistency of MLE or posterior sampling results.

The methods proposed here readily extend to the case of non-binary outcomes. Further development is required to adapt the methods to continuous intermediate variables, such as BMI in its original scale (e.g., as in [25]). Moreover, we did not explore sensitivity to model mis-specification in PS analysis, which can be as crucial as the structural assumptions since model-based PS inference usually involves weakly identified models. Under the Bayesian paradigm, this can be examined via tools like posterior predictive checks; this is a subject for further investigation.

The Matlab code of the EM algorithm and Bayesian data augmentation are available at `www.stat.tamu.edu/~scott/PSsensitivity`.

Acknowledgments

We thank the editor, the associate editor and two reviewers for their constructive comments and suggestions that helped to improve the manuscript significantly. We also thank Clem McDonald, Siu Liu Hiu and Bill Tierney for providing the influenza data, and Olof Nyren, Rino Bellocco and Arvid Sjölander for providing the Swedish NMC data.

References

[1] Rosenbaum P. The consquences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series B* 1984; **147**(5):656–666. URL `http://www.jstor.org/stable/2981697`.

[2] Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 1974; **66**(1):688–701, DOI: 10.1037/h0037350.

[3] Frangakis C, Rubin D. Principal stratification in causal inference. *Biometrics* 2002; **58**(1):21–29, DOI: 10.1111/j.0006-341X.2002.00021.x.

[4] Rubin D. Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics* 2004; **31**(1):161–170, DOI: 10.1111/j.1467-9469.2004.02-123.x.

[5] Sjölander A, Humphreys K, Vansteelandt S, Bellocco R, Palmgren J. Sensitivity analysis for principal stratum direct effects, with an application to a study of physical activity and coronary heart disease. *Biometrics* 2009; **65**(2):514–520, DOI: 10.1111/j.1541-0420.2008.01108.x.

[6] Mattei A, Mealli F. Application of the principal stratication approach to the faenza randomized experiment on breast self-examination. *Biometrics* 2007; **63**:437–446, DOI: 10.1111/j.1541-0420.2006.00684.x.

[7] Grilli L, Mealli F. Nonparametric bounds on the causal effect of university studies on job opportunities using principal stratification. *Journal of Educational and Behavioral Statistics, 33, 1, 111-130.* 2008; **33**(1):111–130, DOI: 10.3102/1076998607302627.

[8] Jo B, Vinokur A. Identifying assumption sensitivity analysis and bounding of causal effects with alternative. *Journal of Educational and Behavioral Statistics* 2011; DOI: 10.3102/1076998610383985. URL `http://jeb.sagepub.com/content/early/2011/01/29/1076998610383985`.

[9] Roy J, Hogan J, Marcus B. Principal stratification with predictors of compliance for randomized trials with 2 active treatments. *Biostatistics* 2008; **9**(2):277–289, DOI: 10.1093/biostatistics/kxm027.

[10] Imbens G, Rubin D. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics* 1997; **25**(1):305–327. URL `http://www.jstor.org/stable/2242722`.

[11] Imbens G, Angrist J. Identification and estimation of local average treatment effects. *Econometrica* 1994; **62**:467–476.

[12] Hirano K, Imbens G, Rubin D, Zhou XH. Assessing the effect of an inuenza vaccine in an encouragement design. *Biostatistics* 2000; **1**(1):69–88, DOI: 10.1093/biostatistics/1.1.69.

[13] Elliott M, Raghunathan T, Li Y. Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics* 2010; **11**(1):353–372, DOI: 10.1093/biostatistics/kxp060.

[14] Griffin B, McCaffrey D, Morral A. An application of principal stratification to control for institutionalization at follow-up in studies of substance abuse treatment programs. *Annals of Applied Statistics* 2008; **2**(1):1034–1055, DOI: 10.1214/08-AOAS179.

[15] Egleston B, Scharfstein D, MacKenzie E. On estimation of the survivor average causal effect in observational studies when important confounders are missing due to death. *Biometrics* 2009; **65**(1):497–504, DOI: 10.1111/j.1541-0420.2008.01111.x.

[16] Small D, Rosenbaum P. War and wages: The strength of instrumental variables and their sensitivity to unobserved biases. *Journal of the American Statistical Association* 2008; **103**(483):924–933, DOI: 10.1198/016214507000001247.

[17] Angrist J, Imbens G, Rubin D. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 1996; **91**(434):444–455. URL `http://www.jstor.org/stable/2291629`.

[18] Rubin D. Comment on 'Randomization analysis of experimental data: The fisher randomization test' by D. Basu. *Journal of the American Statistical Association* 1980; **75**:591–593.

[19] Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983; **70**(1):41–55, DOI: 10.1093/biomet/70.1.41.

[20] McDonald C, Hiu S, Tierney W. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *MD Computing* 1992; **9**(1):304–312.

[21] Guo J, Geng Z. Collapsibility of logistic regression coefficients. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995; **57**(1):263–267. URL `http://www.jstor.org/stable/2346099`.

[22] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 1977; **39**(1):1–38, DOI: 10.1.1.133.4884.

[23] Rosenbaum P. *Observational Studies*. Springer: New York, 2002.

[24] Lagerros Y, Bellocco R, Adami HO, Nyren O. Measures of physical activity and their correlates: The swedish national march cohort. *European Journal of Epidemiology* 2009; **24**:161–169, DOI: 10.1007/s10654-009-9327-x.

[25] Schwartz S, Li F, Mealli F. A Bayesian semiparametric approach to intermediate variables in causal inference. *Journal of the American Statistical Association* 2011; **In press**.

[26] Shephard R, Balady G. Exercise as cardiovascular therapy. *Circulation* 1999; **99**:963–972.

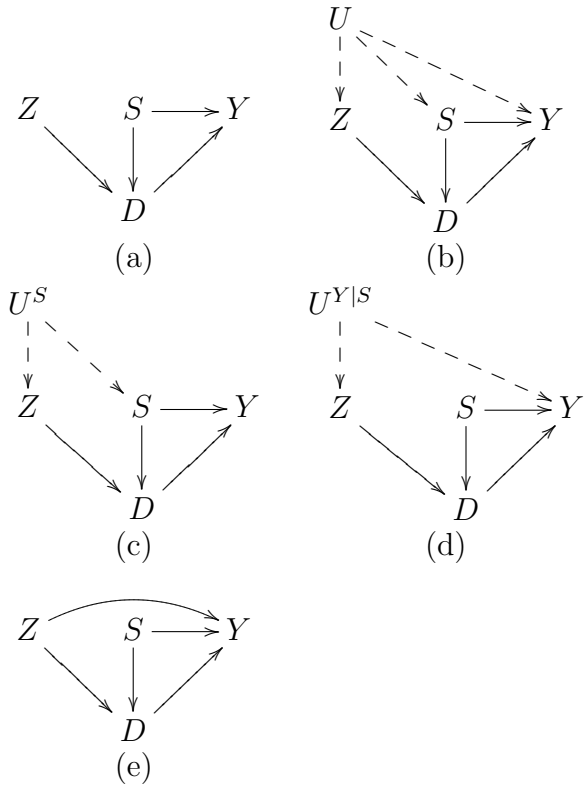

Figure 1: Directed acyclic graphs (DAGs) illustrating the relationsips among the variables in various PS scenarios. An arrow between two variables denotes that the initial variable influences the one it points to, with dashed lines indicating a relationship that is non-negligible but not observed. The relevant structures are: (a) no unmeasured confounding, (b) unmeasured confounding, (c) S-confounding, (d) Y-confounding, (e) no unmeasured confounding, but a direct effect of Z on Y, which is operationally indistinguishable from Y-confounding in (d).

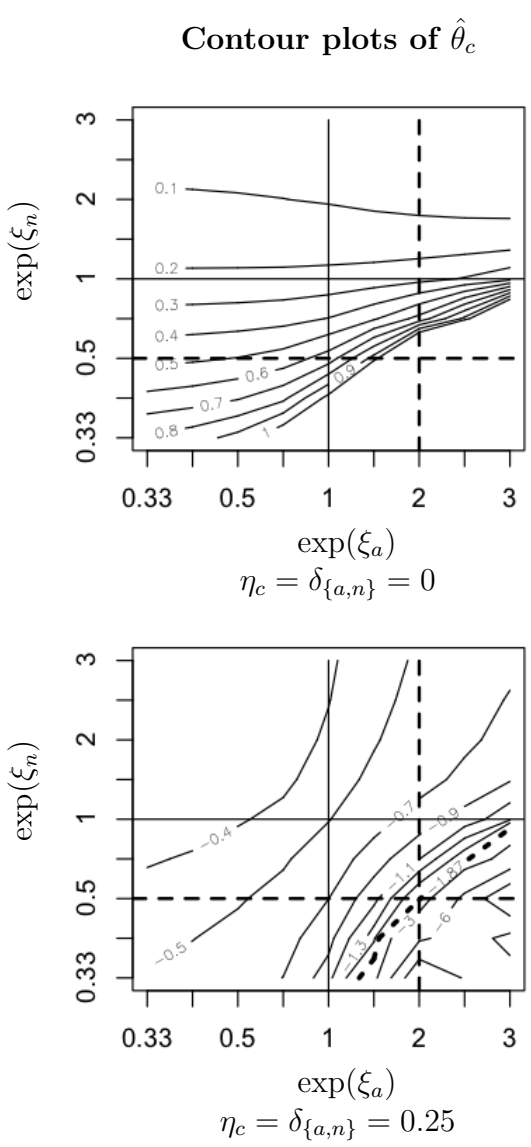

Figure 2: Illustrations of sensitivity contour plots for manipulated McDonald data. The top plot shows MLE contours for $\hat{\theta}_c$ across the possible combinations of ξ_s with $\eta_c = \delta_{\{a,n\}} = 0$. The bottom plot shows the same when $\eta_c = \delta_{\{a,n\}} = 0.25$, which are the true values. The dashed cross-hairs are at the approximately correct S-confounding sensitivity parameter values, $\exp(\xi_a) = 2$ and $\exp(\xi_n) = 1/2$. The dashed curve in the bottom plot indicates where $\hat{\theta}_c$ equals θ_c; this curve does not appear in the top plot because it is off the graph. The plots show that standard PS estimates of θ_c are biased in the presence of unmeasured confounding, and it is possible to recover the true θ_c when correct sensitivity parameters are used.

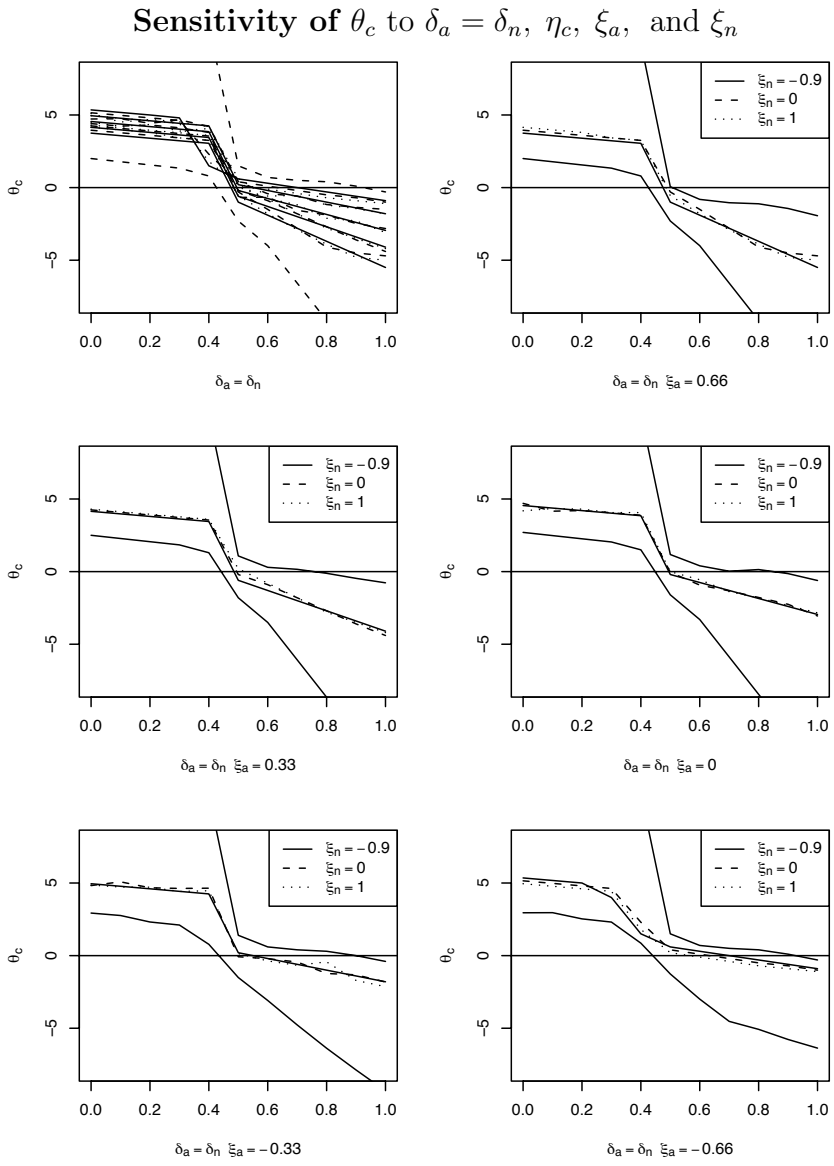

Figure 3: Estimates – and uncertainty assessment – of the sensitivity of the CACE in the NMC data to potential S-confounding and Y-confounding. In all panels, for each level of $\delta_a = \delta_n$, the highest and lowest lines trace the maximum and minimum endpoints of point-wise 95% credible intervals for all values of ξ_a and ξ_n examined in the plot. The top left panel shows the overall sensitivity to confounding for $\delta_a = \delta_n \in [0, 1]$ and $(\xi_a, \xi_n) \in [-0.66, 0.66] \times [-.90, 1.00]$. The remaining five panels decompose the top left panel into its five primary trajectories corresponding to $\xi_a = -.66, -.33, 0, .33, .66$ and re-plots sensitivity to ξ_n and $\delta_a = \delta_n$. For $\eta_c > 0$, the results in the figure are shifted down by η_c.

$p_{00} = .088$	$p_{10} = .112$	$p_{01} = .083$	$p_{11} = .069$
$\pi_{00} = .88$	$\pi_{10} = .12$	$\pi_{01} = .69$	$\pi_{11} = .31$

Table 1: Observed marginal proportions in the influenza study, i.e., ignoring covariates [20].

	$Z=0$	$Z=1$	$Z=0$	$Z=1$
No S-confounding and no Y-confounding				
$S=n$	$\pi_{n0}=.69$	$\pi_{n1}=.69$	$p_{n0}=.083$	$p_{n1}=.083$
$S=c$	$\pi_{c0}=.12$	$\pi_{c1}=.12$	$p_{c0}\approx.117$	$p_{c1}\approx.001$
$S=a$	$\pi_{a0}=.19$	$\pi_{a1}=.19$	$p_{a0}=.112$	$p_{a1}=.112$
S-confounding and no Y-confounding				
$S=n$	$\pi_{n0}\approx.56$	$\pi_{n1}=.69$	$p_{n0}=.083$	$p_{n1}=.083$
$S=c$	$\pi_{c0}\approx.25$	$\pi_{c1}\approx.21$	$p_{c0}\approx.099$	$p_{c1}\approx.046$
$S=a$	$\pi_{a0}=.19$	$\pi_{a1}\approx.10$	$p_{a0}=.112$	$p_{a1}=.112$
Y-confounding and no S-confounding				
$S=n$	$\pi_{n0}=.69$	$\pi_{n1}=.69$	$p_{n0}\approx.102$	$p_{n1}=.083$
$S=c$	$\pi_{c0}=.12$	$\pi_{c1}=.12$	$p_{c0}\approx.010$	$p_{c1}\approx.033$
$S=a$	$\pi_{a0}=.19$	$\pi_{a1}=.19$	$p_{a0}=.112$	$p_{a1}\approx.092$

Table 2: Example of population probabilities with various forms S-confounding and Y-confounding or lack thereof.