University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2015 *Paper* 341

The Statistics of Sensitivity Analyses

Alexander R. Luedtke* Ivan Diaz†
Mark J. van der Laan‡

*University of California, Berkeley, Division of Biostatistics, aluedtke@berkeley.edu
†Google, Inc., ildiazm84@gmail.com
‡University of California, Berkeley, Division of Biostatistics, laan@berkeley.edu

The Statistics of Sensitivity Analyses

Alexander R. Luedtke, Ivan Diaz, and Mark J. van der Laan

Abstract

Suppose one wishes to estimate a causal parameter given a sample of observations. This requires making unidentifiable assumptions about an underlying causal mechanism. Sensitivity analyses help investigators understand what impact violations of these assumptions could have on the causal conclusions drawn from a study, though themselves rely on untestable (but hopefully more interpretable) assumptions. Díaz and van der Laan (2013) advocate the use of a sequence (or continuum) of interpretable untestable assumptions of increasing plausibility for the sensitivity analysis so that experts can have informed opinions about which are true. In this work, we argue that using appropriate statistical procedures when conducting a sensitivity analysis is crucial to drawing valid conclusions about a causal question and understanding what assumptions one would need to make to do so. Conducting a sensitivity analysis typically relies on estimating features of the unknown observed data distribution, and thus naturally leads to statistical problems about which optimality results are already known. We present a general template for efficiently estimating the bounds on the causal parameter resulting from a given untestable assumption. The sequence of assumptions yields a sequence of confidence intervals which, given a suitable statistical procedure, attain proper coverage for the causal parameter if the corresponding assumption is true. We illustrate the pitfalls of an inappropriate statistical procedure with a toy example, and apply our approach to data from the Western Collaborative Group Study to show its utility in practice.

1 Introduction

Many scientific questions are causal rather than associational. Answering causal questions requires making certain unverifiable assumptions (Neyman, 1990; Rubin, 1974; Robins, 1986; Pearl, 2000). One can often attempt to answer a causal question by collecting a random sample of individuals from an observational or experimental study and estimating a statistical parameter which, under these unverifiable assumptions, is equal to the causal quantity of interest. These unverifiable assumptions are referred to as identifiability assumptions because they allow one to identify knowledge of the statistical parameter with knowledge of the causal parameter.

Identifiability assumptions can be implausible in many applications. Sensitivity analyses allow investigators to understand how their analysis would change under varying degrees of violations of these assumptions. Robins et al. (1999) describe a roadmap to evaluate the sensitivity to nonignorable missingness in missing data applications and unmeasured confounding in causal inference applications. In particular, they suggest specifying some semiparametric model for the missingness mechanism or counfounding relationship indexed by a sensitivity parameter, and reporting confidence intervals for the parameter of interest for each value of this sensitivity parameter. Similar approaches were considered in related articles (Rotnitzky et al., 1998; Scharfstein et al., 1999). If the chosen semiparametric form is correct, then such analyses often yield an interpretable sensitivity parameter. Nonetheless, these sensitivity parameters become hard to interpret when (as is typical) this semiparametric form is incorrect.

In response to these earlier approaches, Díaz and van der Laan (2013) advocate the use of sensitivity analyses which avoid the specification of difficult to interpret misspecified semiparametric models. They aim to make the assumptions behind the sensitivity analysis as transparent as possible so that their validity can be judged by non-statisticians. They illustrate the utility of such an approach by demonstrating the effectiveness of a Chagas disease treatment. Causal conclusions were previously thought impossible for this application due to the inevitable informative dropout resulting from the disease's long (30 year) incubation period. This long incubation period also

renders randomized studies prohibitively expensive, so that researchers can only use observational data to evaluate the efficacy of treatment.

VanderWeele and Arah (2011) give general formulas for the bias of the statistical parameter for the causal parameter if one assumes that adjusting for an unmeasured confounder would make the parameter identifiable. This approach yields an interpretable sensitivity parameter under sometimes strong simplifying assumptions, such as that this unmeasured confounder is binary. Obtaining confidence bounds for the causal parameter requires obtaining confidence bounds for a statistical parameter that adjusts for measured confounders, and thus requires modern statistical approaches for these confidence bounds to be valid. Similar approaches were used to get bounds for interaction parameters (VanderWeele et al., 2012) and for direct effects (VanderWeele, 2010). The bounds in Ding and VanderWeele (2015) are defined using this approach, but require far fewer parameters than the earlier methods when the unmeasured confounder is not binary.

These sensitivity analysis procedures are related to the partial identification literature which develops bounds on the difference between the causal and statistical parameter that hold under very weak assumptions (if any), such as bounds on the outcome (Manski, 1990; Horowitz and Manski, 2000; Manski, 2003; MacLehose et al., 2005). Though the bounds resulting from these analyses are convincing when informative, in many cases they can be too conservative to be informative about even the sign of an effect.

In this work we demonstrate the importance of using an appropriate statistical procedure when conducting a sensitivity analysis. In principle this should be obvious, but seems important to point out given the ubiquity of misspecified (semi)parametric working models in the field. The fact that one needs to use a modern statistical procedure when conducting a sensitivity analysis should not be a cause for concern–such methods already exist for many problems of interest, so one can often use pre-existing software packages to implement them.

Given a suitable statistical procedure, one has a valid sensitivity analysis provided the unverifiable assumption corresponding to the sensitivity parameter holds. For this reason we encourage statis-

ticians to make this assumption as interpretable as possible so that subject matter experts can have informed opinions about them.

Our presentation follows that of Díaz and van der Laan (2013), with our emphasis placed on the key role that statistics must play in any sensitivity analysis procedure. Section 2 outlines the objective of sensitivity analyses and shows how they can fail when not paired with an appropriate statistical procedure. Section 3 describes a statistically valid sensitivity analysis procedure. Section 3.2, which details the statistical estimation of bounds on the bias of the statistical parameter for the causal parameter, did not appear in Díaz and van der Laan (2013). Sections 2 and 3 alternate between a discussion that applies to general sensitivity analyses and a discussion of a running toy example. Section 4 details an application of our approach to estimate the additive effect of smoking on coronary heart disease in a real data example.

2 Why statistics is important in sensitivity analyses

Section 2.1 sets up the problem. Section 2.2 presents a sensitivity analysis procedure that yields a confidence interval for the causal parameter of interest provided one has valid confidence intervals for an *a priori* specified statistical parameter. Section 2.3 gives an example of how this procedure fails when one uses an inappropriate procedure to estimate this statistical parameter.

2.1 The problem

General discussion

Suppose one wishes to estimate a causal parameter ψ_{causal}. The observed data distribution is P_0 and, under causal assumptions which cannot be verified using data, we can identify a statistical parameter $\Psi(P_0)$ with the causal parameter ψ_{causal}, i.e. $\Psi(P_0) = \psi_{\text{causal}}$. We start by describing an analysis under these identifiability assumptions, and then discuss how sensitivity analyses can be used to understand what happens under violations of these assumptions.

For simplicity, suppose we observe n observations $O_1, ..., O_n$ drawn i.i.d. from some distribution

P_0. The results also apply to other sampling schemes with little modification. Generally one *attempts* to obtain an asymptotic confidence interval $\text{CI}_{\text{stat}} = [L_n, U_n]$ for $\Psi(P_0)$ with coveraging approaching, e.g., 95% as sample size grows. We focus on 95% coverage throughout this article, though the extension is obvious. The bounds L_n and U_n are data dependent, and generally

$$U_n - L_n \text{ converges to zero (in probability) as sample size grows.} \qquad (\star)$$

We make this assumption throughout. We have emphasized that the practitioner hopes that CI_{stat} has coverage approaching 95%, i.e.

$$\text{the random interval } \text{CI}_{\text{stat}} \text{ contains } \Psi(P_0) \text{ with probability approaching 0.95.} \qquad (\star\star)$$

Using an inappropriate statistical procedure can yield a CI for which ($\star\star$) does not hold, but we make the assumption ($\star\star$) until stated otherwise. Under our identifiability assumptions, $\psi_{\text{causal}} = \Psi(P_0)$ and the CI is also asymptotically valid for ψ_{causal}.

The purpose of sensitivity analyses is to explore what happens under violations of these identifiability assumptions, i.e. to see how dramatically an analysis would change if $\psi_{\text{causal}} \neq \Psi(P_0)$, given that some interpretable condition holds.

Example

We will use the potential outcomes framework when presenting our example, though the same result applies in other frameworks. Define the full data as (W, Y_0, Y_1), where Y_0 is the counterfactual outcome under no treatment, Y_1 is the counterfactual outcome under treatment, and W is a vector of covariates. We assume that all outcomes are bounded in $[0, 1]$. We observe an i.i.d. sample drawn from $(W, A, Y) \sim P_0$, where $Y = Y_A$ is the outcome of interest under the observed treatment A. We make the stable unit treatment value assignment assumption that the potential outcomes for one unit do not change with the treatment of other units and the positivity assumption

that $P_0(A = 1|W) > 0$ with probability 1 over draws of W.

Our causal parameter of interest is $\psi_{\text{causal}} = E[Y_1]$. Our statistical parameter of interest is the G-computation formula $\Psi(P_0) = EE[Y|A = 1, W]$ (Robins, 1986). If A is independent of Y_1 given W, then $\Psi(P_0) = \psi_{\text{causal}}$. In this example we are concerned with the situation where this independence may not hold.

We use a simple data generating distribution for our toy example. Suppose W is discrete and takes on four levels. Let the data be generated as follows, where U is some unobserved confounder:

$$\begin{aligned} U &\sim Bernoulli(0.9) \\ W|U &\sim Unif\{1, 2, 3, 4\} \\ A|U, W &\sim Bernoulli\left(0.75I(W = 1) + 0.98UI(W \neq 1)\right) \\ Y_1|A, U, W &\sim Bernoulli\left(0.8AUI(W \neq 1)\right), \end{aligned}$$

and we let $Y_0 = 0$ with probability 1. One can verify that $\psi_{\text{causal}} = E[Y_1] = 0.54$, while $\Psi(P_0) = EE[Y|A = 1, W] = 0.60$. Thus $\Psi(P_0)$ is not identified with ψ_{causal} due to the confounding by U.

We will return to this toy example throughout. A Monte Carlo simulation for this example can be found in Appendix B to allow readers to avoid the straightforward but tedious calculations needed to verify the numerical results related to this example.

2.2 Sensitivity analyses under condition ($\star\star$)

General discussion

Suppose that $\Psi(P_0)$ is not identified with ψ_{causal}. If no adjustment is made, CI_{stat} has coverage for ψ_{causal} approaching zero as sample size grows. Trivially, the amount that we need to shift our CI to obtain valid coverage for ψ_{causal} is exactly $\psi_{\text{causal}} - \Psi(P_0)$. This shift does not have to do with the

sample. Suppose we know that $\psi_{\text{causal}} - \Psi(P_0)$ belongs to some known interval $\text{CI}_{\text{diff}} = [L_{\text{diff}}, U_{\text{diff}}]$. In Section 3.2 we weaken this assumption and allow CI_{diff} to rely on P_0 and be estimated from the data. Given that $L_{\text{diff}} \leq \psi_{\text{causal}} - \Psi(P_0) \leq U_{\text{diff}}$, we know that

$$\text{CI}_{\text{causal}} = [L_n + L_{\text{diff}}, U_n + U_{\text{diff}}]$$

has asymptotic coverage of at least 0.95 for ψ_{causal}. These bounds are necessarily loose under $(\star)$ and $(\star\star)$. In particular, these two assumptions imply that L_n and U_n converge to $\Psi(P_0)$ in probability, so the non-identiability of ψ_{causal} implies that $L_{\text{diff}} \neq U_{\text{diff}}$: otherwise we could learn ψ_{causal} using data alone in the infinite sample limit. It follows that the width of $\text{CI}_{\text{causal}}$ will not shrink to zero with sample size whenever ψ_{causal} is unidentifiable.

In practice CI_{diff} may be too wide to be informative. In this case, one could use an interval $\text{CI}^{\delta}_{\text{diff}} = [L^{\delta}_{\text{diff}}, U^{\delta}_{\text{diff}}]$, indexed by some low dimensional parameter δ such that, for some *interpretable* Condition δ,

$$\psi_{\text{causal}} - \Psi(P_0) \in \text{CI}^{\delta}_{\text{diff}} \text{ under Condition } \delta.$$

This Condition δ should be interpretable in the entire nonparametric model, in contrast to earlier approaches which develop sensitivity parameters which tied to an untestable (semi)parametric model (see, e.g., Robins et al., 1999). Often one can take δ to be univariate so that $\delta = 1$ indicates no additional assumption, while $\delta = 0$ indicates that $\text{CI}^{\delta}_{\text{diff}} = [0, 0]$ and is thus unachievable given non-identifiability. One can interpret each $\text{CI}^{\delta}_{\text{causal}} = [L_n + L^{\delta}_{\text{diff}}, U_n + U^{\delta}_{\text{diff}}]$ as a valid confidence interval for ψ_{causal} provided Condition δ holds. Díaz and van der Laan (2013) stress that subject matter experts should be able to hypothesize about plausible values of δ.

Following this section, we generally omit δ from the notation for L_{diff}, U_{diff}, and CI_{diff}, with the understanding that these quantities are indexed by a sensitivity parameter δ.

Example

We wish to give an interval CI_{diff} that contains ψ_{causal} under Condition δ. Note that

$$\psi_{\text{causal}} = E\Big[E[Y_1|A=0,W]P_0(A=0|W) + E[Y|A=1,W]P_0(A=1|W)\Big]. \tag{1}$$

Suppose $|E[Y_1|A=0,W] - E[Y|A=1,W]| \leq \delta$ with probability 1 over draws of W. That is, the average potential outcomes under treatment do not differ by more than δ between the treated and untreated individuals in any strata of covariates. This holds trivially for $\delta = 1$ by the bounds on the outcome. Then (1) and some basic calculations yield that

$$-\delta P_0(A=0) \leq \psi_{\text{causal}} - \Psi(P_0) \leq \delta P_0(A=0). \tag{2}$$

The above bound is generally loose in the sense that for some (most) P_0 one cannot exhibit a causal distribution for which P_0 is an observed data distribution resulting from this causal distribution and $\psi_{\text{causal}} - \Psi(P_0)$ attains the prescribed upper or lower bound. For our toy example, $A = 0$ with low probability so this approach yields a reasonably tight bound. In particular, $P_0(A=0) \approx 0.15$, which yields upper and lower bounds on $\psi_{\text{causal}} - \Psi(P_0)$ of $\pm 0.15\delta$. Thus if a subject matter expert knew that the probability that $A = 0$ is no greater than, e.g., 0.2, then (2) yields an informative bound on $\psi_{\text{causal}} - \Psi(P_0)$, namely $\pm 0.2\delta$.

We can also get a tighter bound under an alternative Condition δ which uses relative risks rather than differences. Suppose that, with probability 1 over draws of the covariate W and some $\delta > 0$,

$$-\delta E[Y|A=1,W] \leq E[Y_1|A=0,W] - E[Y|A=1,W] \leq \delta\left(1 - E[Y|A=1,W]\right).$$

Consider the lower bound. Given that $E[Y|A=1,W] > 0$, the above says that $\frac{E[Y_1|A=0,W]}{E[Y|A=1,W]} \geq 1-\delta$ for some $\delta > 0$, i.e. that the relative risk of $Y_1 = 1$ among untreated versus treated people is at

least $1-\delta$ in every strata of covariates. Similarly, the above says that $\frac{E[1-Y_1|A=0,W]}{E[1-Y|A=1,W]} \geq 1-\delta$, which is a similar relative risk bound but now for the risk of $Y_1 = 0$. At $\delta = 1$ the above condition reproduces the known bounds on the outcome. For rare (or highly common) outcomes it is worth using a two-dimensional δ parameter in the above expression, one for the lower bound and one for the upper, but for simplicity we do not explore this here. Plugging the above bound into (1) yields that

$$\begin{aligned} &-\delta E\left[E[Y|A=1,W]P_0(A=0|W)\right] \\ &\quad \leq \psi_{\text{causal}} - \Psi(P_0) \leq \delta E\left[E[1-Y|A=1,W]P_0(A=0|W)\right]. \end{aligned} \tag{3}$$

For a given δ, the bound in (3) will always be half the width of the bound in (2). In our toy example, the lower and upper bounds are approximately -0.07δ and 0.08δ. Nonetheless, this tighter bound relies on P_0 through more than just the marginal censoring mechanism, so estimating the bounds will be more difficult than for (2). In Section 3.2 we discuss how one can estimate these upper and lower bounds and account for this when developing confidence intervals.

Manski (2003) considers bounds on causal parameters which reduce to using the known bounds on the outcome in great detail. That work also gives bounds under other assumptions, such as monotonicity assumptions, which may help inspire other choices of Condition δ.

2.3 Sensitivity analyses when ($\star\star$) does not hold because the estimator is inconsistent

General discussion

We have glossed over another issue which, unlike the sensitivity analysis, can be mitigated by using a proper statistical procedure and gathering more data. We have assumed that we have an asymptotically valid confidence interval CI_{stat} satisfying ($\star$) and ($\star\star$), which immediately implies $L_n \to \Psi(P_0)$ and $U_n \to \Psi(P_0)$. Assuming a misspecified parametric or semiparametric model, for instance, will often yield a point estimate which is consistent for the wrong quantity, namely

$\Psi_1(P_0) \neq \Psi(P_0)$. As the estimate is typically taken to be the center of a CI (or at least contained within the CI), the investigator will report a CI_{stat} for which both the lower and upper bound converge to $\Psi_1(P_0)$ in probability. Thus CI_{stat} will have coverage approaching zero for our statistical parameter $\Psi(P_0)$.

The following decomposition will help to clarify the discussion:

$$\hat{\psi} - \psi_{\text{causal}} = \hat{\psi} - \Psi_1(P_0) + \underbrace{\Psi_1(P_0) - \Psi(P_0)}_{\text{statistical bias}} + \underbrace{\Psi(P_0) - \psi_{\text{causal}}}_{\text{confounding bias}}.$$

We are referring to *zeroth-order bias* when we say "statistical bias", which can (often) be thought the limit of the bias of $\hat{\psi}$ for $\Psi(P_0)$ as sample size grows. Our sensitivity analysis gives us a bound on the confounding bias, telling us that (we are at least confident that) the confounding bias is bounded between L_{diff} and U_{diff}. However, our statistical bias may render $\text{CI}_{\text{causal}}$ invalid. Alternatively, it may just so happen that our bounds on the confounding bias actually bound the sum of the statistical and confounding biases, i.e. $\psi_{\text{causal}} - \Psi_1(P_0) \in \text{CI}_{\text{diff}}$. In this case $\text{CI}_{\text{causal}}$ will be valid provided ($\star\star$) holds, there is no reason to assume that this is the case. We have worked hard to ensure that we believe that $\psi_{\text{causal}} - \Psi(P_0)$ falls within the interval of interest under an interpetable condition. It is often hard to understand what $\Psi_1(P_0)$ even is, i.e. what feature of our observed population of interest Ψ_1 returns. It therefore seems unlikely that we can feel confident that $\psi_{\text{causal}} - \Psi_1(P_0) \in \text{CI}_{\text{diff}}$. The same problem holds if CI_{diff} had been derived to bound $\psi_{\text{causal}} - \Psi_1(P_0)$ from the outset: statisticians cannot communicate this quantity to one another, let alone to subject matter experts.

Example

Suppose one uses maximum likelihood estimation among individuals with $A = 1$ to estimate the following misspecified model for $E[Y|A = 1, W]$:

$$m_\beta(W) = \beta_{134} I(W \neq 2) + \beta_2 I(W = 2).$$

The above model is misspecified because in truth $E[Y|A=1,W] = 0.8I(W \neq 1)$. The infinite sample limits of the maximum likelihood estimators are $(\beta^*_{134}, \beta^*_2) \approx (0.56, 0.80)$. Suppose one also incorrectly estimates $P_0(A=1|W)$ as

$$g_\eta(W) = \eta_{12} I(W \leq 2) + \eta_{34} I(W \geq 3).$$

The above model is misspecified because in truth $P_0(A=1|W) = 0.75 + 0.132I(W \neq 1)$. The infinite sample limits of the maximum likelihood estimators are $(\eta^*_{12}, \eta^*_{34}) \approx (0.82, 0.88)$.

Because the outcome regression and propensity score are misspecified in this example, many standard estimators will be inconsistent for $\Psi(P_0)$: G-estimation and the inverse probability weighted estimator have limits of approximately 0.62, while the double robust one-step estimator has limit approximately 0.61. Targeted minimum loss-based estimators (TMLEs) represent another class of doubly robust estimators. Because both the propensity score and outcome regression are inconsistent in this example, its limit will depend on the fluctuation submodel used to implement the TMLE. Nonetheless, the TMLE will be inconsistent in this example just as is the double robust one-step estimator.

For all of these estimators, $\psi_{\text{causal}} - \Psi_1(P_0) \notin \text{CI}^\delta_{\text{diff}}$ when $\delta = 1$ and $\text{CI}^\delta_{\text{diff}}$ is known. We note that $\psi_{\text{causal}} - \Psi_1(P_0)$ nearly falls in $\text{CI}^\delta_{\text{diff}}$ for $\delta = 1$ when one uses a double robust estimator in this example. Given that we have misspecified the outcome and treatment mechanism, it is perhaps inappropriate to consider coverage of the true $\text{CI}^\delta_{\text{diff}}$ rather than the estimated $\text{CI}^\delta_{\text{diff}}$ as discussed in Section 3.2. Nonetheless, this destroys the interpretability of Condition δ, as Condition δ is interpreted with respect to parameters of the true observed data distribution P_0.

Clearly it would not be difficult in a large sample to correctly estimate the treatment mechanism and outcome regression correctly for a covariate with only four levels. Though contrived, we have presented these estimators to give what we view as one of the simplest examples of how inconsistent estimation of $\Psi(P_0)$ can invalidate an otherwise solid sensitivity analysis. In a general setting, using parametric models to estimate the needed components of the likelihood will often

yield oversmoothed estimators which converge to the wrong limit as sample size grows. If W were high-dimensional, then one might also be overly ambitious in estimating the needed components of the likelihood and obtain undersmoothed estimates with large bias that shrinks slowly as sample size grows. In this case the estimators of $\Psi(P_0)$ will often be consistent but ($\star\star$) will fail to hold because the confidence interval around the estimate of $\Psi(P_0)$ is generally constructed under the assumption that the needed components of the likelihood converge to their limits sufficiently quickly.

3 Conducting a statistically valid sensitivity analysis

In this section we outline how one can overcome the statistical obstacles presented in the previous section. Section 3.1 shows that the results of Section 2.3 do not preclude interpretable sensitivity analyses by exhibiting a large class of estimators which yield valid confidence intervals for the statistical parameter. Section 3.2 shows how we can perform a sensitivity analysis even when the confidence interval exhibited in Section 2.2 is unknown because it depends on features of the observed data distribution.

3.1 Needing to satisfy ($\star\star$) is reasonable

General discussion

In most problems, there are many estimators which are consistent. One such class of estimators is the class of double robust methods such as TMLE and estimating equation methodology can can produce estimators that are consistent for $\Psi(P_0)$ in many problems. Of course, consistency alone does not guarantee ($\star\star$). One also needs to be able to construct a valid confidence interval for the target of interest, which typically relies on the estimator's the bias being small relative to the variance. Fortunately, these double robust methods often also satisfy this property by being so-called asymptotically linear estimators with the property that, for a mean zero, finite variance

function $IC_{\hat{\psi}}(O_i)$ known as the influence curve,

$$\sqrt{n}\left[\hat{\psi} - \Psi(P_0)\right] \approx \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IC_{\hat{\psi}}(O_i) \text{ for large sample sizes.} \tag{4}$$

The approximation becomes arbitrarily precise (in probability) as the sample grows. The dependence of $IC_{\hat{\psi}}$ on P_0 is omitted in the notation. In causal inference applications, influence curves typically rely on P_0 through outcome regressions and the treatment mechanism. The form of the influence curve will vary depending on the parameter of interest. Given (4), the variance of the influence curve will generally be estimable from the data. Because the influence curve is mean zero, CIs for $\Psi(P_0)$ which satisfy ($\star\star$) are given by

$$\text{CI}_{\text{stat}} = \left[\hat{\psi} - 1.96\frac{\hat{\sigma}}{\sqrt{n}}, \hat{\psi} + 1.96\frac{\hat{\sigma}}{\sqrt{n}}\right],$$

where $\hat{\sigma}^2$ is an estimate of the variance of $IC_{\hat{\psi}}(O)$.

Asymptotic linearity will typically only hold when one can estimate the parts of P_0 used to define the influence curve $IC_{\hat{\psi}}$. Developing an estimator which satisfies (4) generally requires estimating one or several outcome regression(s) and the treatment mechanism well. By "well", we mean that these estimators should be estimated consistently with, e.g., mean squared error that shrinks sufficiently quickly as sample size grows. Due to the required consisistency, parametric working models are generally insufficient for estimating these parts of the likelihood. Instead, one can use machine learning methods or ensemble learners that combine multiple candidate estimators to estimate these quantities if asymptotic linearity is desired.

There is a rich literature on the development of asymptotically linear estimators for causal parameters. See van der Laan and Robins (2003) and van der Laan and Rose (2011) for an overview. Sophisticated machine learning and ensemble algorithms have also been developed and integrated into the estimation of causal inference parameters. This allows one to avoid committing to incorrectly specified parametric working models, thereby making (4) more plausible. Using machine

learning algorithms to estimate the nuisance functions is becoming more popular: Setoguchi et al. (2008); Lee et al. (2010) suggest using various machine learning approaches to estimate propensity scores, and van der Laan and Dudoit (2003); van der Laan and Rubin (2006); van der Laan and Rose (2011) suggest using a cross-validation based ensemble algorithm to estimate both the outcome regression and propensity score. In observational studies, estimating both the propensity score and the outcome regression using data adaptive approaches, and then subsequently reducing bias using estimating equations or TMLE, is essential to make the error term in (4) negligible. Using estimators which can optimally select among a collection of candidate estimators (in terms of, e.g., mean squared error) is especially appealing because often this yields tight control of the remainder term in (4). The super-learning methodology satisfies this optimality criteria (van der Laan et al., 2007; Polley and van der Laan, 2010).

We have shown that, under (4), one can develop a CI for ψ_{causal}. More generally, such a CI is valid whenever the CI_{stat} gives valid coverage for $\Psi(P_0)$.

Example

Double robust estimating equation and targeted minimum loss based estimators for $\Psi(P_0)$ have been presented in van der Laan and Robins (2003) and van der Laan and Rose (2011), respectively. When both the outcome regression and treatment mechanism are estimated consistently and at a fast enough rate, these estimators have influence curve

$$IC_{\hat{\psi}}(O) = \frac{A}{P_0(A=1|W)}\left(Y - E[Y|A=1,W]\right) + E[Y|A=1,W] - \Psi(P_0).$$

The assumption that both the outcome regression and treatment mechanism are estimated well enough can be unpleasant in practice. A recent work presents an estimator which is asymptotically linear with valid inference when only one of these objects is estimated well (van der Laan, 2014). This approach can be integrated into a sensitivity analysis, but we omit such discussion here.

3.2 Bounds on the causal bias unknown

General discussion

Suppose that CI_{diff} relies on the observed data distribution P_0. This holds in our example, see (2) and (3). In this section we make this dependence explicit by writing $L_{\text{diff}}(P_0)$ and $U_{\text{diff}}(P_0)$. Note that L_{diff} and U_{diff} are now parameters which take as input an observed data distribution and output a number. Thus, provided L_{diff} is sufficiently smooth in the sense discussed in Section 1.4.1 of van der Laan and Robins (2003), it is reasonable to expect that, for an influence curve $IC_{\hat{\ell}}$, we can develop an asymptotically linear estimators of L_{diff} with the property that

$$\sqrt{n}\left[\hat{\ell}_{\text{causal}} - L_{\text{diff}}(P_0)\right] \approx \frac{1}{\sqrt{n}}\sum_{i=1}^{n} IC_{\hat{\ell}}(O_i) \text{ for large sample sizes.}$$

If U_{diff} is sufficiently smooth, then we would expect to be able to develop an estimator so that the same two expressions above hold with $\hat{\ell}_{\text{causal}}$, $L_{\text{diff}}(P_0)$, and $IC_{\hat{\ell}}$ replaced by $\hat{u}_{\text{causal}}$, $U_{\text{diff}}(P_0)$, and $IC_{\hat{u}}$, respectively. Combining the above and (4) yields

$$\sqrt{n}\left[\hat{\psi} + \hat{\ell}_{\text{causal}} - (\Psi(P_0) + L_{\text{diff}}(P_0))\right] \approx \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[IC_{\hat{\psi}}(O_i) + IC_{\hat{\ell}}(O_i)\right].$$

The right hand side converges to a mean zero normal distribution with variance equal to the variance of the difference of influence curves on the right. An analogous argument yields that $\Psi(P_0) + U_{\text{diff}}(P_0)$ has influence curve $IC_{\hat{\psi}} + IC_{\hat{u}}$. Hence the joint distribution of these two influence curves applied to $O \sim P_0$ converges to a multivariate normal distribution with mean zero and covariance matrix given by that of the two-dimensional random variable with coordinates $IC_{\hat{\psi}}(O) + IC_{\hat{\ell}}(O)$ and $IC_{\hat{\psi}}(O) + IC_{\hat{u}}(O)$. Given a consistent estimate $\hat{\Sigma}$ of the covariance matrix, one can take Monte Carlo draws $(Z_L^1, Z_U^1), ..., (Z_L^m, Z_U^m)$ from the $N(0, \hat{\Sigma})$ distribution. Given these draws, one can then choose $\hat{s}$ to be the 95% quantile of $\max\{Z_L^k, -Z_U^k\}$ among the observa-

tions $k \in \{1, ..., m\}$. In that case, the confidence interval

$$\left[\hat{\psi} + \hat{\ell}_{\text{causal}} - \frac{\hat{s}}{\sqrt{n}}, \hat{\psi} + \hat{u}_{\text{causal}} + \frac{\hat{s}}{\sqrt{n}}\right] \tag{5}$$

will contain the causal parameter with probability approaching 0.95 under Condition δ. One could alternatively replace $\hat{s}$ in the lower bound with a $\hat{s}_L$ and $\hat{s}$ in the upper bound with $\hat{s}_U$ and choose an empirically valid confidence interval (in the Monte Carlo draws) which minimizes $\hat{s}_U + \hat{s}_L$. This may be beneficial when, e.g., the lower bound is significantly easier to estimate than the upper bound and one wants to make the confidence interval lower bound tighter to reflect this.

Horowitz and Manski (2000) used a similar approach to get a confidence region for the bounds in the partial identifiability context. Woutersen (2006) considered how to develop such a confidence region given a asymptotically linear estimators of the upper and lower bound in partial identifiabilty problems. Both of these works actually consider a refined procedure which guarantees coverage for the parameter ψ_{causal}, rather than the entire region $[\Psi(P_0) + L_{\text{diff}}, \Psi(P_0) + U_{\text{diff}}]$ known to contain ψ_{causal}, with probability approaching 0.95. Such refinements will be analogous for sensitivity analyses, but we do not explore them here.

Example

First consider the looser bound in (2). Using that $P_0(A = 0) = E[I(A = 0)]$, we find that an asymptotically linear estimator of the upper bound when $\delta = 1$ is given by $\hat{u}_{\text{causal}} = \frac{1}{n}\sum_{i=1}^{n} I(A_i = 0)$, and an asymptotically linear estimator for the lower bound when $\delta = 1$ is given by $\hat{\ell}_{\text{causal}} = -\hat{u}_{\text{causal}}$. This yields $IC_{\hat{u}}(O) = I(A = 0) - P_0(A = 0)$ and $IC_{\hat{\ell}}(O) = -IC_{\hat{u}}(O)$. For sensitivity parameter values of $\delta < 1$, these estimates and influence curves can simply be scaled by multiplying by δ.

Now consider the tighter bound in (3). We can estimate $P_0(A = 0)$ as we did for (2), but it remains to estimate $E[E[Y|A = 1, W]P_0(A = 0|W)]$. Conveniently, this quantity is a key component of the effect of treatment among the untreated. Similar approaches to those used for the effect

of treatment among the treated can be used to develop an asymptotically linear estimator for this quantity (see Chapter 8 of van der Laan and Rose, 2011).

4 Real data example

We now apply our method to data from the Western Collaborative Group Study (WCGS) (Rosenman et al., 1964, 1975). The WCGS was a prospective cohort study designed to learn about the effect of binary personality type (Type A or B) on coronary heart disease (CHD) within an eight and a half year period. The data is publicly available through the `epitools` package in `R` (Aragon, 2012). Here we will not consider the effect of personality type on CHD due to the difficulties of temporality among the baseline random variables and interpretation of an intervention on personality type. We will instead focus on the effect of smoking status on CHD. This example is particularly useful for testing a sensitivity analysis method because the causal link between smoking and CHD is well-established in the literature. A similar decision was made by Ding and VanderWeele (2015) when evaluating their sensitivity analysis procedure on a historical data set exploring the effect of smoking on lung cancer.

We use the variable definitions given in Table 1, which includes the confounders for which we adjust. Our objective is to estimate the average effect of smoking on CHD events within 8 1/2 years, i.e. $\psi_{\text{causal}} = E[Y_1 - Y_0]$. We do not adjust for baseline blood pressure or cholesterol to avoid potential reverse causality problems. Some readers may be concerned with the interpretation of setting smoking status to one since we have not specified the number of cigarettes smoked per day. In this case we define our intervention as one in which we only intervene on smoking status, and then allow smokers to decide how many cigarettes to smoke per day. Specifically, this corresponds to letting individuals randomly choose the number of cigarettes smoked per day according to the frequency of cigarettes smoked per day among observed smokers in their strata of covariates (van der Laan et al., 2005).

Our objective is to get inference for the additive treatment effect of smoking status on CHD, i.e.

Notation	Measurement
W	Baseline: age, height, weight, indicator of Type A personality
A	Baseline smoking status (indicator of smoking at least one cigarette per day)
Y	Coronary heart disease event within 8 1/2 years

Table 1: Definition of measured potential confounders, treatment, and outcome in the WCGS. Baseline covariates (W) and treatment (A) were measured at baseline, and the outcome is an indicator of a CHD event within an 8 1/2 year period.

$\psi_{\text{causal}} = E[Y_1 - Y_0]$. To do this, we will use the G-computation formula as described in (Robins, 1986). That is, we will attempt to estimate the statistical parameter $\Psi(P_0) = EE[Y|A = 1, W] - EE[Y|A = 0, W]$. Under nearly identical conditions described for the toy example in Section 2.1, $\Psi(P_0)$ is identified with the causal average treatment effect. The only additional assumption is that we require both Y_0 and Y_1 to be independent of A given W, where previously we only required this independence for Y_1.

We use a bivariate sensitivity parameter $\delta = (\delta^-, \delta^+)$, where δ^- and δ^+ fall in $[0, 1]$. Condition δ is satisfied when the following two inequalities hold with probability 1 over draws of the covariate W:

$$-\delta^- E[Y|A = 0, W] \leq E[Y_0|A = 1, W] - E[Y|A = 0, W] \leq \delta^+ E[1 - Y|A = 0, W]$$
$$-\delta^- E[Y|A = 1, W] \leq E[Y_1|A = 0, W] - E[Y|A = 1, W] \leq \delta^+ E[1 - Y|A = 1, W].$$

A straightforward extension of (3) shows that, under Condition δ,

$$L_{\text{diff}}(P_0) = -E\Big[\delta^- E[Y|A = 1, W]P_0(A = 0|W) + \delta^+ E[1 - Y|A = 0, W]P_0(A = 1|W)\Big]$$
$$U_{\text{diff}}(P_0) = E\Big[\delta^+ E[1 - Y|A = 1, W]P_0(A = 0|W) + \delta^- E[Y|A = 0, W]P_0(A = 1|W)\Big]. \quad (6)$$

Below we refer to individuals who smoke in the observed population as "natural smokers" and people who do not smoke in the observed population as "natural nonsmokers". We also refer to

the risk ratio as the RR. Condition δ implies the following four inequalities within each stratum of the covariates:

I1) $\frac{E[Y_1|A=0,W]}{E[Y|A=1,W]} \geq 1-\delta^-$: The RR for a natural nonsmoker versus a natural smoker having a CHD event if, contrary to fact, everyone is intervened upon to be a smoker is at most $1-\delta^-$.

I2) $\frac{E[Y_0|A=1,W]}{E[Y|A=0,W]} \geq 1-\delta^-$: The RR for a natural smoker versus a natural nonsmoker having a CHD event if, contrary to fact, everyone is intervened upon not to smoke is at most $1-\delta^-$.

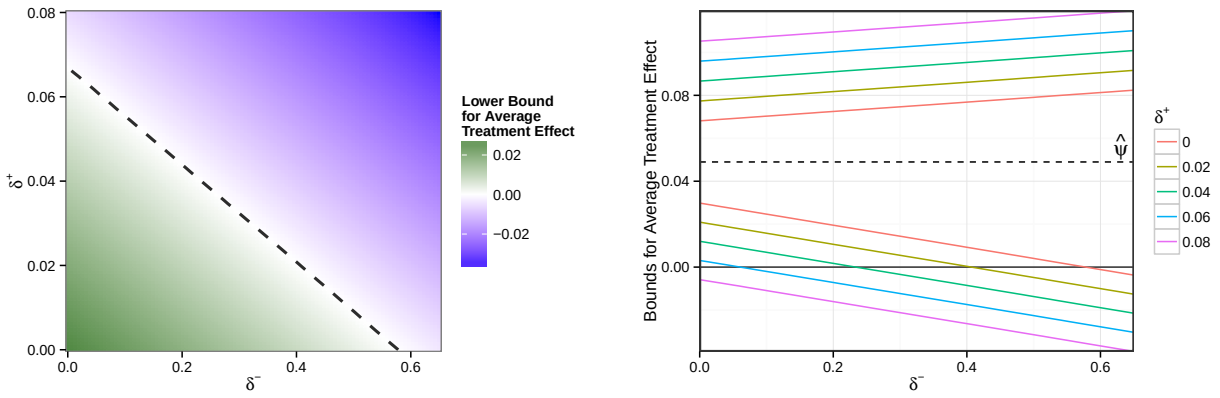
I3) $\frac{E[1-Y_1|A=0,W]}{E[1-Y|A=1,W]} \geq 1-\delta^+$: The RR for a natural nonsmoker versus a natural smoker *not* having a CHD event if, contrary to fact, everyone is intervened upon to be a smoker is at most $1-\delta^+$.

I4) $\frac{E[1-Y_0|A=1,W]}{E[1-Y|A=0,W]} \geq 1-\delta^+$: The RR for a natural smoker versus a natural nonsmoker *not* having a CHD event if, contrary to fact, everyone is intervened upon not to smoke is at most $1-\delta^+$.

The validity of the lower bound $L_{\text{diff}}(P_0)$ only relies on I1) and I4). Under I1), natural nonsmokers are not too protected from CHD events by some unmeasured cause. Under I4), smokers are not too inclined towards CHD events by some unmeasured cause. The prevalence of coronary events is low in our data set ($257/3154 \approx 0.08$), so large values of δ^- should be more plausible than large values of δ^+.

We first estimate the G-computation estimand given by $EE[Y|A=1,W] - EE[Y|A=0,W]$ using the TMLE as presented in Chapter 7 of van der Laan and Rose (2011). We then estimate the interval $\text{CI}_{\text{causal}}$ resulting from (6) using a TMLE algorithm described in Appendix A. Given the broader focus of this paper, we omit a theoretical analysis of the asymptotic properties of this estimator, though refer the reader to van der Laan and Rose (2011) for a general template for how to analyze such an estimator. We use 2.5×10^4 draws from a bivariate normal distribution to implement the confidence bound estimation procedure described in Section 3.2.

Interested readers can reproduce our results using the code in Appendix C. Code to replicate our analysis and generate the figures can be found in Appendix C.



(a) (b)

Figure 1: 95% confidence bounds for the average treatment effect of smoking on CHD events: (a) lower bound at a continuum of δ values, (b) confidence bounds at several δ values.

Figure 1a shows how the lower bound is impacted by different choices of δ. Consider $\delta^+ = 0.02$, and suppose that the probability of not having a CHD event is at most 0.3 within all strata of covariates (according to our estimate of $E[1 - Y|A = 0, W]$, the maximum probability of not having a CHD event is approximately 0.25). In this case I4) is satisfied provided $E[Y_0|A = 1, W] \leq E[Y|A = 1, W] + 0.015$, so that within any stratum of covariates natural smokers could have an at most a 1.5% higher additive heart attack risk than natural nonsmokers if an intervention had set everyone to be nonsmokers at baseline. For the lower bound on the average treamtent effect to remain positive, we then need that δ^- is no more than approximately 0.4. Inequality I2) is irrelevant for the lower bound on the average treamtent effect, so we focus on I1). This says that if we intervened in the population to make everyone a smoker then, within each stratum of covariates, the relative risk of a heart attack between natural nonsmokers and natural smokers is no less than 60%.

Figure 1b provides similar insights for the lower bounds, but also allows one to visualize the upper bound for the average treatment effect under different choices of δ. From these choices of sensitivity parameters, it appears unlikely that the average treatment effect of smoking on CHD events within eight and a half years will be larger than 0.12 in this population.

5 Discussion

Sensitivity analyses shed further light on the plausibility of a causal question. We have followed the approach of Díaz and van der Laan (2013) in advocating the use of an easily interpretable sensitivity analysis parameter that allows subject matter experts to have informed opinions on what values are plausible. If one knew the underlying observed data distribution then one could immediately report a bound on the causal parameter of interest given a value of the sensitivity parameter. Of course one never exactly knows the observed data distribution.

Here enters statistics. We have showed that using an inconsistent statistical procedure can invalidate a sensitivity analysis procedure. We outlined powerful nonparametric approaches which typically allow one to avoid this issue. We also showed how to incorporate these approaches to estimate the bounds that one has on the causal bias given an interpretable but unverifiable assumption. Though our statistical discussion relied on asymptotic arguments, these methods have been shown to perform well in a wide variety of situations. The arguments change very little if one can instead rely on finite sample concentration inequalities to bound the statistical quantity of interest, though such approaches tend to be conservative.

In certain situations sensitivity analyses may fail to yield informative bounds on a causal quantity. There are several ways to deal with this problem. If logistically and ethically feasible, one could gather further covariates on the individuals in the study so that the needed identifiability assumptions are more plausible. Otherwise, one can always interpret the statistical parameter without any reference to causation. Whether or not this parameter is of interest depends on the problem.

The objective of a sensitivity analysis is to better understand the gap between a causal parameter and an unknown but estimable statistical parameter. If one has a bad estimate of the statistical parameter then this endeavor is futile. Thus we advocate for the implementation of cutting edge statistical approaches when conducting sensitivity analyses.

Acknowledgement

This research was supported by NIH grant R01 AI074345-06. Alex Luedtke was supported by the NDSEG Fellowship Program of the U.S. Department of Defense.

References

T J Aragon. *epitools: Epidemiology Tools*, 2012. URL http://cran.r-project.org/package=epitools.

I Díaz and M J van der Laan. Sensitivity Analysis for Causal Inference Under Unmeasured Confounding and Measurement Error Problems. *Int. J. Biostat.*, 9(2):149–160, 2013.

P Ding and T VanderWeele. Sensitivity Analysis Without Assumptions. *in Press at Epidemiol.*, arXiv prep, 2015.

J L Horowitz and C F Manski. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *J. Am. Stat. Assoc.*, 95(449):77–84, 2000.

B K Lee, J Lessler, and E A Stuart. Improving propensity score weighting using machine learning. *Stat. Med.*, 29:337–346, 2010. ISSN 02776715. doi: 10.1002/sim.3782.

R F MacLehose, S Kaufman, J S Kaufman, and C Poole. Bounding causal effects under uncontrolled confounding using counterfactuals. *Epidemiology*, 16(4):548–555, 2005.

C F Manski. Nonparametric bounds on treatment effects. *Am. Econ. Rev.*, pages 319–323, 1990.

C F Manski. *Partial identification of probability distributions*. Springer, Berlin Heidelberg New York, 2003.

J Neyman. On the application of probability theory to agricultural experiments. *Stat. Sci.*, 5: 465–480, 1990.

J. Pearl. Causality: Models, Reasoning, and Inference. *New York Cambridge*, 2000. URL http://adsabs.harvard.edu/abs/2000caus.book.....P.

J Pfanzagl. *Estimation in semiparametric models*. Springer, Berlin Heidelberg New York, 1990.

E C Polley and M J van der Laan. Super Learner in Prediction. Technical Report 266, Division of Biostatistics, University of California, Berkeley, 2010.

Eric Polley and Mark van der Laan. *SuperLearner: super learner prediction*, 2013. URL `http://cran.r-project.org/package=SuperLearner`.

J M Robins. A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. *Math. Model.*, 7(9):1393–1512, 1986.

J M Robins, A Rotnitzky, and D O Scharfstein. Sensitivity Analysis for Selection Bias and Unmeasured Confounding in Missing Data and Causal Inference Models. In M E Halloran and D Berry, editors, *Stat. Model. Epidemiol. Environ. Clin. Trials*, IMA Volumes in Mathematics and Its Applications. Springer, 1999.

R H Rosenman, M Friedman, R Straus, M Wurm, R Kositchek, W Hahn, and N T Werthessen. A predictive study of coronary heart disease: The Western Collaborative Group Study. *J Am Med Assoc*, 189(1):15–22, 1964.

R H Rosenman, R J Brand, C D Jenkins, M Friedman, R Straus, and M Wurm. Coronary heart disease in the Western Collaborative Group Study: Final follow-up experience of 8 1/2 years. *J Am Med Assoc*, 233(8):872–877, 1975.

A Rotnitzky, J M Robins, and D O Scharfstein. Semiparametric Regression for Repeated Outcomes with Nonignorable Nonresponse. *J. Am. Stat. Assoc.*, 93(444):1321–1339, 1998.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*, 66:688–701, 1974.

D O Scharfstein, A Rotnitzky, and J M Robins. Adjusting for Nonignorable Drop-out Using Semiparametric Nonresponse Models, (with Discussion and Rejoinder). *J. Am. Stat. Assoc.*, 94 (448):1096–1120 (1121–1146), 1999.

S Setoguchi, S Schneeweiss, M A Brookhart, R J Glynn, and E F Cook. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiol. Drug Saf.*, 17(6):546, 2008.

M J van der Laan. Targeted estimation of nuisance parameters to obtain valid statistical inference. *Int. J. Biostat.*, 10(1):29–57, 2014.

M J van der Laan and S Dudoit. Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: finite sample oracle inequalities and examples. Technical Report 130, Division of Biostatistics, University of California, Berkeley, 2003.

M J van der Laan and J M Robins. *Unified methods for censored longitudinal data and causality*. Springer, New York Berlin Heidelberg, 2003.

M J van der Laan and S Rose. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer, New York, New York, 2011.

M J van der Laan and Daniel B Rubin. Targeted maximum likelihood learning. *Int. J. Biostat.*, 2 (1):Article 11, 2006.

M J van der Laan, T J Haight, and I B Tager. van der Laan et al. Respond to Hypothetical Interventions to Define Causal Effects. *Am. J. Epidemiol.*, 162(7):621–622, 2005.

M J van der Laan, E Polley, and A Hubbard. Super Learner. *Stat Appl Genet Mol*, 6(1):Article 25, 2007. ISSN 1.

T J VanderWeele. Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology*, 21(4):540, 2010.

T J VanderWeele and O A Arah. Unmeasured confounding for general outcomes, treatments, and confounders: Bias formulas for sensitivity analysis. *Epidemiology*, 22(1):42, 2011.

T J VanderWeele, B Mukherjee, and J Chen. Sensitivity analysis for interactions under unmeasured confounding. *Stat. Med.*, 31(22):2552–2564, 2012.

T Woutersen. A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters. *Tech. Report, Johns Hopkins Univ.*, 2006.

Appendices

A TMLE for causal bounds for the average treatment effect

A.1 Estimator

Suppose we have obtained an estimate $\hat{\psi}$ of $\Psi(P_0) = EE[Y|A=1,W] - EE[Y|A=0,W]$ using the TMLE presented in Chapter 7 of van der Laan and Rose (2011). The procedure also yields an estimate of the influence curve, which we term $\widehat{IC}_{\hat{\psi}}$. This influence curve estimate takes as input an observation and outputs a number. The TMLE is defined in such a way that the empirical mean of $\widehat{IC}_{\hat{\psi}}(O)$ is zero.

We now present a TMLE procedure to estimate the causal bounds for the average treatment effect at a fixed δ. We remind the reader from (6) that these bounds are defined as

$$L_{\text{diff}}(P_0) = -E\Big[\delta^- E[Y|A=1,W]P_0(A=0|W) + \delta^+ E[1-Y|A=0,W]P_0(A=1|W)\Big]$$
$$U_{\text{diff}}(P_0) = E\Big[\delta^+ E[1-Y|A=1,W]P_0(A=0|W) + \delta^- E[Y|A=0,W]P_0(A=1|W)\Big].$$

We note that the choice of δ has no effect on the algorithm until Step 5.

1. Regress A_i against W_i, $i = 1, ..., n$, to obtain an estimate of the probability of treatment given covariates, and refer to this estimate of $P_0(A=1|W)$ as $g_n(W)$.

2. Regress Y_i against A_i and W_i, $i = 1, ..., n$, to obtain the regression function $E_n[Y|A,W]$, which takes as input A and W.

3. Run an intercept-free logistic regression of Y_i against the two covariates $H_0(O_i) = (1 - A_i)\frac{g_n(W_i)}{1-g_n(W_i)}$ and $H_1(O_i) = A_i\frac{1-g_n(W_i)}{g_n(W_i)}$ with offset $\text{logit}\, E_n[Y_i|A_i,W_i]$, $i = 1, ..., n$. Label the estimated coefficients on H_0 and H_1 as ϵ_0 and ϵ_1, respectively.

4. Define the following fluctuation of $E_n[Y|A,W]$:

$$E_n^*[Y|A,W] = \text{logit}^{-1}\left(\text{logit}\, E_n[Y|A,W] + \epsilon_0(1-A)\frac{g_n(W)}{1-g_n(W)} + \epsilon_1 A\frac{1-g_n(W)}{g_n(W)}\right).$$

5. Define our estimates of $L_{\text{diff}}(P_0)$ and $U_{\text{diff}}(P_0)$ as

$$\hat{\ell}_{\text{causal}} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{-\left[\delta^- E_n[Y|A=1,W_i]\,(1-A_i) + \delta^+\,(1-E_n[Y|A=0,W_i])\,A_i\right]}_{=f^{\delta}_{\hat{\ell}}(O_i)}$$

$$\hat{u}_{\text{causal}} = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left[\delta^+\,(1-E_n[Y|A=1,W_i])\,(1-A_i) + \delta^- E_n[Y|A=0,W_i]A_i\right]}_{=f^{\delta}_{\hat{u}}(O_i)}.$$

6. For each i, define the influence curve estimate for $\hat{\ell}_{\text{causal}}$ and $\hat{u}_{\text{causal}}$ as

$$\widehat{IC}_{\hat{\ell}}(O_i) = \frac{\delta^+(1-A_i)g_n(W_i) - \delta^- A_i(1-g_n(W_i))}{A_i g_n(W_i) + (1-A_i)(1-g_n(W_i))}(Y_i - E[Y|A_i,W_i]) + f^{\delta}_{\hat{\ell}}(O_i) - \hat{\ell}_{\text{causal}}$$

$$\widehat{IC}_{\hat{u}}(O_i) = \frac{\delta^-(1-A_i)g_n(W_i) - \delta^+ A_i(1-g_n(W_i))}{A_i g_n(W_i) + (1-A_i)(1-g_n(W_i))}(Y_i - E[Y|A_i,W_i]) + f^{\delta}_{\hat{u}}(O_i) - \hat{u}_{\text{causal}}.$$

7. Define $\hat{\Sigma}$ as the empirical covariance matrix of the function $(\widehat{IC}_{\hat{\psi}}+\widehat{IC}_{\hat{\ell}}, \widehat{IC}_{\hat{\psi}}+\widehat{IC}_{\hat{u}})$ applied to the n observations.

8. Construct the confidence interval as in (5), using the sample size n and the estimates $\hat{\psi}$, $\hat{\ell}_{\text{causal}}$, $\hat{u}_{\text{causal}}$, and $\hat{\Sigma}$.

In our application to the WCGS study we implement Steps 1 and 2 using the ensemble algorithm known as super-learner (van der Laan et al., 2007). We used the implementation in the `SuperLearner` package (Polley and van der Laan, 2013), and included as candidate algorithms `SL.glm`, `SL.randomForest`, `SL.earth`, `SL.glmnet`, and `SL.gam`.

Our estimator has been defined so that the influence curve estimates in Step 6 have empirical mean zero when applied to the data. Careful readers will notice that we have not written our estimator

using the usual TMLE setup in Step 5. That is, we have not defined some distribution P_n^* and then written $\hat{\ell}_{\text{causal}} = L_{\text{diff}}(P_n^*)$, and likewise we have not done this for $\hat{u}_{\text{causal}}$. Nonetheless, our estimator respects the bounds on the parameter space and thus there exists a distribution P_n^* such that $\hat{\ell}_{\text{causal}} = L_{\text{diff}}(P_n^*)$, and likewise there exists a P_n^* such that $\hat{u}_{\text{causal}} = U_{\text{diff}}(P_n^*)$. We refer readers interested in a more traditional substitution-type estimator to the implementation of the estimator for a closely related problem in Chapter 8 of van der Laan and Rose (2011).

A.2 How we derived our estimator

We now briefly describe how we came up with our estimator for $L_{\text{diff}}(P_0)$. The estimator for $U_{\text{diff}}(P_0)$ was derived analogously. Fix δ. First, we found the canonical gradient of the parameter L_{diff} which takes as input a distribution P and outputs a real number $L_{\text{diff}}(P)$. The canonical gradient is given by

$$IC(P)(O) = \underbrace{\frac{\delta^+(1-A)g_P(W) - \delta^- A(1-g_P(W_i))}{Ag_P(W) + (1-A)(1-g_P(W))}(Y - E_P[Y|A,W])}_{D_Y(E_P[Y|\cdot,\cdot],g_P)} + f^{\delta}_{\hat{\ell},P}(O) - \Psi(P),$$

where $g_P(W) = P(A=1|W)$ and

$$f^{\delta}_{\hat{\ell},P}(O) = -\left[\delta^- E_P[Y|A=1,W]\,(1-A) + \delta^+\,(1 - E_P[Y|A=0,W])\,A\right].$$

We refer the reader to Pfanzagl (1990) for details on the importance of the canonical gradient in semiparametric efficiency theory. We fluctuated our initial estimate $E_n[Y|A,W]$ of the outcome regression so that

$$\frac{1}{n}\sum_{i=1}^{n} D_Y(E_n^*[Y|\cdot,\cdot], g_n)(O_i) = 0.$$

We then chose our estimator of $L_{\text{diff}}(P_0)$ so that

$$\frac{1}{n}\sum_{i=1}^{n}\left[f^{\delta}_{\hat{\ell},P}(O_i) - \hat{\ell}_{\text{causal}}\right] = 0.$$

Let P_n^* be any distribution with $E_{P_n^*}[Y|A,W] = E_n^*[Y|A,W]$ and $g_{P_n^*} = g_n$. Straightforward calculations show that our procedure yields that that

$$\hat{\ell}_{\text{causal}} - L_{\text{diff}}(P_0) \approx \frac{1}{n}\sum_{i=1}^{n}\left[IC(P_n^*)(O_i) - E_{P_0}[IC(P_n^*)(O)]\right],$$

where the approximation holds up to a double robust remainder term which is small if the outcome regression and treatment mechanism are estimated well.

B Toy example code

```
rm(list=ls())
# Number of simulations to draw
n=1e7

# True E[Y|A,U,W]
EYgivenAUW = function(AA,UU,WW){0.8*AA*UU*(WW>=2)}

# Simulate the data
U = rbinom(n,1,0.9)
W = sample(1:4,n,replace=TRUE)
A = rbinom(n,1,0.75*(W==1) + 0.98*U * (W>=2))
Y = rbinom(n,1,EYgivenAUW(A,U,W))

# Data set restricted to treated individuals
A1.inds = which(A==1)
U1 = U[A1.inds]
W1 = W[A1.inds]
A1 = A[A1.inds]
Y1 = Y[A1.inds]

# Approximate E[Y1]
psi.causal = mean(EYgivenAUW(1,U,W))
# Approximate EE[Y|A=1,W]
PsiP0 = mean(W==1)*mean(Y1[which(W1==1)]) + mean(W==2)*mean(Y1[which(W1==2)])
    +
  mean(W==3)*mean(Y1[which(W1==3)]) + mean(W==4)*mean(Y1[which(W1==4)])


# Estimating Psi(P0)
# Estimate the misspecified m_{\beta}(W)
```

```
mW = mean(Y1[W1==1 | W1==3 | W1==4]) * (W==1 | W==3 | W==4) + mean(Y1[W1==2])
    * (W==2)
# Estimate the misspecified g_{\eta}(W)
gW = mean(A[W<=2]) * (W<=2) + mean(A[W>=3]) * (W>=3)

# $G$-estimation
Psi1P0.gest = mean(mW)
# Inverse probability of treatment weighted (IPTW) estimator
Psi1P0.iptw = mean(A*Y/gW)
# Double robust estimating equation
Psi1P0.dr = mean(A*(Y-mW)/gW + mW)


# Developing CI_{diff}
# EY.A1Wk = E[Y|A=1,W=k]
EY.A1W1 = mean(Y1[W1==1])
EY.A1W2 = mean(Y1[W1==2])
EY.A1W3 = mean(Y1[W1==3])
EY.A1W4 = mean(Y1[W1==4])

# E[Y|A=1,W]
EY.A1W = EY.A1W1*(W==1) + EY.A1W2*(W==2) + EY.A1W3*(W==3) + EY.A1W4*(W==4)

# EA.Wk = Pr(A=1|W=k)
EA.W1 = mean(A[W==1])
EA.W2 = mean(A[W==2])
EA.W3 = mean(A[W==3])
EA.W4 = mean(A[W==4])

# Pr(A=1|W)
EA.W = EA.W1*(W==1) + EA.W2*(W==2) + EA.W3*(W==3) + EA.W4*(W==4)

# Tight bounds
# LB_{diff}
lb.diff = -mean(EY.A1W*(1-EA.W))
# UB_{diff}
ub.diff = mean(A==0) - lb.diff



print('The causal parameter is')
psi.causal
print('The statistical parameter is')
PsiP0
print('The estimand using the misspecified $G$-estimation is')
Psi1P0.gest
print('The estimand using the misspecified IPTW estimator is')
Psi1P0.iptw
print('The estimand using the misspecified double robust estimator is')
Psi1P0.dr

print('P0(A=0) is approximately')
1-mean(A)
```

```
# psi_{causal}-Psi(P0)>LB_{diff} when delta=1
print(paste(round(psi.causal-PsiP0,5),'is greater than',round(lb.diff,5)))
# psi_{causal}-Psi_1(P0)>LB_{diff} when delta=1 for the misspecified $G$-
    estimator
print(paste(round(psi.causal-Psi1P0.gest,5),'is NOT greater than',round(lb.
    diff,5)))
# psi_{causal}-Psi_1(P0)>LB_{diff} when delta=1 for the misspecified IPTW
    estimator
print(paste(round(psi.causal-Psi1P0.iptw,5),'is NOT greater than',round(lb.
    diff,5)))
if(psi.causal-Psi1P0.dr<=lb.diff){
  # psi_{causal}-Psi_1(P0)>LB_{diff} when delta=1 for the misspecified double
      robust estimator
  print(paste(round(psi.causal-Psi1P0.dr,5),'is NOT greater than',round(lb.
      diff,5)))
} else {
  print('Not enough Monte Carlo draws to show that psi_{causal}-Psi_1(P0)>LB_{
      diff} when delta=1 for the misspecified double robust estimator.')
  print('Closed form calculations show psi_{causal}=0.54, Psi(P0) approx
      0.6114, and LB_{diff}=-0.0708.')
}
```

C Real data example code

Utility function file: `functions.R`

```
# Estimate EE[Y|A=1,W] - EE[Y|A=0,W] using a TMLE
tmlepsi <- function(q1, q0, g, a, y){

  g[g < 0.01] <- 0.01
  g[g > 0.99] <- 0.99

  q <- a * q1 + (1-a) * q0

  eps <- coef(glm(y ~ 0 + offset(qlogis(q)) + I(a/g - (1-a)/(1-g)),
                  family = binomial()))

  q1 <- plogis(qlogis(q1) + eps / g)
  q0 <- plogis(qlogis(q0) - eps / (1-g))
  q <- a * q1 + (1-a) * q0

  psi <- mean(q1 - q0)
  eif <- (a/g - (1-a)/(1-g)) * (y - q) + (q1 - q0) - psi

  return(list(psi = psi, eif = eif))

}

# Return a function which estimates the upper and lower bounds for a given
# value of the sensitivity parameter.
tmlebounds <- function(q1, q0, g, a, y, psihat, normaldraws){
```

```
  psi <- psihat$psi

  g[g < 0.01] <- 0.01
  g[g > 0.99] <- 0.99

  q <- a * q1 + (1-a) * q0

  eps0 <- coef(glm(y ~ 0 + offset(qlogis(q0)) + I(g/(1-g)),
                   family = binomial(), subset = a == 0))
  eps1 <- coef(glm(y ~ 0 + offset(qlogis(q1)) + I((1-g)/g),
                   family = binomial(), subset = a == 1))

  q0 <- plogis(qlogis(q0) + eps0 * g/(1-g))
  q1 <- plogis(qlogis(q1) + eps1 * (1-g)/g)

  q <- a * q1 + (1-a) * q0

  fun <- function(dM, dP, confid){
    # Efficient influence function of E[E[YY|AA=1,WW]P(AA=0|WW)]
    # qq = E[YY|AA=1,WW]
    # gg = P(YY=1|WW)
    eif.fun <- function(qq,gg,aa,yy){
      est.eq = (aa * (1 - gg) / gg) * (yy - qq) + (1 - aa - (1 - gg)) * qq +
          qq * (1-gg)
      est = mean(est.eq)
      eif = est.eq - est
      return(list(eif=eif,est=est))
    }

    l1 <- eif.fun(-q1,g,a,-y)
    u1 <- eif.fun(1-q1,g,a,1-y)
    l0 <- eif.fun(-q0,1-g,1-a,-y)
    u0 <- eif.fun(1-q0,1-g,1-a,1-y)

    lb <- dM * l1$est - dP * u0$est
    ub <- dP * u1$est - dM * l0$est

    eiflb <- dM * l1$eif - dP * u0$eif + psihat$eif
    eifub <- dP * u1$eif - dM * l0$eif + psihat$eif

    n <- length(y)
    dd <- data.frame(eiflb, eifub)
    sigma <- cov(dd)
    sigma.eig <- eigen(sigma)
    sigma.sqrt <- sigma.eig$vectors %*% diag(sqrt(sigma.eig$values)) %*% solve
        (sigma.eig$vectors)
    Zs <- (normaldraws%*%sigma.sqrt)/sqrt(n)

    s <- quantile(pmax(Zs[,1],-Zs[,2]),probs=0.95)

    return(c(psi + lb - s, psi + ub + s))

  }
```

```
  return(fun)

}

# Estimate confidence bounds for the causal average treatment effect at
    different values of the
# sensitivity parameters.
tmlecausal <- function(q1, q0, g, a, y, dMlim = c(0, 0.65), dPlim = c(0, 0.08)
    ,
                       npoints = 101, confid = 0.95, num.mc=2.5e4){

  dM <- seq(dMlim[1], dMlim[2], length.out = npoints)
  dP <- seq(dPlim[1], dPlim[2], length.out = npoints)

  psihat <- tmlepsi(q1, q0, g, a, y)

  # Use the same Monte Carlo draw for all settings
  normaldraws <- cbind(rnorm(num.mc),rnorm(num.mc))
  boundfun <- tmlebounds(q1, q0, g, a, y, psihat, normaldraws)

  ds  <- expand.grid(dM, dP)
  cis <- t(apply(ds, 1, function(x)boundfun(x[[1]], x[[2]], confid)))

  colnames(ds) <- c('dM', 'dP')
  colnames(cis) <- c('ll', 'ul')

  res <- cbind(ds, cis)
  return(res)
}
```

Code to run analysis and generate plots

```
# Preliminaries
rm(list = ls())
library(epitools)
library(SuperLearner)
library(randomForest)
library(earth)
library(glmnet)
library(gam)
library(ggplot2)
library(dplyr)
library(reshape2)

set.seed(123)

# Replace with directory functions.R is in
setwd('.')

# Read in the functions
```

```
source('functions.R')
# Read in the data
data(wcgs)

# Select variables to use
y <- select(wcgs, chd69)[, 1]
a <- as.numeric(select(wcgs, ncigs0)[, 1]>0)
w <- select(wcgs, age0, height0, weight0, dibpat0)

# Super-learner library
lib <- c('SL.glm', 'SL.randomForest', 'SL.earth', 'SL.glmnet', 'SL.gam')
# Outcome regression fit
fity <- SuperLearner(Y = y, X = data.frame(a=a, w), family = binomial(),
                     SL.library = lib)
# Treatment mechanism fit
fita <- SuperLearner(Y = a, X = w, family = binomial(), SL.library = lib)

# Estimate of E[Y|A=1,W]
q1 <- predict(fity, newdata = data.frame(a = 1, w))$pred[, 1]
# Estimate of E[Y|A=9,W]
q0 <- predict(fity, newdata = data.frame(a = 0, w))$pred[, 1]
# Estimate of P(A=1|W)
g  <- predict(fita, newdata = w)$pred[, 1]

summary(q1)
summary(q0)
summary(g)

# Estimate EE[Y|A=1,W] - EE[Y|A=0,W]
psihat <- tmlepsi(q1, q0, g, a, y)

# Return bounds for causal parameter at different values of the sensitivity
    parameters
sensanalysis <- tmlecausal(q1, q0, g, a, y)

# Make plots
dd <- melt(sensanalysis, id.vars = c('dP', 'dM'),
           measurevars = c('ul', 'll'))

table(dd$variable)
levels(as.factor(dd$variable))

dd$variable <- as.factor(dd$variable)
levels(dd$variable) <- c('Lower Limit', 'Upper Limit')

pp <- ggplot(dd[as.character(dd$variable)=='Lower Limit',], aes(x=dM,y=dP,z=
    value,fill=value,colour=value)) +
    theme_bw() + theme(panel.grid.major = element_blank(), panel.grid.minor =
        element_blank()) +
    scale_x_continuous(expand=c(0,0)) + scale_y_continuous(expand=c(0,0)) +
    geom_tile() + stat_contour(breaks=0,linemitre=10,size=1,colour='#2C2C2C',
        linetype='dashed') +
    scale_fill_gradient2(name='Lower Bound\nfor Average\nTreatment Effect',low
        ='blue', high='darkgreen',mid='white') +
```

```
    scale_colour_gradient2(low='blue', high='darkgreen',mid='white',guide=
        FALSE) +
    xlab(expression(delta ^ '-')) + ylab(expression(delta ^ '+'))
ggsave(pp,file='lower.pdf',width=6,height=4)
pp

dPquartiles <- quantile(dd$dP,(0:4)/4)
dd2 <- dd[is.element(dd$dP,dPquartiles),]
dd2$dP <- factor(dd2$dP)
pp <- ggplot(dd2, aes(x=dM,y=value,
  group=interaction(variable, dP),colour=dP)) + theme_bw() +
  theme(panel.border = element_rect(fill = NA,colour = "gray20",size=1)) +
  geom_hline(yintercept=0,colour='gray20') + geom_line() +
  geom_hline(yintercept=psihat$psi,linetype='dashed') +
  annotate('text',label='hat(psi)',x=max(dd2$dM)*0.97,y=psihat$psi+0.007,parse
      =TRUE) +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,0)) +
  xlab(expression(delta ^ '-')) + ylab('Bounds for Average Treatment Effect')
      +
  scale_colour_discrete(name = expression(delta ^ '+')) +
  theme(legend.title=element_text(size=12))
ggsave(pp,file='confidencebounds.pdf',width=6,height=4)
pp
```