INFERENCE ABOUT TREATMENT EFFECTS USING BOUNDS, SENSITIVITY ANALYSIS AND INSTRUMENTAL VARIABLES

Amy Richardson

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill
2014

Approved by:

Michael G. Hudgens

M. Alan Brookhart

Stephen R. Cole

Jason P. Fine

Haibo Zhou

ABSTRACT

AMY RICHARDSON: Inference about Treatment Effects Using Bounds, Sensitivity Analysis and Instrumental Variables
(Under the direction of Michael G. Hudgens)

This dissertation considers conducting inference about the effect of a treatment (or exposure) on an outcome of interest. In the ideal setting where treatment is assigned randomly, under certain assumptions the treatment effect is identifiable from the observable data and inference is straightforward. However, in many other settings observable data may only partially identify treatment effects or may identify treatment effects only for some subset of the population. In this case three approaches are often employed: (i) bounds are derived for the treatment effect under minimal assumptions, (ii) additional untestable assumptions are invoked that render the treatment effect identifiable and then sensitivity analysis is conducted to assess how inference changes as the untestable assumptions are varied, or (iii) instrumental variables are used to identify treatment effects for a subset of the population of interest. In this dissertation, first we review approaches (i) and (ii) in various settings, including assessing principal strata effects, direct and indirect effects, and effects of time-varying exposures. Methods for drawing formal inference about partially identified parameters are also discussed. Second, we derive the large sample properties of instrumental variable-based treatment effect estimators and test statistics when the outcome is subject to right censoring and competing risks. These results are applied to a real data example about the use of antiretroviral therapy to reduce mother to child transmission of HIV. Third, we derive identification results for direct, indirect and total effects of treatment in presence of interference (i.e., settings where the treatment of one individual may be affected by the treatment of other individuals). These results are applied to a real data example about rotavirus vaccination. All derived asymptotic results are supported by simulation studies.

ACKNOWLEDGMENTS

This work was supported in part by grant R01-AI085073 from the US National Institutes of Health (NIH). The content is solely the responsibility of the author and does not necessarily represent the official views NIH. The author would like to thank her dissertation advisor, Michael G. Hudgens, the committee M. Alan Brookhart, Stephen R. Cole, Jason P. Fine and Haibou Zhou, the BAN investigators for access to the data from their study, and the MarketScan Research Databases (Thomson Truven Healthcare, Inc.) for access to the data on rotavirus incidence and vaccination among U.S. infants.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

The goal of many health and epidemiological studies is to gain insight into the mechanisms that cause disease or other health outcomes and then use that insight to prevent disease and/or better health outcomes. These mechanisms may consist of one or more causal pathways in which different exposure states or risk factors lead to various effects (Rothman, 1976). Epidemiological studies often seek to investigate not only whether or not various sets of exposures or risk factors are on any of the causal pathways to a health outcome of interest Y but also the size and nature of the effects of these exposures on this causal pathway.

In hopes of gaining some information on causal effects of exposure states suspected of being on some causal pathway, data is collected and analyzed on some set of subjects or units in either a controlled experiment or under observational settings. However, as often noted by statistical scientists, effects estimated using conventional statistical methods can only measure association and do not have a causal interpretation. To estimate causal effects using more conventional methods, assumptions that are strong and often empirically untestable are needed. If these assumptions are dubious, resulting causal effect estimates are subject to biases and inferences may be misleading. In order to untangle differences between associational effects and causal effects precise mathematical notation and language is essential. The counterfactual or potential outcomes framework dating back to Neyman (1923) and formalized by Rubin (1974) allows for precise definitions of a myriad of causal effects and allows for distinction between associational and causal effects (Holland, 1986).

1.1.1 The Potential Outcomes Model

Using the potential outcomes approach from Rubin (1974), let z denote the various levels of the exposure or factor for which causal comparisons are being drawn, and Z the observed value of the exposure z which is observed prior to the outcome of interest Y. The potential outcome $Y(z)$ is defined as the value of the outcome under exposure status z; for example, z might be a medical intervention or treatment with $z = 1$ denoting treatment received and $z = 0$ denoting control or treatment not received. The potential outcome $Y(1)$ would then be the outcome had treatment $z = 1$ been received and $Y(0)$ the outcome had control been received.

Causal effects at the unit level can be defined as some function of the potential outcomes $Y(z)$ that compare different levels of the exposure z. In order to define sensible causal effects each level of the exposure in the effect must be able to be observed in the units under study. For example, an individual causal treatment effect might be defined as the difference between an individual's outcome under treatment compared to control, $Y(1) - Y(0)$, which is not well defined if it is not possible for an individual or unit to experience both $z = 0$ and $z = 1$ (Holland, 1986). Often the target of inference in health and epidemiological studies is some population level parameter or population level causal effect such as the mean difference between the potential outcomes under treatment and control, $E[Y(1] - Y(0)]$ or average treatment effect (ATE). Here $E[Y(1)]$ is the mean potential outcome of the target population if everyone were treated ($Z = 1$), and $E[Y(0)]$ the mean potential outcome if no one were treated ($Z = 0$).

In order to make inferences about these population causal effects, assumptions regarding the relationships between the observed outcome Y, the observed exposure Z, and the potential outcomes $Y(z)$ must be made. One of the first assumptions regularly made when studying causal effects under the potential outcomes framework is that each level z of the exposure Z maps to one fixed potential outcome $Y(z)$. A second basic assumption connects the observed outcome Y to the potential outcomes $Y(z)$; specifically it is assumed that $Y(Z) = Y$, which means that the observed outcome is equal to the potential outcome under the observed

exposure Z. This assumption has been referred to as a consistency assumption in the literature (Cole and Frangakis, 2009) and will be termed causal consistency here to avoid confusion with concepts of statistical consistency. With this assumption, one potential outcome for each individual is observed and known, but the potential outcomes $Y(z')$ for $Z \neq z'$ are termed counterfactual and are unobserved. A third basic assumption frequently made in studying causal effects is that the exposure of one unit under study does effect the outcomes of other individuals or units, this is referred to as an assumption of no interference between units. Collectively these three assumptions are often referred to as the stable unit treatment value assumption or SUTVA (Rubin, 1980).

The assumptions contained in SUTVA will be reasonable in many situations, but unfortunately are not strong enough to allow for estimation of most population level causal effects such as the average treatment effect, $E[Y(1) - Y(0)]$. Under SUTVA, the observation of Y and Z for some population of units allows us to estimate $E[Y(z)|Z = z]$, thus allowing for estimation of the associational effect

$$E[Y(1)|Z = 1] - E[Y(0)|Z = 0], \tag{1.1}$$

but estimation of the ATE, a causal effect, is not possible without further assumptions. Under experimental settings the observed exposure or treatment Z might be under the control of the experimenter and random assignment of Z to the units under study would give plausibility to the assumption that

$$Y(z) \amalg Z \text{ for } z = 0, 1 \tag{1.2}$$

(here $\amalg$ denotes statistical independence). Under (1.2) $E[Y(z)|Z = z] = E[Y(Z)]$ and thus the average treatment effect may be consistently estimated using the estimated associational effect (1.1). In absence of random assignment of Z, (1.2) may be dubious and an estimator based on (1.1) is subject to bias. Specifically, the associational effect (1.1) between the exposure Z and the outcome Y may have resulted from some unknown or unmeasured factor(s) X that is associated with both the exposure and the outcome. The factors in X are said to confound the effect of Z on Y. Epidemiologists often seek to measure different variables in X, if this

can be accomplished then (1.2) might be replaced by

$$Y(z) \amalg Z|X \text{ for } z = 0, 1, \tag{1.3}$$

which will be plausible if all factors that confound the causal effect of Z on Y are measured in X. Under (1.3) the ATE may be consistently estimated using by weighting by the inverse probability of exposure z

$$\frac{E[Y(1)|Z = 1; X]}{\Pr[Z = 1|X]} - \frac{E[Y(0)|Z = 0; X]}{\Pr[Z = 0|X]} \tag{1.4}$$

which is a function of associational parameters that may be consistently estimated from the data. Estimators based on (1.4) are referred to as inverse probability of treatment weighted estimators.

1.1.2 Instrumental Variables

Measuring all factors X such that (1.3) holds is one of the biggest challenges of causal inference in epidemiological research, particularly because it is not possible to provide evidence that (1.3) holds using empirical statistical tests. If there are factors U not measured in X that confound the effect of Z on $Y(z)$ then the resulting inverse probability estimators will be biased. A method to potentially avoid this problem of unmeasured confounding entails the use of instrumental variables. A variable R is considered an instrumental variable if it meets the following three criteria: i) R has a causal effect on the exposure of interest Z, ii) R affects the outcome Y only through its effect on Z and iii) R does not share common causes with Y (Hernán and Robins, 2006).

Specifically, under a set of assumptions that may be more reasonable than (1.2) or (1.3), the instrumental variable allows for estimation of a causal effect known as a local average treatment effect or a principle treatment effect. To illustrate, let $Z(r)$ be the potential values of the exposure Z for different levels of the instrument r, without loss of generality assume that the exposure Z, the outcome Y and the instrument R are all binary. Define S^{P_0} as the vector

of the two potential values of $Z(r)$, $S^{P_0} = (Z(0), Z(1))$, stratification of the potential outcomes $Y(z)$ by S^{P_0} is commonly referred to as principal stratification (Frangakis and Rubin, 2002). A local average treatment effect or principle effect is the average treatment effect in one of the strata defined by S^{P_0}, where causal effects in the strata defined by $S^{P_0} = (0, 1)$ are commonly of interest. Imbens and Angrist (1994) showed the local average treatment effect (LATE) defined as $E[Y(1) - Y(0)|S^{P_0} = (0, 1)]$ is identifiable under four assumptions: independent treatment instrument

$$R \amalg \{Y(z), Z(r)\} \text{ for } z.r = 0, 1, \tag{1.5}$$

monotonicity with respect to Z

$$\Pr[Z(1) \geq Z(0)] = 1, \tag{1.6}$$

exclusion restriction

$$Y(0) = Y(1) \text{ if } Z(0) = Z(1), \tag{1.7}$$

and if there is a nonzero causal effect of R on Z, namely

$$E[Z(1) - Z(0)] \neq 0. \tag{1.8}$$

The monotonicity assumption (2.21) states there are no individuals such that $Z(0) = 1$ and $Z(1) = 0$, meaning that the principal strata $S^{P_0} = (1, 0)$ is empty. Assumption (1.7) states that Z has no effect on Y in individuals who are always exposed $S^{P_0} = (1, 1)$ or are never exposed $S^{P_0} = (0, 0)$. Assumption (1.8) indicates that the instrument R has a causal effect on the exposure Z. Under these four assumptions the LATE can be expressed as

$$\frac{E[Y|R = 1] - E[Y|R = 0]}{E[Z|R = 1] - E[Z|R = 0]}. \tag{1.9}$$

Under these four assumptions, (1.9) is simply the ratio of the average associational effect of R on Y and the average associational effect of R on Z; (1.9) is referred to as the instrumental variable estimand (Angrist et al., 1996; Hernán and Robins, 2006). To see that

$E[Y(1)-Y(0)|S^{P_0}=(0,1)]$ equals (1.9), first note under the assumptions that the numerator of (1.9) equals $E[Y(1)-Y(0)] = E[\{Y(1)-Y(0)\}\{Z(1)-Z(0)\}] = E[Y(1)-Y(0)|Z(1) > Z(0)]\Pr[Z(1) > Z(0)]$. Similarly, the denominator of (1.9) equals $\Pr[Z(1)=1] - \Pr[Z(0)=1] = \Pr[Z(0)=0, Z(1)=1] = \Pr[Z(1) > Z(0)]$, which is non-zero under (1.8).

Obtaining a valid instrument such that (1.5–1.8) hold can prove to be a difficult task. An example of a variable R that might satisfy (i)-(iii) such that (1.5–1.8) hold is the calendar time for the FDA approval of a novel treatment for a disease, where here Z would be the novel treatment. Let $R = 1$ denote that diagnosis of the disease was after the calendar time for the approval of the new treatment, and $R = 0$ indicate diagnosis was before this calendar time; let $Z = 1$ denote that the novel treatment was selected and $Z = 0$ denote that treatment was not selected. The principal strata vector would indicate a subject's treatment selection before and after the calendar time FDA approval for the novel treatment, for instance $S^{P_0} = (1,1)$ would be represent individuals that would take the treatment regardless of the FDA calendar time approval. In this situation Y might represent survival to a given time point after having been diagnosed with the disease, thus the local average treatment defined as $E[Y(1) - Y(0)|S^{P_0} = (0,1)]$ would represent the difference in the proportion surviving amongst those who took the novel treatment versus those whom did not in the principal stratum wherein individuals would take treatment only after the FDA calendar time approval.

The assumption in (1.5) states that there are no factors that confound the relationship between the outcome Y and the instrument R; randomization of the instrument R can insure that this assumption is met making instruments that can be randomized attractive. In a randomized clinical trial with noncompliance a commonly used instrumental variable is treatment assignment (here the exposure would be the actual treatment taken). Another example might be randomized treatment assignment in an encouragement randomized trial where subjects are randomly assigned to be enrolled in programs that encourage (or discourage) the exposure Z, while others are randomized to control, or no encouragement program. In some situations natural randomization processes, such as Mendelian randomization, might provide a valid instrumental variable. For instance, a study investigating a causal effect between low

serum cholesterol and cancer might use a genetic determinant of serum cholesterol as the instrumental variable (Martens et al., 2006).

1.1.3 Interference

When studying causal effects of a treatment or exposure, sometimes the treatment or exposure received by one individual may affect the outcomes of other individuals under study. In the causal inference literature this is referred to as interference and is most frequently encountered in settings in which outcomes are largely dependent on social happenings. Some well known examples of settings where this might occur include the study of infectious diseases and vaccination, educational interventions and effects of housing voucher programmes. Until recently most of the causal inference literature has operated under the assumption that there is no interference between units (Cox, 1958, this assumption is included in SUTVA). In the aforementioned settings this assumption is not only undoubtedly violated, but the pattern of interference between units is often a target of inference useful in determining important social and public health policies.

Though most of the causal inference literature operates under the assumption of no interference, Rubin (1980) noted that the potential outcomes framework could be extended to accommodate interference between units. Specifically, assume that there are $N > 1$ groups or communities for which data are observed, with each group having n_i individuals for $i = 1, \ldots, N$. Let $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{in_i})$ denote the treatment selections or exposures of those n_i individuals for each group i. Assume that Z_{ij} is a dichotomous, taking values 0 for no exposure or treatment not selected, and 1 for exposure or treatment selected. Let $\boldsymbol{Z}_{i,-j} = (Z_{i1}, \ldots Z_{ij-1}, Z_{ij+1}, \ldots, Z_{in_i})$ denote the $n_i - 1$ subvector of treatment selections for group i with entry j deleted. Define $\mathcal{Z}(n_i)$ as the set of possible treatment selections for a group of size n_i. $\boldsymbol{Z}_{i,-j}$ takes on values in the set $\mathcal{Z}(n_i - 1)$. There are 2^{n_i} different realizations of the vector $\boldsymbol{Z}_i$, and $2^{n_i - 1}$ realizations of $\boldsymbol{Z}_{i,-j}$. For each subject in each group we extend the potential outcomes such that there is a separate potential outcome for each permutation of the treatment allocation vector $\boldsymbol{Z}_i$. Denote the potential outcome for the jth person in the

ith group for treatment allocation vector $\boldsymbol{Z}_i$ by $Y_{ij}(\boldsymbol{Z}_i)$. Denote the potential outcomes for all members of group i as $\boldsymbol{Y}_i(\boldsymbol{Z}_i)$. This allows for interference between members of the same group, but does not allow for interference between members of different groups. This is an assumption, and is referred to as partial interference in the literature. In Manski (2013) this assumption is a specific case of a more general class of assumptions which he calls constant treatment response assumptions (CTR). This assumption is reasonable if interaction between members of different groups is minimal or nonexistent.

Halloran and Struchiner (1995) took Rubin's suggestion and defined several new causal effects unique to studying interference: direct, indirect, total and overall effects. The individual direct, indirect, total and overall effects are defined as

$$\begin{aligned}
DE_{ij}(\boldsymbol{z}_{i,-j}) =& Y_{ij}(\boldsymbol{z}_{i,-j}, z_{ij}=0) - Y_{ij}(\boldsymbol{z}_{i,-j}, z_{ij}=1),\\
IE_{ij}(\boldsymbol{z}_{i,-j}, \boldsymbol{z}'_{i,-j}) =& Y_{ij}(\boldsymbol{z}_{i,-j}, z_{ij}=0) - Y_{ij}(\boldsymbol{z}'_{i,-j}, z_{ij}=0),\\
TE_{ij}(\boldsymbol{z}_{i,-j}, \boldsymbol{z}'_{i,-j}) =& Y_{ij}(\boldsymbol{z}_{i,-j}, z_{ij}=0) - Y_{ij}(\boldsymbol{z}'_{i,-j}, z_{ij}=1), \text{ and}\\
OE_{ij}(\boldsymbol{z}_i, \boldsymbol{z}'_i) =& Y_{ij}(\boldsymbol{z}_i) - Y_{ij}(\boldsymbol{z}'_i).
\end{aligned}$$

The direct effect compares potential outcomes that keep the treatment of other members of the same group constant and comparing the effect of treatment in the individual. The indirect effect compares two different treatment allocation vectors given to other members of the group while holding the treatment given to the individual constant at $z_{ij}=0$. The total effect compares both the different treatment allocation vectors and the effect of treatment in the individual. The overall effect compares any two treatment allocation vectors for the whole group and may correspond to a direct effect, indirect effect or a total effect, or another effect.

Often it is still of interest to compare 2 specific treatment allocation strategies or laws, denoted $\boldsymbol{\pi}_i(\boldsymbol{Z}_i, \alpha_0)$ and $\boldsymbol{\pi}_i(\boldsymbol{Z}_i, \alpha_1)$, say for example comparing causal effects when vaccinating 1/3 of the population versus vaccinating 2/3 thirds of the population. The parameters α_0 and α_1 index the two treatment allocation law of which comparisons of causal effects are desired.

For the purposes here, a Bernoulli type parametrization of $\boldsymbol{\pi}_i(\boldsymbol{Z}_i, \alpha_0)$ will be assumed:

$$\boldsymbol{\pi}_i(\boldsymbol{Z}_i, \alpha_0) = \prod_{j=1}^{n_i} \alpha_0^{Z_{ij}}(1-\alpha_0)^{1-Z_{ij}} \tag{1.10}$$

Sobel (2006) introduced the idea of defining causal estimands that average over all possible treatment assignment vectors according to some treatment allocation law in a paper assessing the comparative effectiveness of different housing voucher programmes. Specifically, let

$$\overline{Y}_{ij}(z, \alpha_0) = \sum_{\boldsymbol{s} \in \mathcal{Z}(n_i-1)} Y_{ij}(\boldsymbol{z}_{i,-j} = \boldsymbol{s}, z_{ij} = z)\Pr_{\alpha_0}(\boldsymbol{Z}_{i,-j} = \boldsymbol{s} | Z_{ij} = z)$$

and

$$\overline{Y}_{ij}(\alpha_0) = \sum_{\boldsymbol{s} \in \mathcal{Z}(n_i)} Y_{ij}(\boldsymbol{z}_i = \boldsymbol{s})\boldsymbol{\pi}_i(\boldsymbol{Z}_{i,j} = \boldsymbol{s}; \alpha_0);$$

then the individual average direct, indirect, total and overall effects are defined as

$$\begin{aligned}
\overline{DE}_{ij}(\alpha_0) =& \overline{Y}_{ij}(0; \alpha_0) - \overline{Y}_{ij}(1; \alpha_0), \\
\overline{IE}_{ij}(\alpha_0, \alpha_1) =& \overline{Y}_{ij}(0; \alpha_0) - \overline{Y}_{ij}(0; \alpha_1), \\
\overline{TE}_{ij}(\alpha_0, \alpha_1) =& \overline{Y}_{ij}(0; \alpha_0) - \overline{Y}_{ij}(1; \alpha_1), \\
\overline{OE}_{ij}(\alpha_0, \alpha_1) =& \overline{Y}_{ij}(\alpha_0) - \overline{Y}_{ij}(\alpha_1).
\end{aligned}$$

Now the direct effect compares the effect of treatment in individual ij while holding the treatment allocation strategy constant, the indirect effect compares the effect of the treatment allocation strategy constant while holding the treatment to the individual constant at $z_{ij} = 0$. The total effect compares both the treatment allocation strategies and treatment given to individual ij. Group average direct, indirect, total and overall effects can be defined by taking the mean for group i of $\overline{DE}_{ij}(\alpha_0)$, $\overline{IE}_{ij}(\alpha_0, \alpha_1)$, $\overline{TE}_{ij}(\alpha_0, \alpha_1)$ and $\overline{OE}_{ij}(\alpha_0, \alpha_1)$ (i.e. $\overline{DE}_i(\alpha_0) = \sum_{j=1}^{n_i} \overline{DE}_{ij}(\alpha_0)$ and so forth). Population average direct, indirect, total and overall effects can be defined by taking the mean of the group average direct, indirect, total and overall effects (i.e. $\overline{DE}(\alpha_0) = \sum_{i=1}^{N} \overline{DE}_i(\alpha_0)$ and so forth).

The so called gold standard for achieving accurate estimates of causal estimands comparing these two allocation strategies is 2-stage randomization, where both individual treatment assignment is randomized as well as treatment allocation strategy to various groups or communities. The majority of the inferential methods developed for the population average direct, indirect, total and overall effects rely on the assumption that there are two levels of randomization. Rosenbaum (2007) developed nonparametric inferential methods for assessing treatment effects in presence of interference under 2 stage randomized treatment assignment. Hudgens and Halloran (2008) formalized the definitions of direct, indirect, total and overall effects averaged over all possible treatment assignment vectors and developed unbiased estimators and corresponding variance upper bounds for these causal treatment effects under 2 stage randomization and an additional assumption referred to as stratified interference. Stratified interference assumes

$$Y_{ij}(\boldsymbol{z}_{i,-j}, z_{ij}) = Y_{ij}(\boldsymbol{z}'_{i,-j}, z_{ij}) \text{ for } \sum_j z_{ij} = \sum_j z'_{ij}.$$

which means that potential outcomes will remain constant when the same number of other members of the group are treated and treatment to the individual remains constant. This reduces the number of potential outcomes for each individual from 2^{n_i} to n_i. Tchetgen Tchetgen and Vanderweele (2012) improved upon the variance bounds developed by Hudgens and Halloran (2008) for these effects under 2 stage randomization by relaxing the stratified interference assumption.

Hong and Raudenbush (2006) consider interference effects in the context of educational performance. In this setting randomization is not present at either the individual level or the group level. The independent treatment assignment assumption that 2 stage randomization makes plausible

$$\{Y(\boldsymbol{z}_i)\}_{\boldsymbol{z}_i \in \mathcal{Z}(n_i)} \amalg \boldsymbol{Z}_i$$

for $i = 1, \ldots, N$ is replaced by

$$\{Y(\boldsymbol{z}_i)\}_{\boldsymbol{z}_i \in \mathcal{Z}(n_i)} \amalg \boldsymbol{Z}_i | \boldsymbol{X}_i$$

for some set of covariates $\boldsymbol{X}_i$. Tchetgen Tchetgen and Vanderweele (2012) derived the following inverse probability of treatment weighted estimators for $\overline{Y}_i(z, \alpha_0)]$ (the group average) to be used for inference of the population average direct, indirect, and total effects

$$\widehat{Y}_i^{IPW}(z, \alpha_0) = \frac{\sum_{j=1}^{n_i} \pi_i(\boldsymbol{Z}_{i,-j}; \alpha_0) Y_{ij}(\boldsymbol{Z}_i) I[Z_{ij} = z]}{n_i f_{\boldsymbol{Z}|\boldsymbol{X},i}(\boldsymbol{Z}_i|\boldsymbol{X}_i)}$$

for a the Bernoulli type parametrization of $\pi(\boldsymbol{Z}_i, \alpha_0)$ given above and where $f_{\boldsymbol{Z}|\boldsymbol{X},i}(\boldsymbol{z}_i|\boldsymbol{X}_i) = \Pr[\boldsymbol{Z}_i = \boldsymbol{z}_i]$ is the estimated probability of treatment allocation vector $\boldsymbol{X}_i$ given covariates $\boldsymbol{X}_i$. Tchetgen Tchetgen and Vanderweele (2012) suggest using a logistic-normal mixed effects model to estimate $f_{\boldsymbol{Z}|\boldsymbol{X},i}(\boldsymbol{Z}_i|\boldsymbol{X}_i)$.

Though the advancements made account for interference and allow for definition of causal effects specific to interference, many of the results obtained are limited by the need for 2 stage randomized designs requiring randomization at the individual level within each group, as well as randomization of groups to different treatment allocation strategies. Such designs are difficult, all but infeasible to implement in practice, thus there is a strong need for adaptations to be used for observational data or for randomized designs not necessarily having achieving randomization at both the group and the individual level. Both Hong and Raudenbush (2006) and Tchetgen Tchetgen and Vanderweele (2012) have obtained results for observational data under the assumption that conditional on measured covariates, treatment assignment is independent of the potential outcomes. Hong and Raudenbush (2006) developed results for estimators of interference effects within strata defined by different levels of $\boldsymbol{X}_i$ and Tchetgen Tchetgen and Vanderweele (2012) derived inverse probability of treatment weighted estimators of interference effects. Both the results of Hong and Raudenbush (2006) and Tchetgen Tchetgen and Vanderweele (2012) enjoy the same results obtained under 2 stage randomization in terms of inference, but are limited by the fact that they rely on a strong conditional independent treatment assignment assumption and require measurement of all covariates required for conditional independence of the treatment and potential outcomes.

Manski (2013) develops bounds for interference effects under assumptions that are weaker and maybe more plausible than the conditionally independent treatment allocation of Hong

and Raudenbush (2006) and Tchetgen Tchetgen and Vanderweele (2012), but such bounds are only for effects pertaining to group level or population level means in presence of interference and are not for causal estimands averaged according to some treatment allocation law.

CHAPTER 2: NONPARAMETRIC BOUNDS AND SENSITIVITY ANALYSIS OF TREATMENT EFFECTS

2.1 Introduction

In many areas of science, interest often lies in assessing the causal effect of a treatment (or exposure) on some particular outcome of interest. For example, researchers may be interested in estimating the difference between the average outcomes when all individuals are treated (exposed) versus when all individuals are not treated (unexposed). When treatment is assigned randomly and there is perfect compliance to treatment assignment, such treatment effects are identifiable and inference about the effect of treatment proceeds in a straightforward fashion. On the other hand, if the treatment assignment mechanism is not known to the analyst or compliance is not perfect, then these treatment effects are not identifiable from the observable data.

A statistical parameter is considered identifiable if different values of the parameter give rise to different probability distributions of the observable random variables. A parameter is partially identifiable if more than one value of the parameter gives rise to the same observed data law, but the set of such values is smaller than the parameter space. Traditionally, statistical inference has been restricted to the situation when parameters are identifiable. More recent research has considered methods for conducting inference about partially identifiable parameters. This research has been motivated to some extent by methods to evaluate causal effects of treatment, which are frequently partially identifiable. For instance, causal estimands are typically only partially identifiable in observational studies where the treatment selection mechanism is not known to the analyst. Noncompliance in randomized trials may also render treatment effects partially identifiable and a large amount of research has been devoted to drawing inference about treatment effects in the presence of noncompliance. Partial identifi-

ability also arises when drawing inference about treatment effects within principal strata or effects describing relationships between an outcome and a treatment that are mediated by some intermediate variable.

In order to conduct inference about treatment effects that are partially identifiable, two approaches are often employed: (i) bounds are derived for the treatment effect under minimal assumptions, or (ii) additional untestable assumptions are invoked under which the treatment effect is identifiable and then sensitivity analysis is conducted to assess how inference about the treatment effect changes as the untestable assumptions are varied. Below (i) and (ii) are illustrated in five settings. In Section 2.2 we consider treatment effect bounds and sensitivity analysis when the treatment assignment mechanism is unknown. In Section 2.3 partial identifiability of principal strata causal effects are discussed. In Section 2.4 the setting of non-compliance is considered where there is interest in assessing the effect of treatment if there was perfect compliance. In Section 2.5 bounds and sensitivity analysis for direct and indirect effects in mediation analysis are presented, and in Section 2.6 longitudinal treatment effects are considered. Much of the literature on bounds and sensitivity analysis focuses on ignorance due to partial identifiability and tends to ignore uncertainty due to sampling error. Section 2.7 presents some methods that appropriately quantify this uncertainty when drawing inference about partially identifiable treatment effects. Section 2.8 concludes with a discussion.

2.2 Treatment Selection

2.2.1 Minimal Assumptions Bounds

Suppose we have a random sample of individuals where each potentially receives treatment or control. Unless otherwise indicated, let Z indicate treatment received where $Z = 1$ denotes treatment and $Z = 0$ denotes control. Denote the observed outcome of interest by Y. In order to define a treatment effect on the outcome Y, we first define potential outcomes for an individual when receiving treatment, denoted $Y(1)$, and when receiving control, denoted $Y(0)$.

Throughout this paper we invoke the stable unit treatment value assumption (SUTVA; Rubin, 1980), i.e., there is no interference between units and there are no hidden (unrepresented) forms of treatment such that each individual has two potential outcomes $\{Y(0), Y(1)\}$. The no hidden forms of treatment guarantees that the observed outcome is equal to the potential outcome corresponding to the observed treatment, namely that $Y = Y(z)$ for $Z = z$. Here this will be referred to as the causal consistency assumption; for further discussion of causal consistency see Pearl (2010) and references therein. Once an individual receives treatment Z, the potential outcome $Y(Z)$ is observed and the other potential outcome (or counterfactual) $Y(1-Z)$ becomes missing. Assume that n iid copies of (Z, Y) are observed and denoted by (Z_i, Y_i) for $i = 1, \ldots, n$.

In this section we consider treatment effect bounds when the treatment assignment mechanism is unknown. Here Z can be thought of as treatment selection by the individual or by nature, rather than random treatment assignment as in an experiment. Define the average treatment effect ATE to be $E[Y(1)-Y(0)] = E[Y(1)]-E[Y(0)]$ where E denotes the expected value. The ATE can be decomposed as

$$\sum_{z=0}^{1} E[Y(1)|Z = z] \Pr[Z = z] - \sum_{z=0}^{1} E[Y(0)|Z = z] \Pr[Z = z]. \tag{2.1}$$

Note $E[Y(z)|Z = z] = E[Y|Z = z]$ by the causal consistency assumption. Thus from the observed data $E[Y(z)|Z = z]$ and $\Pr[Z = z]$ are identifiable and can be consistently estimated by their empirical counterparts. On the other hand, the observed data provide no information about $E[Y(z)|Z = 1 - z]$, such that (2.1) is only partially identifiable without additional assumptions.

Bounds on $E[Y(1)-Y(0)]$ can be obtained by entertaining the smallest and largest possible values for $E[Y(z)|Z = 1-z]$. If $Y(1)$ and $Y(0)$ are not bounded then bounds on $E[Y(1)-Y(0)]$ will be completely uninformative, ranging from $-\infty$ to ∞. Thus informative bounds are only possible if $Y(0)$ and $Y(1)$ are bounded. Because any bounded variable can be rescaled to take values in the unit interval, without loss of generality assume $Y(z) \in [0, 1]$ for $z = 0, 1$. Then $0 \leq E[Y(z)|Z = 1 - z] \leq 1$ and from (2.1) it follows that $E[Y(1) - Y(0)]$ is bounded below

by setting $E[Y(1)|Z=0]=0$ and $E[Y(0)|Z=1]=1$, which yields the lower bound

$$E[Y(1)|Z=1]\Pr[Z=1]-E[Y(0)|Z=0]\Pr[Z=0]-\Pr[Z=1]. \tag{2.2}$$

Similarly $E[Y(1)-Y(0)]$ is bounded above by setting $E[Y(1)|Z=0]=1$ and $E[Y(0)|Z=1]=0$, which yields the upper bound

$$E[Y(1)|Z=1]\Pr[Z=1]-E[Y(0)|Z=0]\Pr[Z=0]+\Pr[Z=0]. \tag{2.3}$$

These bounds were derived independently by Robins (1989) and Manski (1990). The lower and upper bounds (2.2) and (2.3) are sharp in the sense that it is not possible to derive narrower bounds without additional assumptions. Note the interval formed by (2.2) and (2.3) is contained in $[-1,1]$ and is of width 1. Thus the bounds are informative in that the treatment effect is now restricted to half of the otherwise possible range $[-1,1]$. On the other hand, the bounds will always contain the null value 0 corresponding to no average treatment effect. That is, without additional assumptions the sign of the treatment effect cannot be determined from the observable data.

2.2.2 Additional Assumptions

The bounds (2.2) – (2.3) are sometimes called the "no assumptions" or "worst case" bounds because no assumptions are made about the effect of treatment in the population (Lee, 2005; Morgan and Winship, 2007). The only assumptions made in deriving (2.2) and (2.3) are SUTVA and that the observed data constitute a random sample. If additional assumptions are invoked, the treatment effect bounds may become tighter (i.e., narrower) or even collapse to a point (i.e., the treatment effect may become identifiable). Sometimes these additional assumptions will have implications that are testable based on the observed data. Should the observed data provide evidence against an assumption under consideration, then bounds should be computed without making this assumption.

An example of an additional assumption is mean independence, i.e.,

$$E[Y(z)|Z=0] = E[Y(z)|Z=1] \text{ for } z = 0, 1. \tag{2.4}$$

Under (2.4) ATE is identifiable. Specifically the upper and lower bounds (2.2) and (2.3) both equal $E[Y(1)|Z=1] - E[Y(0)|Z=0]$, which is identifiable from the observable data and can be consistently estimated by the "naive" estimator given by the difference in sample means between the groups of individuals receiving treatment and control. Assumption (2.4) will hold in experiments where treatment is randomly assigned as in a randomized clinical trial. Moreover, in randomized experiments the stronger assumption

$$Y(z) \amalg Z \text{ for } z = 0, 1, \tag{2.5}$$

will hold, which in turn implies (2.4).

In some settings it may be reasonable to consider additional assumptions that are not as strong as (2.4) or (2.5) but nonetheless lead to tighter bounds than (2.2) and (2.3). For example, monotonicity type assumptions might be considered, such as monotone treatment selection (MTS)

$$E[Y(z)|Z=1] \geq E[Y(z)|Z=0] \text{ for } z = 0, 1. \tag{2.6}$$

MTS assumes individuals who select treatment will on average have outcomes greater than or equal to that of individuals who do not select treatment under the counterfactual scenario all individuals selected the same z. Manski and Pepper (2000) consider MTS when examining the effect of returning to school on wages later in life. For this example, MTS implies individuals who choose to return to school will have higher wages on average compared to individuals who choose to not return to school under the counterfactual scenario no individuals return to school. Alternatively, one might assume monotone treatment response (MTR)

$$\Pr[Y(1) \geq Y(0)] = 1$$

(Manski, 1997). MTR assumes that under treatment each individual will have a response greater than or equal to that under control. For instance, suppose $Z = 1$ if an individual elects to get the annual influenza vaccine and $Z = 0$ otherwise, and let $Y(z) = 1$ if an individual subsequently does not develop flu-like symptoms when $Z = z$, and $Y(z) = 0$ otherwise. MTR asserts that each individual is more or as likely to not develop flu-like symptoms if they are vaccinated versus if they are unvaccinated. Given to date there is no evidence that the annual flu vaccine enhances the probability of acquiring influenza, MTR might be plausible for this example.

Assuming MTS or MTR can lead to narrower bounds than (2.2) and (2.3) because they imply additional constraints on unobserved counterfactual expectations. For example, assuming MTS, $E[Y(0)|Z = 1]$ is bounded below by $E[Y(0)|Z = 0]$ and $E[Y(1)|Z = 0]$ is bounded above by $E[Y(1)|Z = 1]$, implying the upper bound on $E[Y(1) - Y(0)]$ is

$$E[Y(1)|Z = 1] - E[Y(0)|Z = 0], \tag{2.7}$$

for which the naive estimator is consistent. Under MTS the lower bound remains (2.2). In contrast to the no assumptions bounds, assuming MTS the bounds may exclude 0, specifically when (2.7) is negative. MTR implies $E[Y(1)] \geq E[Y(0)]$ which in turn implies that the ATE lower bound is 0. Under MTR the upper bound remains (2.3).

2.2.3 AZT Example

To illustrate the bounds above consider a hypothetical study of 2000 HIV patients (from Figure 2 of Robins, 1989) where 1400 individuals elected to take the drug AZT and 600 elected not to take AZT (this is a simplified version of the problem Robins considers). The outcome of interest is death or survival at a given time point. Of the 2000 patients, 1000 died with exactly 500 from each group. Let $Z = 1$ if the patient elected to take AZT and $Z = 0$ otherwise; let $Y = 1$ if the individual died and 0 otherwise. The naive estimator, i.e., the difference in sample means between $Z = 1$ and $Z = 0$, equals 500/1400-500/600 $\approx$ -0.48. The

empirical estimates of the no assumptions bounds (2.2) and (2.3) equal -0.7 and 0.3. In this setting, the MTS assumption (2.6) supposes that individuals who elected to take AZT would have been more or as likely to die as individuals who did not take AZT in the counterfactual scenarios where everyone receives treatment or everyone does not receive treatment. This might be reasonable if it is thought that those who took AZT were on average less healthy than those who did not. Assuming MTS, the upper bound (2.7) is estimated to be -0.48. Thus in this example the MTS bounds are substantially tighter than the no assumption bounds. The estimated MTS bounds lead to the conclusion (ignoring sampling variability, a point which we return to later) that AZT reduces the probability of death by at least 0.48 whereas without the MTS assumption we cannot even conclude whether the effect of treatment is non-zero.

2.2.4 Sensitivity Analysis

Assumptions such as (2.4) or (2.5) which identify the ATE, or assumptions such as MTS which sharpen the bounds, cannot be tested empirically because such assumptions pertain to the counterfactual distribution of $Y(z)$ given $Z = 1 - z$. Robins and others (e.g., see Robins et al., 1999; Scharfstein et al., 1999) have argued that a data analyst should conduct sensitivity analysis to explore how inference varies as a function of departures from any untestable assumptions.

For instance, a departure from assumption (2.5) might be due to the existence of an unmeasured variable U associated with both treatment selection Z and the potential outcomes $Y(z)$ for $z = 0, 1$; a variable such as U is often referred to as an unmeasured confounder. Under this scenario, one might postulate that $Y(z) \amalg Z|U$ for $z = 0, 1$ rather than (2.5). Sensitivity analysis proceeds by examining how inference drawn about ATE varies as a function of the magnitude of the association of U with Z, $Y(0)$, and $Y(1)$. This idea has roots as early as Cornfield et al. (1959), who demonstrated the plausibility of a causal effect of cigarette smoking (Z) on lung cancer (Y) by arguing that the absence of such a relationship was only possible if there existed an unmeasured factor U associated with cigarette use that was at

least as strongly associated with lung cancer as cigarette use. This idea was further developed by Schlesselman (1978); Rosenbaum and Rubin (1983); Lin et al. (1998); Hernán and Robins (1999); and VanderWeele and Arah (2011) among others.

To illustrate this approach, suppose in the AZT example above that the analyst first assumes (2.5) holds and thus estimates the effect of AZT to be -0.48. To proceed with sensitivity analysis, the analyst posits the existence of an unmeasured binary variable U and assumes that $Y(z) \amalg Z|U$ for $z = 0, 1$. Similar to VanderWeele and Arah (2011), let

$$c(z) = \{E[Y(z)|U = 1] - E[Y(z)|U = 0]\}\{\Pr[U = 1|Z = z] - \Pr[U = 1]\}.$$

Then under the assumption that $Y(z) \amalg Z|U$ for $z = 0, 1$, the naive estimator converges in probability to $E[Y(1)] - E[Y(0)] + c(1) - c(0)$. Thus the naive estimator is asymptotically unbiased if and only if $c(1) = c(0)$. For an alternative decomposition of the asymptotic bias of the naive estimator see Morgan and Winship (2007, §2.6.3)

Sensitivity analysis proceeds by making varying assumptions about the unidentifiable associations of U with $Y(0)$, $Y(1)$, and Z. Under the most extreme of these assumptions the bounds (2.2) and (2.3) are recovered. In particular, the upper bound in (2.3) is achieved when $\Pr[U = 1|Z = 1] = 0$, $\Pr[U = 1|Z = 0] = 1$, $E[Y(1)|U = 1] = 1$ and $E[Y(0)|U = 0] = 0$, meaning that the confounder U is perfectly negatively correlated with treatment Z and that if the confounder is present ($U = 1$), then a treated individual will die, whereas if the confounder is absent ($U = 0$), then an untreated individual will survive. The lower bound (2.2) is achieved under the opposite conditions.

In practice the extreme associations of U with $Y(0)$, $Y(1)$, and Z leading to the bounds might be considered unrealistic. Instead the analyst might consider associations only in a range deemed plausible by subject matter experts. In order to arrive at an accurate range, care should be taken in communicating the meaning of these associations and eliciting this range should be done in a manner that avoids data driven choices. Alternatively, the degree of associations required to change the sign of the effect of interest might be determined. For

instance, suppose the analyst further assumes that $E[Y(z)|U = 1] - E[Y(z)|U = 0]$ does not depend on z. This assumption will hold if the effect of Z on Y is the same if $U = 0$ or $U = 1$. Letting $\gamma_0 = E[Y(z)|U = 1] - E[Y(z)|U = 0]$ and $\gamma_1 = \Pr[U = 1|Z = 1] - \Pr[U = 1|Z = 0]$, the asymptotic bias of the naive estimator is then given by $\gamma_0\gamma_1$ and a bias adjusted estimator is found by subtracting $\gamma_0\gamma_1$ from the naive estimator. Sensitivity analysis may proceed by determining the values of γ_0 and γ_1 for which the bias adjusted estimator of the ATE will have the opposite sign of the naive estimator. For the AZT example, the bias adjusted estimator will have the opposite sign of the naive estimator if $\gamma_0\gamma_1 < -0.48$. This indicates that the product of (i) the difference in the mean potential outcomes between levels of the confounder for both treatment and control and (ii) the difference in the prevalence of the unmeasured confounder between the treatment and control groups must be less than -0.48. Such magnitudes might be considered unlikely in the opinion of subject matter experts, in which case the sensitivity analysis would support the existence of a beneficial effect of AZT on survival among HIV+ men (ignoring sampling variability). Note the observed data distribution places some restrictions on the possible values of (γ_0, γ_1), i.e., (γ_0, γ_1) is partially identifiable. For instance, if $\gamma_1 = 1$ then $\Pr[U = 1|Z = 1] = 1$ and $\Pr[U = 1|Z = 0] = 0$ which implies $E[Y(z)|U = u] = E[Y(z)|Z = u]$ and therefore $\max\{E[Y(1)|Z = 1] - 1, -E[Y(0)|Z = 0]\} \leq \gamma_0 \leq \min\{E[Y(1)|Z = 1], 1 - E[Y(0)|Z = 0]\}$. Such considerations should be taken into account when determining the range of values of (γ_0, γ_1) in sensitivity analysis.

Because the data provide no evidence about U, VanderWeele (2008) and VanderWeele and Arah (2011) recommend choosing U and any simplifying assumptions based on what is considered plausible by relevant subject-matter experts. Such sensitivity analyses are most applicable when the existence of unmeasured confounders is known, but these factors could not be measured for logistical or other reasons. General bias formulas to be used for sensitivity analyses of unmeasured confounding for categorical or continuous outcomes, confounders, and treatments can be found in VanderWeele and Arah (2011).

In other settings there might not be any known unmeasured confounders, or it may be thought that there are numerous unmeasured confounders, in which cases the sensitivity

analysis strategy described above would not be applicable or feasible. One general alternative approach entails making additional untestable assumptions regarding the unobserved potential outcome distributions. Typically these assumptions (or models) are indexed by one or more sensitivity analysis parameters conditional upon which the causal estimand of interest is identifiable (e.g., Scharfstein et al., 1999; Brumback et al., 2004). Sensitivity analysis then proceeds by examining how inference changes as assumed values of the parameters are varied over plausible ranges. Examples of such sensitivity analyses are given below in Sections 2.3.4 and 2.6.3.

2.2.5 Covariate Adjustment

Typically in observational studies baseline (pre-treatment) covariates X will be collected in addition to Z and Y. Incorporating information from observed covariates can help sharpen inferences about partially identified treatment effects. For example, incorporating covariates will generally lead to narrower bounds (Scharfstein et al., 1999). This follows because any treatment effect consistent with the distribution of observed variables (X, Y, Z) must also be consistent with the distribution of (Y, Z), i.e., the observable variables if we do not observe or choose to ignore X (Lee, 2009). Covariate adjusted bounds are discussed further in Section 2.3.3 below.

Additionally, incorporating covariates may lend plausibility to some of the bounding assumptions discussed in Section 2.2.2. For example, in the absence of randomized treatment assignment (2.4) or (2.5) may be dubious. Instead of (2.4) it might be more plausible to assume

$$E[Y(z)|Z = 0, X = x] = E[Y(z)|Z = 1, X = x] \text{ for } z = 0, 1. \tag{2.8}$$

Similarly, assumption (2.5) might be replaced by

$$Y(z) \amalg Z|X \text{ for } z = 0, 1, \tag{2.9}$$

i.e., each potential outcome is independent of treatment selection conditional on some set

of covariates. Assumption (2.9) is commonly referred to as no unmeasured confounders. Assumptions such as (2.8) or weaker inequalities similar to (2.6) such as

$$E[Y(z)|Z = 1, X = x] \geq E[Y(z)|Z = 0, X = x] \text{ for } z = 0, 1,$$

may be deemed plausible for certain levels of X, but not for others. Availability of covariates also allows for the consideration of new types of assumptions (e.g., see Chiburis, 2010).

To conduct covariate adjusted sensitivity analysis, departures from identifying assumptions such as (2.9) can be explored. Similar to the previous section, a departure from (2.9) might entail positing the existence of an unmeasured variable U associated with both treatment selection Z and the potential outcomes $Y(z)$ for $z = 0, 1$. Under this scenario, one might postulate that $Y(z) \amalg Z|\{X, U\}$ for $z = 0, 1$ rather than (2.9) and sensitivity analysis proceeds by examining how inference varies as a function of the magnitude of the association of U with Z, $Y(0)$, and $Y(1)$ given X. Similar to covariate adjusted bounds, smaller associations or tighter regions of the values of the sensitivity parameters may be deemed plausible within certain levels of X, potentially yielding sharper inferences from the sensitivity analyses. However, as cautioned by Robins (2002), care should be taken in clearly communicating the meaning of such sensitivity parameters and their relationship to covariates when eliciting plausible ranges from subject matter experts. In some scenarios plausible regions for sensitivity parameters may in fact be wider when conditioning on X than when not conditioning on X.

2.3 Principal Stratification

2.3.1 Background

Even if treatment is randomly assigned (e.g., as in a clinical trial), the causal estimand of interest may still be only partially identifiable. For example, in many studies it is often of interest to draw inference about treatment effects on outcomes that only exist or are meaningful after the occurrence of some observable intermediate variable. For instance, in studies where

some individuals die, investigators might be interested in treatment effects only among individuals alive at the end of the study. Unfortunately, estimands defined by contrasting mean outcomes under treatment and control that simply condition on this observable intermediate variable do not measure a causal effect of treatment without additional assumptions. One approach that may be employed in this scenario entails principal stratification (Frangakis and Rubin, 2002). Principal stratification uses the potential outcomes of the intermediate post-randomization variable to define strata of individuals. Because these "principal strata" are not affected by treatment assignment, treatment effect estimands defined within principal strata have a causal interpretation and do not suffer from the complications of standard post-randomization adjusted estimands. The simple framework of principal stratification has a wide range of applications. For a recent discussion of the utility (and lack thereof) of principal stratification, see Pearl (2011) and corresponding reader reactions.

As a motivating example for this section, we consider evaluating vaccine effects on post-infection outcomes. In vaccine studies, uninfected subjects are enrolled and followed for infection endpoints, and infected subjects are subsequently followed for post-infection outcomes such as disease severity or death due to infection with the pathogen targeted by the vaccine; often interest is in assessing the effect of vaccination on these post-infection endpoints (Hudgens and Halloran, 2006). For example, Preziosi and Halloran (2003) present data from a pertussis vaccine field study in Niakhar, Senegal. In this study 3845 vaccinated children and 1020 unvaccinated children were followed for one year for pertussis. In the vaccine group 548 children contracted pertussis, of whom 176 had severe infections; in the unvaccinated group 206 children contracted pertussis, of whom 129 had severe infections. In this setting investigators are interested in assessing whether or not the vaccine had an effect on the severity of infection.

When assessing such post-infection effects, a data analyst might consider contrasts between study arms including all individuals randomized, or, alternatively, only those who become infected. Though including all individuals in the study has the advantage of providing valid inference about the overall effect of vaccination (assuming perfect compliance), such

an approach does not distinguish vaccine effects on susceptibility to infection from effects on the post-infection endpoint of interest. An analysis that conditions on infection attempts to distinguish these effects and may be more sensitive in detecting post-infection vaccine effects. However, because the set of individuals who would become infected under control are not likely to be the same as those who would become infected if given the vaccine, conditioning on infection might result in selection bias. For example, those who would become infected under vaccine may tend to have weaker immune systems than those who would become infected under control, and thus are more susceptible to severe infection. Because of this potential selection bias, comparisons between infected vaccinees and infected controls do not necessarily have causal interpretations.

2.3.2 Principal Effects

In this section treatment is vaccination, with $Z = 1$ corresponding to vaccination and $Z = 0$ corresponding to not being vaccinated. Assume that assignment to vaccine is equivalent to receipt of vaccine, i.e., there is no non-compliance. Denote the potential infection outcome by $S(z)$, where $S(z) = 0$ if uninfected and $S(z) = 1$ if infected. Here the focus is on evaluating the causal effect of vaccine on Y, a post-infection outcome. For simplicity we consider the case where Y is binary, indicating the presence of severe disease. If $S(z) = 1$, define the potential post-infection outcome $Y(z) = 1$ if the individual would have the worse (or more severe) post-infection outcome of interest given z, and $Y(z) = 0$ otherwise. If an individual's potential infection outcome for treatment z is uninfected, (i.e., $S(z) = 0$), then we adopt the convention that $Y(z)$ is undefined. In other words, it does not make sense to define the severity of an infection in an individual who is not infected. This convention is similar to that employed in other settings. For instance, in the analysis of quality of life studies it might be assumed that quality of life metrics are not well defined in those who are not alive (Rubin, 2000).

Define a *basic principal stratification* P_0 according to the joint potential infection outcomes $S^{P_0} = (S(0), S(1))$. The four basic principal strata or response types are defined by the

joint potential infection outcomes, $(S(0), S(1))$, and are composed of immune (not infected under both vaccine and placebo), harmed (infected under vaccine but not placebo), protected (infected under placebo but not vaccine), and doomed individuals (infected under both vaccine and placebo). Note the only stratum where both potential post-infection endpoints are well defined is in the doomed basic principal stratum, $S^{P_0} = (1, 1)$. Thus defining a post-infection causal vaccine effect is only possible in the doomed principal stratum $S^{P_0} = (1, 1)$. Such a causal estimand will describe the effect of vaccination on disease severity in individuals who would become infected whether vaccinated or not. For instance, the vaccine effect on disease severity may be defined by

$$E[Y(1)|S^{P_0} = (1, 1)] - E[Y(0)|S^{P_0} = (1, 1)]. \tag{2.10}$$

Frangakis and Rubin call treatment effect estimands such as (2.10) "principal effects."

2.3.3 Bounds

Assume we observe n iid copies of (Z, S, Y) denoted by (Z_i, S_i, Y_i) for $i = 1, \ldots, n$. Also assume that the doomed principal strata is non-empty, $\Pr[S^{P_0} = (1, 1)] > 0$, so that the principal effect in (2.10) is well defined. Bounds for (2.10) are presented below under two additional assumptions: independent treatment assignment, i.e.,

$$Z \amalg \{Y(z), S(z)\} \text{ for } z = 0, 1 \tag{2.11}$$

and monotone treatment response with respect to S, i.e.,

$$\Pr[S(0) \geq S(1)] = 1. \tag{2.12}$$

Assumption (2.11) will hold in randomized vaccine trials. Monotonicity (2.12) assumes that the vaccine does no harm at the individual level, i.e., there are no individuals who would be infected if vaccinated but uninfected if not vaccinated. Monotonicity is equivalent to assuming the harmed principal stratum is empty. Note no such monotonicity assumption is being made

regarding Y. Under (2.11), assumption (2.12) implies $P(S = 1|Z = 1) \leq P(S = 1|Z = 0)$, which is testable using the observed data. For the pertussis example, the proportion infected in the vaccine group was less than in the unvaccinated group; thus, assuming (2.11), the data do not provide evidence against (2.12).

Assuming independent treatment assignment and monotonicity, (2.10) is partially identifiable from the observable data. The left term of (2.10) can be written

$$\begin{aligned} E[Y(1)|S^{P_0} = (1,1)] &= E[Y(1)|S(1) = 1] \\ &= E[Y(1)|S(1) = 1, Z = 1] \\ &= E[Y|S = 1, Z = 1], \end{aligned} \tag{2.13}$$

where the first equality holds under (2.12), the second equality under (2.11), and the third by causal consistency. On the other hand, the right term of (2.10) is only partially identifiable. To see this, note

$$\begin{aligned} E[Y(0)|S(0) = 1] = \quad & E[Y(0)|S^{P_0} = (1,1)] \Pr[S(1) = 1|S(0) = 1] + \\ & E[Y(0)|S^{P_0} = (1,0)] \Pr[S(1) = 0|S(0) = 1]. \end{aligned} \tag{2.14}$$

In (2.14), only $E[Y(0)|S(0) = 1]$ and $\Pr[S(1) = s|S(0) = 1]$ for $s = 0, 1$ are identifiable. In particular, $E[Y(0)|S(0) = 1] = E[Y|S = 1, Z = 0]$ by similar reasoning to (2.13), and

$$\Pr[S(1) = 1|S(0) = 1] = \frac{\Pr[S(1) = 1]}{\Pr[S(0) = 1]} = \frac{\Pr[S = 1|Z = 1]}{\Pr[S = 1|Z = 0]},$$

where the first equality holds under (2.12) and the second under independent treatment assignment (and causal consistency). The other two terms in (2.14), namely $E[Y(0)|S^{P_0} = (1,1)]$ and $E[Y(0)|S^{P_0} = (1,0)]$, are only partially identifiable. In words, infected controls are a mixture of individuals in the protected and doomed principal stratum and without further assumptions the observed data do not identify exactly which infected controls are doomed. Therefore the probability of severe disease when not vaccinated in the doomed principal stratum is not identified. Under (2.12), the data do however indicate what proportion of infected controls are doomed and this information provides partial identification of

$E[Y(0)|S^{P_0} = (1,1)]$ and hence (2.10).

For fixed values of $E[Y(0)|S(0) = 1]$ and $\Pr[S(1) = 1|S(0) = 1]$, any pair of expectations $(E[Y(0)|S^{P_0} = (1,1)], E[Y(0)|S^{P_0} = (1,0)]) \in [0,1]^2$ satisfying (2.14) will give rise to the same observed data distribution. Equation (2.14) describes a line segment with non-positive slope intersecting the unit square as illustrated in Figure 1. An upper bound of $E[Y(0)|S^{P_0} = (1,1)]$ and thus a lower bound for (2.10) is achieved when the line intersects the right or lower side of the square, i.e., when either

$$E[Y(0)|S^{P_0} = (1,1)] = 1 \text{ or } E[Y(0)|S^{P_0} = (1,0)] = 0. \tag{2.15}$$

Together (2.14) and (2.15) imply $E[Y(0)|S^{P_0} = (1,1)]$ is bounded above by

$$\min\left\{1, \frac{E[Y(0)|S(0) = 1]}{\Pr[S(1) = 1|S(0) = 1]}\right\}. \tag{2.16}$$

Similarly, $E[Y(0)|S^{P_0} = (1,1)]$ is bounded below by

$$\max\left\{0, \frac{E[Y(0)|S(0) = 1] - \Pr[S(1) = 0|S(0) = 1]}{\Pr[S(1) = 1|S(0) = 1]}\right\}. \tag{2.17}$$

Combining (2.17) with (2.13) yields the upper bound on the principal effect of interest (2.10) and combining (2.16) with (2.13) yields the lower bound. These bounds were derived by Rotnitzky and Jemiai (2003); Zhang and Rubin (2003); and Hudgens et al. (2003). Consistent estimates of (2.16) and (2.17) can be computed by replacing $E[Y(0)|S(0) = 1]$ with $\sum_i Y_i I(S_i = 1, Z_i = 0)/\sum_i I(S_i = 1, Z_i = 0)$ and $\Pr[S(1) = 1|S(0) = 1]$ with

$$\min\left\{1, \frac{\sum_i I(S_i = Z_i = 1)/\sum_i I(Z_i = 1)}{\sum_i I(S_i = 1, Z_i = 0)/\sum_i I(Z_i = 0)}\right\}.$$

Returning to the pertussis vaccine study, the estimated lower and upper bounds of (2.10) are -0.57 and -0.15. These estimated bounds exclude zero, leading to the conclusion (ignoring sampling variability) that vaccination lowers the risk of severe pertussis in individuals who will become infected regardless of whether they are vaccinated.

Note if $\Pr[S(1) = 1|S(0) = 1] = 1$, i.e., the vaccine has no protective effect against infection, then the protected principal stratum $S^{P_0} = (1, 0)$ is empty and both (2.16) and (2.17) equal $E[Y(0)|S(0) = 1]$ meaning that (2.10) is identifiable and equals $E[Y|Z = 1, S = 1] - E[Y|Z = 0, S = 1]$. Intuitively the lack of vaccine effect against infection eliminates the potential for selection bias.

As discussed in Section 2.2.5, incorporation of covariates can tighten bounds. For covariates X with finite support, one simple approach of adjusting for covariates entails determining bounds within strata defined by the levels of X and then taking a weighted average of the within strata bounds over the distribution of X. For the bounds in (2.16) and (2.17), adjustment for covariates will always lead to bounds that are at least as tight as bounds unadjusted for covariates (Lee, 2009; Long and Hudgens, 2013).

If the observed data provide evidence contrary to monotonicity (2.12), then bounds may be obtained under only (2.11). Without monotonicity (2.12) the proportion of infected controls that are in the doomed principal stratum is no longer identified but may be bounded in order to arrive at bounds for $E[Y(0)|S^{P_0} = (1, 1)]$. In addition, the harmed principal stratum defined by $S^{P_0} = (0, 1)$ is no longer empty and thus $E[Y(1)|S^{P_0} = (1, 1)]$ is no longer identifiable from the observed data and may also be bounded in a similar fashion to $E[Y(0)|S^{P_0} = (1, 1)]$. Details regarding these bounds without the monotonicity assumption may be found in Zhang and Rubin (2003) and Grilli and Mealli (2008).

2.3.4 Sensitivity Analysis

The bounds (2.16) and (2.17) are useful in bounding the vaccine effect on Y in the doomed stratum. However, these bounds may be rather extreme. An alternative approach is to make an untestable assumption that identifies the post-infection vaccine effect on Y and then consider how sensitive the resulting inference is to departures from this assumption. For instance, assuming

$$\Pr[Y(0) = 1|S^{P_0} = (1, 1)] = \Pr[Y(0) = 1|S^{P_0} = (1, 0)], \tag{2.18}$$

identifies (2.10). Hudgens and Halloran (2006) refer to this as the no selection model. To examine how inference varies according to departures from (2.18), following Scharfstein et al. (1999), and Robins et al. (1999), consider the following sensitivity parameter

$$\exp(\gamma) = \frac{\Pr[Y(0) = 1 | S^{P_0} = (1,1)] / \Pr[Y(0) = 0 | S^{P_0} = (1,1)]}{\Pr[Y(0) = 1 | S^{P_0} = (1,0)] / \Pr[Y(0) = 0 | S^{P_0} = (1,0)]}. \tag{2.19}$$

In words, $\exp(\gamma)$ compares the odds of severe disease when not vaccinated in the doomed versus the protected principal stratum. Assuming (2.18) corresponds to $\gamma = 0$. A sensitivity analysis entails examining how inference about (2.10) varies as γ becomes farther from 0. For any fixed value of γ, (2.10) is identified (see Figure 1) and can be consistently estimated by maximum likelihood estimation without any additional assumptions (Gilbert et al., 2003). The lower and upper bounds (2.17) and (2.16) are obtained by letting $\gamma \to \infty$ and $\gamma \to -\infty$. To see this, note that as $\gamma \to \infty$ (2.19) implies in the limit that either

$$\Pr[Y(0) = 1 | S^{P_0} = (1,1)] = 1 \text{ or } \Pr[Y(0) = 1 | S^{P_0} = (1,0)] = 0,$$

which is equivalent to (2.15). Sensitivity analysis can be conducted by letting γ range over a set of values Γ.

Tighter bounds can be achieved by placing restrictions on Γ, perhaps based on prior beliefs about γ elicited from subject matter experts. For example, Shepherd et al. (2007) surveyed 10 recognized HIV experts in order to elicit a plausible range for a sensitivity parameter representing a departure from the assumption of no selection bias between vaccinated and unvaccinated individuals who acquired HIV during an HIV vaccine trial. Included in this survey was the analysis approach, a brief explanation of the potential for selection bias, the definition of the sensitivity parameter being employed, examples of the implications of certain sensitivity parameter values on selection bias, and possible justification for believing certain values of the sensitivity parameter. The expert responses to the survey were fairly consistent and several written justifications for the respondents' chosen ranges indicated a high level of understanding of both the counterfactual nature of the sensitivity parameter and the need to

account for selection bias.

2.4 Randomized Studies with Partial Compliance

2.4.1 Global Average Treatment Effect

In a placebo controlled randomized trial where (2.5) holds but there is non-compliance (i.e., individuals are randomly assigned to treatment or control but they do not necessarily adhere or comply with their assigned treatment), the naive estimator is a consistent estimator of the average effect of treatment *assignment.* However, in this case parameters other than the effect of treatment assignment may be of interest. As in the last section, a principal effect may be defined using compliance as the intermediate post-randomization variable over which to define principal strata; namely the principal strata would consist of individuals who would comply with their randomization assignment if assigned treatment or control or "compliers," individuals who would always take treatment regardless of randomization or "always takers," individuals who never take treatment "never takers," and individuals who take treatment only if assigned control or "defiers." A principal effect of interest might be the effect of treatment in the complier principal stratum (Imbens and Angrist, 1994; Angrist et al., 1996), in which case bounds and sensitivity analyses similar to those in Section 2.3 are applicable. However, as several authors including Robins (1989) and Robins and Greenland (1996) have pointed out, such principal effects may not be of ultimate public health interest because they only apply to the subpopulation of compliers in clinical trials, which may differ from the population that elect to take treatment once licensed. For example, once efficacy is proved, a larger subpopulation of people may be willing to take the treatment. Effects defined on the subpopulation of compliers are also of limited decision-making utility because individual principal stratum membership is generally unknown prior to treatment assignment (Joffe, 2011).

Robins and Greenland (1996) suggested that in settings where the trial population could be persuaded to take the treatment once licensed, a more relevant public health estimand is

the global average treatment effect, defined as the average effect of actually taking treatment versus not taking treatment given treatment assignment z. This causal estimand is similar to the average treatment effect defined in Section 2.2, but requires generalizing the potential outcome definitions used previously to include separate potential outcomes for each of the four combinations of treatment assignment and actual treatment received. For further discussion regarding causal models in presence of noncompliance see Chickering and Pearl (1996) and Dawid (2003) among others.

Suppose we observe data from a clinical trial where each individual is randomly assigned to treatment or control. Let Z indicate treatment assignment where $Z = 1$ denotes treatment and $Z = 0$ denotes control. Suppose individuals do not necessarily comply with their randomization assignment and let S be a variable indicating whether or not treatment was actually taken, where $S = 1$ denotes treatment was taken and $S = 0$ otherwise. Thus an individual is compliant with their randomization assignment if $S = Z$. Let Y be a binary outcome of interest. Denote the potential treatment taken by $S(z)$ for $z = 0, 1$, where $S(z) = 1$ indicates taking treatment when assigned treatment z and $S(z) = 0$ denotes not taking treatment when assigned z. Let $Y(z, s)$ denote the potential outcome if an individual is assigned treatment z but actually takes treatment s. Conceiving of these potential outcomes depends on a supposition that trial participants who did not comply in the trial could be persuaded to take the treatment under other circumstances. Given this supposition, the global average treatment effect for each treatment assignment $z = 1$ and $z = 0$ is defined as $\text{GATE}_z = E[Y(z, 1) - Y(z, 0)]$. For instance, GATE_1 is the difference in the average outcomes under the counterfactual scenario everyone was assigned vaccine and did comply versus the counterfactual scenario everyone was assigned vaccine but did not comply.

Bounds for GATE_z are given below under three assumptions: independent treatment assignment

$$Z \amalg \{S(0), S(1), Y(0, 0), Y(0, 1), Y(1, 0), Y(1, 1)\}; \tag{2.20}$$

monotonicity with respect to S

$$\Pr[S(1) \geq S(0)] = 1; \tag{2.21}$$

and the exclusion restriction

$$Y(0,s) = Y(1,s) \text{ for } s = 0,1. \tag{2.22}$$

Assumption (2.22) indicates treatment assignment has no effect when the actual treatment taken is held fixed. Under (2.22), $\text{GATE}_0 = \text{GATE}_1$ which we denote by GATE. In this case each individual has two potential outcomes according to $s = 0$ and $s = 1$ (which could be denoted by $Y(s) = Y(0,s) = Y(1,s)$ for $s = 0,1$) and GATE is equivalent to the ATE discussed in Section 2.2 with z replaced by s. Robins (1989) derived bounds for GATE under several different combinations of (2.20) – (2.22) as well as some additional assumptions such as monotonicity with respect to S, i.e., $Y(z,1) \geq Y(z,0)$ for $z = 0,1$. Manski (1990) independently derived related results. Under (2.20) – (2.22) the sharp lower and upper bounds on GATE are

$$-1 + \max_z\{\Pr[Y = 1, S = 1|Z = z]\} + \max_z\{\Pr[Y = 0, S = 0|Z = z]\}, \tag{2.23}$$

and

$$1 - \max_z\{\Pr[Y = 0, S = 1|Z = z]\} - \max_z\{\Pr[Y = 1, S = 0|Z = z]\}. \tag{2.24}$$

Balke and Pearl (1997) derived sharp bounds for GATE under a variety of assumptions, including (2.20) – (2.22), by recognizing that the derivation of the bounds is equivalent to a linear programming optimization problem. To see that bounds can be formulated as a linear programming optimization problem, first note that GATE can be expressed as a linear combination of probabilities of the joint distribution of $L = (Y(0,0), Y(0,1), Y(1,0), Y(1,1), S(0), S(1))$

$$\sum_{l_1 \in \mathcal{L}_1} \Pr[L = l_1] - \sum_{l_0 \in \mathcal{L}_0} \Pr[L = l_0] \tag{2.25}$$

where $\mathcal{L}_s$ is the set of possible realizations of L where $Y(0,s) = Y(1,s) = 1$ for $s = 0,1$. Under independent treatment assignment, there exists a linear transformation between the probabilities in the joint distribution of L and the probabilities in the conditional distribution

of the observable random variables Y and S given Z, namely

$$\Pr[Y = y, S = s|Z = z] = \sum_{l \in \mathcal{O}_{ys \cdot z}} \Pr[L = l] \tag{2.26}$$

where $\mathcal{O}_{ys \cdot z}$ is the set of possible realizations of L where $S(z) = s$ and $Y(z, s) = y$ for $z, y, s = 0, 1$. To find the sharp bounds, the objective function (2.25) is minimized (or maximized) subject to the constraints (2.26), $\Pr[L = l] \geq 0$ for every $l \in \mathcal{L}$, and $\sum_{l \in \mathcal{L}} \Pr[L = l] = 1$ where $\mathcal{L}$ is the set of all possible realizations of L assuming (2.21) and (2.22). Optimization may be accomplished using the simplex algorithm and the dimension of this problem permits obtaining a closed form solution involving probabilities of the observed data distribution (Balke and Pearl, 1993), namely (2.23) and (2.24).

If in addition to assumptions (2.20) and (2.22), it is assumed that

$$E[Y(z, 1) - Y(z, 0)|Z = 1, S = s] = E[Y(z, 1) - Y(z, 0)|Z = 0, S = s] \tag{2.27}$$

for $s, z = 0, 1$ then GATE is identified and equals

$$\frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[S|Z = 1] - E[S|Z = 0]} \tag{2.28}$$

(Hernán and Robins, 2006). For $s = 0$ assumption (2.27) is known as a no current treatment interaction assumption (Robins, 1994), and expression (2.28) is known as the instrumental variables estimand (Imbens and Angrist, 1994; Angrist et al., 1996). Sensitivity analyses may be conducted by defining sensitivity parameters representing departures from (2.20), (2.22) or (2.27) and then examining how inference about GATE varies as values of these parameters change. For instance, Robins et al. (1999) define current treatment interaction functions which represent a departure from (2.27) for $s = 0$.

2.4.2 Cholestyramine Example

To illustrate the GATE, we consider data presented in Pearl (2009, §8.2.6) on 337 subjects who participated in a randomized trial to assess the effect of cholestyramine on cholesterol reduction. Let $Z = 1$ denote assignment to cholestyramine and $Z = 0$ assignment to placebo. Let $S = 1$ if cholestyramine was actually taken by the participant and $S = 0$ otherwise. Let $Y = 1$ if the participant had a response and $Y = 0$ otherwise, where response is defined as reduction in the level of cholesterol by 28 units or more. Pearl reported the following observed proportions

$$\begin{aligned}
&\hat{\Pr}[Y = 0, S = 0|Z = 0] = 0.919 \quad \hat{\Pr}[Y = 0, S = 0|Z = 1] = 0.315 \\
&\hat{\Pr}[Y = 0, S = 1|Z = 0] = 0.000 \quad \hat{\Pr}[Y = 0, S = 1|Z = 1] = 0.139 \\
&\hat{\Pr}[Y = 1, S = 0|Z = 0] = 0.081 \quad \hat{\Pr}[Y = 1, S = 0|Z = 1] = 0.073 \\
&\hat{\Pr}[Y = 1, S = 1|Z = 0] = 0.000 \quad \hat{\Pr}[Y = 1, S = 1|Z = 1] = 0.473
\end{aligned}$$

No participants assigned placebo actually took cholestyramine, suggesting the monotonicity assumption (2.21) is reasonable. On the other hand, 38.8% of individuals assigned treatment did not actually take cholestyramine.

From (2.23) and (2.24) the bounds on GATE assuming (2.21), (2.20) and (2.22) are estimated to be $-1 + \max\{0.000, 0.473\} + \max\{0.919, 0.315\} = 0.392$ and $1 - \max\{0, 0.139\} - \max\{0.081, 0.073\} = 0.780$. The positive sign of the estimated bounds indicates the treatment is beneficial. Pearl interprets the estimated bounds as follows: "although 38.8% of the subjects deviated from their treatment protocol, the experimenter can categorically state that, when applied uniformly to the population, the treatment is guaranteed to increase by at least 39.2% the probability of reducing the level of cholesterol by 28 points or more." Such an interpretation does not account for sampling variability, the topic of Section 2.7.

2.5 Mediation Analysis

2.5.1 Natural Direct and Indirect Effects

As demonstrated in Sections 2.3 and 2.4, independent treatment assignment does not guarantee that the causal estimand of interest will be identifiable. Another setting where this occurs is in mediation analysis, where researchers are interested in whether or not the effect of a treatment is mediated by some intermediate variable. Even in studies where treatment is assigned randomly and there is perfect compliance, confounding may exist between the intermediate variable and the outcome of interest such that effects describing the mediated relationships will not in general be identifiable. Thus bounds and sensitivity analysis may be helpful in drawing inference.

To illustrate, let Y be an observed binary outcome of interest, and S a binary intermediate variable observed some time between treatment assignment Z and the observation of Y. The goal is to assess whether and to what extent the effect of Z on Y is mediated by or through S. Denote the potential outcome of the intermediate variable under treatment z by $S(z)$ for $z = 0, 1$ such that $S = S(Z)$, and the potential outcomes under treatment z and intermediate s as $Y(z, s)$ such that $Y = Y(Z, S(Z))$. Here, as in the previous section, it is assumed that both Z and S can be set to particular fixed values, such that there are four potential outcomes for Y per individual. Unless otherwise specified, independent treatment assignment (2.20) will be assumed throughout this section.

Define the total effect of treatment to be $E[Y(1, S(1)) - Y(0, S(0))]$, which is equivalent to the ATE defined in Section 2.2.1. The total effect of treatment can be decomposed in the following way

$$\begin{aligned} E[Y(1, S(1)) - Y(0, S(0))] = E[Y(1, S(z)) - Y(0, S(z))] \\ + E[Y(z', S(1)) - Y(z', S(0)] \end{aligned} \tag{2.29}$$

for $z = 0, 1$ and $z' = 1 - z$. The right side of (2.29) decomposes the total effect into the sum of

two separate effects. The first expectation on the right side of (2.29) is the natural direct effect for treatment z, $\text{NDE}_z = E[Y(1, S(z)) - Y(0, S(z))]$ (Robins and Greenland, 1992; Pearl, 2001; Robins, 2003; Kaufman et al., 2009; Robins and Richardson, 2010). The natural direct effect is the average effect of the treatment on the outcome when the intermediate variable is set to the potential value that would occur under treatment assignment z. The second expectation on the right side of (2.29) is the natural indirect effect, $\text{NIE}_z = E[Y(z, S(1)) - Y(z, S(0))]$ (Pearl, 2001; Robins, 2003; Imai et al., 2010). The natural indirect effect is the difference in the average outcomes when treatment is set to z and the intermediate variable is set to the value that would have occurred under treatment compared to if the intermediate variable were set to the value that would have occurred under control.

Though the total effect is identifiable assuming (2.20), the natural direct and indirect effects are not identifiable since they entail $E[Y(z, S(1-z))]$ which depends on unobserved counterfactual distributions. Sjölander (2009) derived bounds for the natural direct effects assuming only independent treatment assignment (2.20) using the linear programming technique of Balke and Pearl (1997). This results in the following sharp lower and upper bounds for NDE_0 and NDE_1

$$\max\left\{\begin{array}{c} -p_{11\cdot 0} - p_{10\cdot 0}, \\ p_{11\cdot 1} + p_{01\cdot 0} - 1 - p_{10\cdot 0}, \\ p_{10\cdot 1} + p_{00\cdot 0} - 1 - p_{11\cdot 0} \end{array}\right\} \leq \text{NDE}_0 \leq \min\left\{\begin{array}{c} p_{01\cdot 0} + p_{00\cdot 0}, \\ 1 - p_{00\cdot 1} + p_{01\cdot 0} - p_{10\cdot 0}, \\ 1 - p_{01\cdot 1} + p_{00\cdot 0} - p_{11\cdot 0} \end{array}\right\} \tag{2.30}$$

$$\max\left\{\begin{array}{c} -p_{01\cdot 1} - p_{00\cdot 1}, \\ p_{00\cdot 0} - 1 - p_{01\cdot 1} + p_{10\cdot 1}, \\ p_{01\cdot 0} - 1 - p_{00\cdot 1} + p_{11\cdot 1} \end{array}\right\} \leq \text{NDE}_1 \leq \min\left\{\begin{array}{c} p_{11\cdot 1} + p_{10\cdot 1}, \\ 1 - p_{01\cdot 1} + p_{10\cdot 1} - p_{11\cdot 0}, \\ 1 - p_{00\cdot 1} + p_{11\cdot 1} - p_{10\cdot 0} \end{array}\right\} \tag{2.31}$$

where $p_{ys\cdot z} = \Pr(Y = y, S = s | Z = z)$. These bounds may exclude 0, indicating a natural direct effect of treatment z when the intermediate variable is set to $S(z)$ (ignoring sampling variability). There are instances where the bounds in (2.30) and (2.31) may collapse to a single point, e.g., if $p_{10\cdot 0} = p_{10\cdot 1} = 1$. Using (2.29), bounds for NIE_0 and NIE_1 can be obtained by

subtracting the bounds for NDE_1 and NDE_0 from the total effect, which is identified under (2.20) and equal to $(p_{11\cdot1} + p_{10\cdot1}) - (p_{10\cdot0} - p_{11\cdot0})$.

Just as in Sections 2.2–2.4, monotonicity assumptions can be made to tighten the above bounds. For instance, if

$$\begin{aligned}
&\Pr[S(0) \leq S(1)] = 1 \\
&\Pr[Y(0,s) \leq Y(1,s)] = 1 \text{ for } s = 0, 1 \text{ and} \\
&\Pr[Y(z,0) \leq Y(z,1)] = 1 \text{ for } z = 0, 1,
\end{aligned}$$

are assumed, then $\Pr[L = l] = 0$ for all l such that (i) $S(0) = 1$ and $S(1) = 0$, (ii) $Y(0,s) = 1$ and $Y(1,s) = 0$ for $s = 0$ or 1, or (iii) $Y(z,0) = 1$ and $Y(z,1) = 0$ for $s = 0$ or 1, which restricts the feasible region of the linear programming problem. The resulting sharp bounds for the natural direct effect are

$$\max \left\{ \begin{array}{c} 0,\, p_{01\cdot0} - p_{01\cdot1},\, p_{10\cdot1} - p_{10\cdot0}, \\ p_{01\cdot0} - p_{01\cdot1} + p_{10\cdot1} - p_{10\cdot0} \end{array} \right\} \leq \text{NDE}_z \leq p_{10\cdot1} + p_{11\cdot1} - p_{10\cdot0} - p_{11\cdot0} \tag{2.32}$$

(Sjölander, 2009). The bounds (2.32) are always at least as narrow as (2.30) and (2.31). Interestingly these narrower bounds do not depend on z. The bounds in (2.32) may also collapse to a single point, e.g., if $p_{10\cdot0} = p_{10\cdot1}$ and $p_{01\cdot0} - p_{01\cdot1} = p_{11\cdot1} - p_{11\cdot0}$.

The natural direct effect provides insight into whether or not treatment yields additional benefit on the outcome of interest when the influence of treatment on the intermediate variable is eliminated. However, researchers might also be interested in what benefit is provided by treatment if the effect of the intermediate variable on the outcome is eliminated or held constant. This question suggests a different causal estimand known as the controlled direct effect. Bounds for the controlled direct effect can be found in Pearl (2001); Cai et al. (2008); Sjölander (2009); and VanderWeele (2011).

2.5.2 Sensitivity Analysis

As in other settings where the effect of interest is not identifiable, sensitivity analysis in the mediation setting may be conducted by making untestable assumptions that identify the direct or indirect effects. Then sensitivity of inference to departures from these assumptions can be examined. For example, if (2.20) holds, then the natural direct and indirect effects are identified under the following additional assumptions

$$Y(z,s) \amalg S|Z \text{ for } z,s=0,1 \text{ and} \tag{2.33}$$

$$Y(z,s) \amalg S(z') \text{ for } z,z',s=0,1 \tag{2.34}$$

(Pearl, 2001; VanderWeele, 2010). Assumption (2.33) would be valid if subjects were randomly assigned S within different levels of treatment assignment Z. In settings where S is not randomly assigned, (2.33) might be considered plausible if it is believed that conditional on Z there are no variables which confound the mediator–outcome relationship. Both assumptions (2.33) and (2.34) will not hold in general if Z has an effect on some other intermediate variable, say R, which in turn has an effect on both S and Y. Thus (2.33) and (2.34) may fail unless the mediator S occurs shortly after treatment Z. Under assumptions (2.20), (2.33) and (2.34),

$$\text{NDE}_z = (-1)^z \sum_s \{E[Y|Z=1-z,S=s] - E[Y|Z=z,S=s]\} \Pr[S=s|Z=z]$$

and

$$\text{NIE}_z = (-1)^z \sum_s E[Y|Z=z,S=s]\{\Pr[S=s|Z=1-z] - \Pr[S=s|Z=z]\}.$$

Because assumptions (2.33) and (2.34) cannot be empirically tested, sensitivity analysis should be conducted. Similar to Section 2.2.4, sensitivity analysis might proceed by positing the existence of an unmeasured confounding variable U associated with the potential mediator values $S(z)$ and the potential outcomes $Y(z,s)$ for $z,s=0,1$. Assumption (2.33) would then replaced by $Y(z,s) \amalg S|\{Z,U\}$ and (2.34) by $Y(z,s) \amalg S(z')|U$ for $s,z,z'=0,1$. Sensitivity

analysis would then proceed by exploring how inference about the natural direct and indirect effects changes as the magnitude of the associations of U with $S(z)$ and $Y(s, z')$ for $z, z', s = 0, 1$ vary. For further details regarding bounds and sensitivity analysis in mediation analysis see Imai et al. (2010); VanderWeele (2010); and Hafeman (2011).

2.6 Longitudinal Treatment

2.6.1 Background

In Sections 2.2–2.5 treatment is assumed to remain fixed across follow up time and outcomes are one dimensional. However, frequently researchers are interested in assessing causal effects comparing longitudinal outcomes for patients on different treatment regimens where treatment may vary in time. As the number of times at which an individual may receive treatment increases, the number of possible treatment regimens increases exponentially. Because each treatment regimen corresponds to a separate potential (longitudinal) outcome and only one potential outcome is ever observed, the fraction of potential outcomes that are unobserved quickly grows close to one as the number of possible treatment times increases. As in other settings, unless treatment regimens are randomly assigned, regimen effects will not be identifiable without additional assumptions. In the longitudinal setting bounds will typically be largely uninformative because of the large proportion of unobserved potential outcomes. Therefore analyses usually proceed by invoking modeling assumptions that render treatment effects identifiable and then conducting sensitivity analysis corresponding to key untestable modeling assumptions.

Models for potential outcomes as functions of covariates (such as treatment) and possibly other potential outcomes are often referred to as structural models. For longitudinal potential outcomes and treatments, popular models include structural nested models and marginal structural models (Robins et al., 1999; Robins, 1999; van der Laan and Robins, 2003; Brumback et al., 2004). In Section 2.6.2 below we consider a marginal structural model where the treatment effect is identified assuming conditionally independent treatment assignment.

Sensitivity analyses exploring departures from this assumption are then considered in Section 2.6.3.

2.6.2 Marginal Structural Model

Consider a study where individuals possibly receive treatment at τ fixed time points (i.e., study visits). In general let $\bar{A}(t) = (A(0), \ldots, A(t))$ represent the history of variable A up to time t and $\bar{A}$ be the entire history of variable A such that $\bar{A} = \bar{A}(\tau)$. Let $z(t) = 1$ indicate treatment at visit t, and $z(t) = 0$ otherwise such that $\bar{z}$ represents a treatment regimen for visits $0, \ldots, \tau$. Denote the observed treatment regimen up to time t as $\bar{Z}(t)$. Let Y be some outcome of interest that may be categorical or continuous, and denote the potential outcome of Y at visit t for regimen $\bar{z}$ by $Y(\bar{z}, t)$ and the observed outcome by $Y(t)$. Let $\bar{X}(t)$ denote the history of some set of time varying covariates up to time t, where $X(0)$ denotes the baseline covariates. Assume for simplicity there is no loss to follow-up or non-compliance such that we observe n iid copies of $(\bar{Z}, \bar{Y}, \bar{X})$.

Consider the following marginal structural model of the mean potential outcome were the entire population to follow regimen $\bar{z}$ up to time t

$$g(E[Y(\bar{z}, t)|X(0) = x(0)]) = \beta_0 + \beta_1 \text{cum}[\bar{z}(t-1)] + \beta_2 t + \beta_3 x(0) \tag{2.35}$$

for $t \in \{1, \ldots, \tau\}$, where $\text{cum}[\bar{z}(t-1)] = \sum_{k=1}^{t-1} z(k)$ and $g(\cdot)$ is an appropriate link function. The causal estimand of interest is β_1, the regression coefficient for $\text{cum}[\bar{z}(t-1)]$, which is the effect of having received treatment at one additional visit prior to time t conditional on baseline covariates $X(0)$. Because (2.35) involves counterfactual outcome distributions, β_1 is not identifiable without additional assumptions. One additional assumption is conditionally independent treatment assignment

$$Y(\bar{z}, t) \amalg Z(k)|\{\bar{Z}(k-1), \bar{X}(k)\} \text{ for all } \bar{z} \text{ and } t > k \tag{2.36}$$

(Robins et al., 1999; Robins, 1999; Brumback et al., 2004). This assumption is true if the

potential outcome at visit t under treatment regimen $\bar{z}$ is independent of the observed treatment at visit k given the history of treatment up to visit $k-1$ and the covariate history up to visit k. Assuming both a correctly specified model (2.35) and conditionally independent treatment assignment (2.36), fitting the following model to the observed data

$$g(E[Y(t)|\bar{Z}(t-1)=\bar{z}(t-1), X(0)=x(0)]) = \eta_0 + \eta_1 \text{cum}[\bar{z}(t-1)] + \eta_2 t + \eta_3 x(0),$$

using generalized estimating equations with an independent working correlation matrix and time varying inverse probability of treatment weights (IPTW) yields an estimator $\hat{\eta}_1$ that is consistent for β_1 (Tchetgen Tchetgen et al., 2012a,b).

2.6.3 Sensitivity Analysis

If assumption (2.36) does not hold, then the IPTW estimator $\hat{\eta}_1$ is not necessarily consistent. Because (2.36) is not testable from the observed data, sensitivity analysis might be considered to assess robustness of inference to departures from (2.36). Following Robins (1999) and Brumback et al. (2004), let

$$\begin{aligned} c(t,k,\bar{z}(t-1),\bar{x}(k)) = E[Y(\bar{z},t)|\bar{Z}(k)=\bar{z}(k),\bar{X}(k)=\bar{x}(k)]- \\ E[Y(\bar{z},t)|Z(k)=1-z(k),\bar{Z}(k-1)=\bar{z}(k-1),\bar{X}(k)=\bar{x}(k)] \end{aligned}$$

for $t > k$ and $\bar{z}$ such that $\Pr[Z(k)=z(k)|\bar{Z}(k-1)=\bar{z}(k-1)]$ is bounded away from 0 and 1. The function c quantifies departures from the conditional independent treatment assignment assumption (2.36) at each visit $t > k$, where $c(t,k,\bar{z}(t-1),\bar{x}(k)) = 0$ for all $\bar{z}$ and $t > k$ if (2.36) holds. For the identity link, a bias adjusted estimator of the causal effect β_1 may be obtained by recalculating the IPTW estimator with the observed outcome $Y(t)$ replaced by $Y^{\gamma}(t) = Y(t) - b(\bar{Z}(t-1),\bar{X}(t-1))$ where

$$b(\bar{Z}(t-1),\bar{X}(t-1)) = \sum_{k=0}^{t-1} c(t,k,\bar{Z}(t-1),\bar{X}(k)) f[1-Z(k)|\bar{Z}(k-1),\bar{X}(k)]$$

and $f[z(k)|\bar{z}(k-1), \bar{x}(k)] = \hat{\Pr}[Z(k) = z(k)|\bar{Z}(k-1) = \bar{z}(k-1), \bar{X}(k) = \bar{x}(k)]$ is an estimate of the conditional probability of the observed treatment based on some fitted parametric model (Brumback et al., 2004). Provided this parametric model and c are both correctly specified, this bias adjusted estimator, say $\tilde{\eta}_1$, is consistent for β_1. Sensitivity analysis proceeds by examining how $\tilde{\eta}_1$ changes when varying sensitivity parameters in $c(t, k, \bar{z}(t-1), \bar{x}(k))$.

Because $c(t, k, \bar{z}(t-1), \bar{x}(k))$ is not identifiable from the observable data, Robins (1999) recommends choosing a particular c that is easily explainable to subject matter experts to facilitate eliciting plausible ranges of the sensitivity parameters. As an example of a particular c, Brumback et al. (2004) suggest $c(t, k, \bar{z}(t-1), \bar{x}(k)) = \gamma\{2z(k) - 1\}$ where γ is an unidentifiable sensitivity analysis parameter. Note that $c(t, k, \bar{z}(t-1), \bar{x}(k)) = \gamma$ for $z(k) = 1$ and $c(t, k, \bar{z}(t-1), \bar{x}(k)) = -\gamma$ for $z(k) = 0$. Thus $\gamma > 0$ ($\gamma < 0$) corresponds to subjects receiving treatment at time k having greater (smaller) mean potential outcomes at future visit t than those who did not receive treatment at visit k. When $\gamma = 0$, $Y(t) = Y^{\gamma}(t)$ and therefore $\tilde{\eta}_1 = \hat{\eta}_1$. The function c might depend on the baseline covariates $X(0)$ or the time-varying covariates $\bar{X}(k)$. In this case, as in Section 2.2.5, care should be taken in clearly communicating the sensitivity parameters' relationship to these covariates when eliciting plausible ranges from subject matter experts. Another consideration when choosing a function c is whether it will allow for the sharp null of no treatment effect, i.e., for all individuals $Y(\bar{z}, t) = Y(\bar{z}', t)$ for all $\bar{z}, \bar{z}'$, t. The example function c presented above allows for the sharp null. See Brumback et al. (2004) for other example c functions and further discussion of sensitivity analysis for marginal structural models.

2.7 Ignorance and Uncertainty Regions

Treatment effect bounds describe ignorance due to partial identifiability but do not account for uncertainty due to sampling error. This section discusses some methods to appropriately quantify uncertainty due to sampling variability when drawing inference about partially identifiable treatment effects. Over the past decade a growing body of research, especially in econometrics, has considered inference of partially identifiable parameters. The approach

presented below draws largely upon Vansteelandt et al. (2006), who considered methods for quantifying uncertainty in the general setting where missing data causes partial identifiability. As questions about treatment (or causal) effects can be viewed as missing data problems, the approach of Vansteelandt et al. generally applies (under certain assumptions) to the type of problems considered throughout this paper. This approach builds on earlier work by Robins (1997) and others.

2.7.1 Ignorance Regions

Let L be a vector containing the potential outcomes for an individual, let O denote the observed data vector, and let R be a vector containing indicator variables denoting whether the corresponding component of L is observed. For example, $L = (Y(1), Y(0))$, $O = (Z, Y)$, and $R = (Z, (1-Z))$ for the scenario described in Section 2.2 and $L = (Y(1), Y(0), S(1), S(0))$, $O = (Z, Y, S)$ and $R = (Z, (1-Z), Z, (1-Z))$ for the scenario described in Section 2.3. Denote the distribution of (L, R) by $f(L, R)$ and let $f(L) = \int f(L, R)dR$. The goal is to draw inference about a parameter vector β which is a functional of the distribution of potential outcomes L; this is sometimes made explicit by writing $\beta = \beta\{f(L)\}$. Denote the true distribution of (L, R) by $f_0(L, R)$ and the true value of β by $\beta_0 = \beta\{f_0(L)\}$. For example, $\beta_0 = E[Y(1) - Y(0)]$ for the scenario described in Section 2.2 and $\beta_0 = E[Y(1) - Y(0)|S^{P_0} = (1, 1)]$ for the scenario described in Section 2.3. Denote the true observed data distribution by $f_0(O) = \int f_0(L, R)dL_{(1-R)}$ where $L_{(1-r)}$ denotes the missing part of L when $R = r$ (i.e., the unobserved potential outcomes). The challenge in drawing inference about β_0 is that there may be multiple full data distributions $f(L, R)$ that marginalize to the true observed data distribution, i.e., $f_0(O) = \int f(L, R)dL_{(1-R)}$ for some $f \neq f_0$. When this occurs, β may be only partially identifiable from O, in which case bounds can be derived for β_0 as illustrated in the sections above.

The set of values of $\beta\{f(L)\}$ such that $f(L, R)$ marginalizes to the true observed data distribution is sometimes called the ignorance region or the identified set. These ignorance regions or intervals are distinct from traditional confidence intervals in that even as the sample

size tends to infinity these intervals will not shrink to a single point when β is partially identifiable. The ignorance region for β can be defined formally as follows. Following Robins (1997), define a class $\mathcal{M}(\gamma)$ of full data laws indexed by some sensitivity parameter vector γ to be non-parametrically identified if for each observed data law $f(O)$ there exists a unique law $f(L, R; \gamma) \in \mathcal{M}(\gamma)$ such that $f(O) = \int f(L, R; \gamma) dL_{(1-R)}$. In other words, the class $\mathcal{M}(\gamma)$ contains a unique distribution that marginalizes to each possible observed data distribution. For example, for the sensitivity analysis approach in Section 2.3.4, Hudgens and Halloran (2006, §4.3.3) defined a class of full data laws indexed by γ given in (2.19) that is non-parametrically identified. The ignorance region for β is formally defined to be

$$\text{ir}_{f_0}(\beta, \Gamma) = \Bigg\{ \beta\{f(L)\} \,:\, f(L) = \int f(L, R; \gamma) dR \text{ for some } f(L, R) \in \mathcal{M}(\Gamma) \text{ such that } \int f(L, R; \gamma) dL_{(1-R)} = f_0(O) \Bigg\}, \tag{2.37}$$

where Γ is the set of all possible values of γ under whatever set of assumptions is being invoked and $\mathcal{M}(\Gamma) = \cup_{\gamma \in \Gamma} \mathcal{M}(\gamma)$. Assume $\mathcal{M}(\Gamma)$ contains the true full data distribution, i.e., $f_0(L, R) = f(L, R, \gamma_0)$ for some $\gamma_0 \in \Gamma$. (For considerations when $\mathcal{M}(\Gamma)$ does not contain the true full data distribution, see Todem et al. (2010).) Because $\mathcal{M}(\gamma)$ is non-parametrically identified, for each $\gamma \in \Gamma$ there is a single $\beta(\gamma) = \beta\{\int f(L, R; \gamma) dR)\}$ in the ignorance region (2.37). If $\mathcal{M}(\Gamma)$ includes all possible full data distributions that marginalize to any possible observed data distribution, then the ignorance region will contain the bounds.

In practice the ignorance region will be unknown because it depends on the unknown true observed data distribution $f_0(O)$. For γ fixed, $\beta(\gamma)$ is identifiable from the observed data and the ignorance region can be estimated by estimating $\beta(\gamma)$ for each value of $\gamma \in \Gamma$, denoted by $\hat{\beta}(\gamma)$. The resulting estimator of $\text{ir}_{f_0}(\beta, \Gamma)$ is then $\{\hat{\beta}(\gamma) : \gamma \in \Gamma\}$. For scalar $\beta(\gamma)$, let $\hat{\beta}_l = \inf_{\gamma \in \Gamma}\{\hat{\beta}(\gamma)\}$ and $\hat{\beta}_u = \sup_{\gamma \in \Gamma}\{\hat{\beta}(\gamma)\}$ such that the estimated ignorance region is contained in the interval $[\hat{\beta}_l, \hat{\beta}_u]$.

2.7.2 Uncertainty Regions

Estimated ignorance regions convey ignorance due to partial identifiability and do not reflect sampling variability in the estimates. Indeed much of the literature on bounds and sensitivity analysis of treatment effects tends to report estimated ignorance regions and either ignores sampling variability or employs ad-hoc inferential approaches such as pointwise confidence intervals conditional on each value of the unidentifiable sensitivity parameter. More recent developments have provided a formal framework for conducting inference in partial identifiability settings (e.g., see Imbens and Manski, 2004; Vansteelandt et al., 2006; Romano and Shaikh, 2008; Bugni, 2010; Todem et al., 2010). The main focus in this research has been the construction of confidence regions for either the parameter β_0 or the ignorance region $\text{ir}_{f_0}(\beta_0, \Gamma)$.

Following Vansteelandt et al. (2006), a $(1-\alpha)$ pointwise uncertainty region for β_0 is defined to be a region $\text{UR}_p(\beta, \Gamma)$ such that

$$\inf_{\gamma \in \Gamma} \Pr_{f_0} \{\beta(\gamma) \in \text{UR}_p(\beta, \Gamma)\} \geq 1 - \alpha,$$

where $\Pr_{f_0}\{\cdot\}$ denotes probability under $f_0(O)$. That is, $\text{UR}_p(\beta, \Gamma)$ contains $\beta(\gamma)$ with at least probability $1-\alpha$ for all $\gamma \in \Gamma$. In particular, assuming $\gamma_0 \in \Gamma$, then $\text{UR}_p(\beta, \Gamma)$ will contain $\beta_0 = \beta(\gamma_0)$ with at least probability $1-\alpha$.

An appealing aspect of pointwise uncertainty regions is that they retain the usual duality between confidence intervals and hypothesis testing. Namely, one can test the null hypothesis $H_0 : \beta_0 = \beta_c$ versus $H_a : \beta_0 \neq \beta_c$ for some specific β_c at the α significance level by rejecting H_0 when the $(1-\alpha)$ pointwise uncertainty region $\text{UR}_p(\beta, \Gamma)$ excludes β_c. This is easily shown by noting for $\beta_c = \beta(\gamma_0)$

$$\begin{aligned} \Pr_{f_0}[\text{reject } H_0] &= 1 - \Pr_{f_0} \{\beta(\gamma_0) \in \text{UR}_p(\beta, \Gamma)\} \\ &\leq 1 - \inf_{\gamma \in \Gamma} \Pr_{f_0} \{\beta(\gamma) \in \text{UR}_p(\beta, \Gamma)\} \leq \alpha, \end{aligned}$$

where the last inequality follows because $\text{UR}_p(\beta, \Gamma)$ is a $(1-\alpha)$ pointwise uncertainty region.

Various methods under different assumptions have been proposed for constructing pointwise uncertainty regions. Imbens and Manski (2004) and Vansteelandt et al. (2006) proposed a simple method for constructing pointwise uncertainty regions for a scalar β with ignorance region $[\beta_l, \beta_u]$. Let $\gamma_l, \gamma_u \in \Gamma$ be the values of the sensitivity parameter such that $\beta_l = \beta(\gamma_l)$ and $\beta_u = \beta(\gamma_u)$. Assume

$$\begin{aligned} &\text{There exist } \hat{\beta}_l \text{ such that } \sqrt{n}(\hat{\beta}_l - \beta_l) \to^d N(0, \sigma_l^2) \text{ and } \hat{\beta}_u \text{ such that} \\ &\sqrt{n}(\hat{\beta}_u - \beta_u) \to^d N(0, \sigma_u^2). \end{aligned} \tag{2.38}$$

$$\text{The values } \gamma_l \text{ and } \gamma_u \text{ are the same for all possible observed data laws.} \tag{2.39}$$

Under assumptions (2.38) and (2.39) an asymptotic $(1-\alpha)$ pointwise uncertainty interval for β_0 is

$$\text{UR}_p(\beta, \Gamma) = \left[\hat{\beta}_l - c_\alpha \widehat{\sigma}_l/\sqrt{n},\ \hat{\beta}_u + c_\alpha \widehat{\sigma}_u/\sqrt{n}\right], \tag{2.40}$$

where c_α satisfies

$$\Phi\left(c_\alpha + \frac{\sqrt{n}\left(\hat{\beta}_u - \hat{\beta}_l\right)}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}}\right) - \Phi(-c_\alpha) = 1 - \alpha, \tag{2.41}$$

$\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal variate, and $\widehat{\sigma}_l$ and $\widehat{\sigma}_u$ are consistent estimators of σ_l and σ_u respectively (Imbens and Manski, 2004; Vansteelandt et al., 2006). Note if $\hat{\beta}_u - \hat{\beta}_l > 0$ and n is large such that the left side of (2.41) is approximately equal to $1 - \Phi(-c_\alpha)$, then $c_\alpha \approx z_{1-\alpha}$, the $(1-\alpha)$ quantile of a standard normal distribution. In contrast, if $\hat{\beta}_u = \hat{\beta}_l$, then $c_\alpha = z_{1-\alpha/2}$.

In addition to the pointwise uncertainty region, Horowitz and Manski (2000) and Vansteelandt et al. (2006) define a $(1-\alpha)$ strong uncertainty region for β_0 to be a region $\text{UR}_s(\beta, \Gamma)$ such that

$$\Pr_{f_0}\left\{\text{ir}_{f_0}(\beta, \Gamma) \subseteq \text{UR}_s(\beta, \Gamma)\right\} \geq 1 - \alpha,$$

i.e., $\text{UR}_s(\beta, \Gamma)$ contains the entire ignorance region with probability at least $1-\alpha$. Whereas the pointwise uncertainty region can be viewed as a confidence region for the partially identifiable

target parameter β_0, the strong uncertainty region is a confidence region for the ignorance region $\text{ir}_{f_0}(\beta, \Gamma)$. Clearly any strong uncertainty region will also be a (conservative) pointwise uncertainty region as $\beta_0 \in \text{ir}_{f_0}(\beta, \Gamma)$. Under assumptions (2.38) and (2.39) an asymptotic $(1-\alpha)$ strong uncertainty interval for scalar β_0 is simply

$$\text{UR}_s(\beta, \Gamma) = \left[\hat{\beta}_l - z_{1-\alpha/2}\widehat{\sigma}_l/\sqrt{n},\ \hat{\beta}_u + z_{1-\alpha/2}\widehat{\sigma}_u/\sqrt{n}\right]. \tag{2.42}$$

Note that (2.42) is equivalent to the union of all pointwise $(1-\alpha)$ confidence intervals for $\beta(\gamma)$ under $\mathcal{M}(\gamma)$ over all $\gamma \in \Gamma$, which is a simple approach often employed when reporting sensitivity analysis. Because strong uncertainty intervals are necessarily pointwise intervals, this simple approach is also a valid method for computing pointwise intervals, although intervals based on (2.40) will always be as or more narrow.

The two key assumptions (2.38) and (2.39) may not hold in general. For example, (2.38) may not hold for all possible observed data distributions, particularly for extreme values of γ_l or γ_u. Assumption (2.39) may not hold if different observed data distributions place different constraints on the possible range of γ or if Γ is chosen by the data analyst on the basis of the observed data. If (2.38) or (2.39) does not hold, alternative inferential methods are needed (e.g., see Vansteelandt and Goetghebeur, 2001; Horowitz and Manski, 2006; Chernozhukov et al., 2007; Romano and Shaikh, 2008; Stoye, 2009; Todem et al., 2010; Bugni, 2010).

A third approach to quantifying uncertainty due to sampling variability is to consider $\beta(\cdot)$ as function of γ and construct a $(1-\alpha)$ simultaneous confidence band for the function $\beta(\cdot)$. That is, a random function $\text{CB}(\cdot)$ is found such that

$$\Pr_{f_0}\{\beta(\gamma) \in \text{CB}(\gamma) \text{ for all } \gamma \in \Gamma\} \geq 1-\alpha.$$

It follows immediately that $\cup_{\gamma\in\Gamma}\text{CB}(\gamma)$ is a strong uncertainty region (and thus a pointwise uncertainty region as well). Todem et al. (2010) suggest a bootstrap approach to constructing confidence bands.

Whether pointwise uncertainty regions, strong uncertainty regions, or confidence bands

are preferred will be context specific. Typically it is of interest to draw inference about a single target parameter and not the entire ignorance region. Thus, in general pointwise uncertainty regions may have greater utility than strong uncertainty regions. Because strong uncertainty regions are necessarily conservative pointwise uncertainty regions, the strong regions can be useful in settings where determining a pointwise region is more difficult. Additionally, in some settings it may be of interest to assess whether β is non-zero, e.g., if β denotes the effect of treatment. In these settings computing a confidence band CB$(\cdot)$ has the advantage of providing the subset of Γ where the null hypothesis $\beta(\gamma) = 0$ can be rejected. This is especially appealing if γ is scalar, in which case a confidence band (as in Figure 3 of Todem et al., 2010) provides a simple approach to reporting sensitivity analysis results. On the other hand, if γ is multidimensional, visualizing confidence bands can be difficult and instead reporting the (pointwise or strong) uncertainty region may be more practical.

2.7.3 Data Example

Returning to the pertussis vaccine study described in Section 2.3, an analysis that ignores the potential for selection bias might entail computing a naive estimator (the difference in empirical means of Y between the vaccinated and unvaccinated amongst those infected) along with a 95% Wald confidence interval, which would be -0.31 (95% CI -0.38, -0.23). If the sensitivity analysis approach in Section 2.3.4 is applied, the parameter of interest $\beta(\gamma) = E[Y(1) - Y(0)|S^{P_0} = (1,1)]$ is identified for fixed values of the sensitivity analysis parameter γ given in (2.19). For fixed γ, $E[Y(0)|S^{P_0} = (1,1)]$ equals the intersection of the negative sloped line (2.14) and the curve (2.19), which is illustrated in Figure 2.1 for the pertussis data. Because $E[Y(0)|S^{P_0} = (1,1)]$ increases with γ, $\beta(\gamma)$ is a monotonically decreasing function of γ. Therefore γ_l and γ_u equal the maximum and minimum values of Γ regardless of the observed data law, indicating (2.39) holds provided that Γ is chosen by the analyst independent of the observed data.. For γ fixed and finite, $\beta(\gamma)$ can be estimated via nonparametric maximum likelihood (i.e., without any additional assumptions). This estimator will be consistent and asymptotically normal under standard regularity conditions if $\Pr[S(0) > S(1)] > 0$ (i.e., the

Table 2.1: Pertussis vaccine study data: Estimated ignorance regions and 95% pointwise and strong uncertainty regions of $\beta = E[Y(1) - Y(0)|S^{P_0} = (1,1)]$ for different Γ.

Γ	$\text{ir}_{f_0}(\beta, \Gamma)$	$\text{UR}_p(\beta, \Gamma)$	$\text{UR}_s(\beta, \Gamma)$
$[-3, 3]$	[-0.49, -0.17]	[-0.58, -0.07]	[-0.59, -0.06]
$[-5, 5]$	[-0.55, -0.15]	[-0.66, -0.05]	[-0.69, -0.03]
$[-10, 10]$	[-0.57, -0.15]	[-0.70, -0.04]	[-0.73, -0.02]
$(-\infty, \infty)$	[-0.57, -0.15]	[-0.70, -0.04]	[-0.73, -0.02]

vaccine has a protective effect against infection). For $\gamma = \pm\infty$ and $\Pr[S(0) > S(1)] > 0$, Lee (2009) proved that the estimators of the bounds similar to those given in Section 2.3.3 are consistent and asymptotically normal for a continuous outcome Y. The limiting distribution of the estimator of the upper bound ($\gamma = -\infty$) for a binary outcome will be normal if in addition

$$1 - E[Y|S = 1, Z = 0] \neq \frac{\Pr[S = 1|Z = 1]}{\Pr[S = 1|Z = 0]}, \tag{2.43}$$

and similarly the estimator of the lower bound ($\gamma = \infty$) will be asymptotically normal if in addition

$$E[Y|S = 1, Z = 0] \neq \frac{\Pr[S = 1|Z = 1]}{\Pr[S = 1|Z = 0]}. \tag{2.44}$$

Likelihood ratio tests for the null hypotheses that (2.43) and (2.44) do not hold yield p-values $p < 10^{-4}$ and $p = 0.18$ respectively, indicating strong evidence that (2.43) holds and equivocal evidence regarding (2.44). Assuming (2.43) and (2.44) both hold implies (2.38), such that (2.40) and (2.42) can be used to construct $(1 - \alpha)$ pointwise and strong uncertainty intervals for β_0. Estimated ignorance and uncertainty intervals of β_0 for different choices of Γ are given in Table 2.1 and Figure 2.2, with standard error estimates obtained using the observed information. Even for $\Gamma = (-\infty, \infty)$ both the pointwise and strong uncertainty intervals exclude zero, indicating a significant effect of vaccination. In particular, with 95% confidence we can conclude the vaccine decreased the risk of severe disease among individuals who would have become infected regardless of vaccination.

2.8 Discussion

This paper considers conducting inference about the effect of a treatment (or exposure) on an outcome of interest. Unless treatment is randomly assigned and there is perfect compliance, the effect of treatment may be only partially identifiable from the observable data. Through the five settings in Sections 2.2 – 2.6, we discussed two approaches often employed to address partial identifiability: (i) bounding the treatment effect under minimal assumptions, or (ii) invoking additional untestable assumptions that render the treatment effect identifiable and then conducting sensitivity analysis to assess how inference about the treatment effect changes as the untestable assumptions are varied. Incorporating uncertainty due to sampling variability was discussed in Section 2.7, and throughout large-sample frequentist methods were considered. Analogous Bayesian approaches to partial identification (Gustafson, 2010; Moon and Schorfheide, 2012; Richardson et al., 2011) and sensitivity analysis (McCandless et al., 2007; Gustafson et al., 2010) have also been developed.

Determining treatment effect bounds is essentially a constrained optimization problem, where the constraints are determined by the relationship between the distributions of the observable random variables and of the potential outcomes under whichever assumptions are being made. In simple cases, such as in Section 2.2.1, bounds can easily be derived from first principles and may have simple closed forms; in more complicated settings, such as in Section 2.4, bounds may be determined using linear programming or other optimization methods. In many cases calculating bounds under minimal assumptions may seem to be a meaningless exercise because the bounds are often quite wide and may not exclude the null of no treatment effect as seen with the "no assumptions" bounds in Section 2.2. On the contrary, in settings like this Robins and Greenland (1996) write: "Some argue against reporting bounds for nonidentifiable parameters, because bounds are often so wide as to be useless for making public health decisions. But we view the latter problem as a reason *for* reporting bounds in conjunction with other analyses: Wide bounds make clear that the degree to which public health decisions are dependent on merging the data with strong prior beliefs."

Bounds may be narrowed by reducing the feasible region of the optimization problem. This

may be accomplished by considering further assumptions that place restrictions on either the distributions of the potential outcomes, the distributions of the observable random variables, or both. Assumptions that place restrictions on the observable random variables may have implications which are testable. If the observed data provide evidence against any assumptions being considered, bounds should be computed without making these assumptions. Those assumptions without testable implications can only be determined to be plausible or not by subject matter experts.

A potentially less conservative approach to computing bounds is to make untestable assumptions which identify the causal estimand and then assess the robustness of inference drawn to departures from these assumptions in a sensitivity analysis. A general guideline for specifying the sensitivity analysis parameters representing these departures is to choose parameters that are easily interpretable to subject matter experts. Parameter specification will depend on whether or not sensitivity analysis is conducted by directly modeling the association of an unmeasured confounder U with treatment selection and the potential outcomes. Sensitivity analyses based on this approach are applicable when the existence of U is known and there is some historical knowledge of the magnitude association of U with Z and the potential outcomes (Robins, 1999; Brumback et al., 2004). Otherwise, alternative approaches based on directly modeling the unobserved potential outcome distributions may be preferred. A second guiding principle should be to avoid specifications of sensitivity parameters that place restrictions on the distributions of observable random variables that are not empirically supported. A third consideration when conducting sensitivity analysis concerns determining a plausible region of the sensitivity parameters. That the region be chosen prior to data analysis is in general necessary for inference, such as described in Section 2.7, to be valid. Choice of the region of the sensitivity parameters may be dictated by whether one wants to consider only mild or also severe departures from the identifying assumptions. If the identifying assumption in question is considered plausible, then it may be that only mild departures from the assumption are deemed necessary for the sensitivity analysis. In this case, subject matter experts can be consulted to determine, prior to data analysis, a plausible region for the sensitivity parameters. If, on the other hand, severe departures from untestable identifying

assumptions are to be entertained, sensitivity analyses should be conducted over all possible values of the sensitivity parameters. Sensitivity analyses which consider all possible full data distributions that marginalize to the observed data distribution will yield ignorance regions containing the bounds.

Though the examples presented here demonstrate the broad scope of scenarios where bounds and sensitivity analysis methods have been derived and employed to draw inference about treatment effects, they certainly are not exhaustive of all settings where these methods have been developed. For instance, VanderWeele et al. (2011) consider sensitivity analysis to unmeasured confounding for causal interaction effects. Bounds and sensitivity analysis methods have also recently been considered in the presence of interference, i.e., in settings where treatment of one individual may affect the outcome of another individual, such as in social networks (Ver Steeg and Galstyan, 2010; Vanderweele, 2011; Manski, 2013). For studies where sensitivity analyses are planned or anticipated, Rosenbaum and colleagues have examined how aspects of study design and the choice of statistical tests or estimators may affect the power or precision of the sensitivity analyses to be conducted (Heller et al., 2009; Rosenbaum, 2010a,b, 2011).

Bounds and sensitivity analyses of treatment effects have been utilized in various substantive settings, such as biomedical research (e.g., Cole et al., 2005; Rerks-Ngarm et al., 2009; VanderWeele and Hernández-Diaz, 2011; Hu et al., 2012) and economics (e.g., Heckman, 2001; Sianesi, 2004; Armstrong et al., 2010). Nonetheless, despite the wide range of settings in which these methods are applicable, their use in substantive settings remains somewhat limited in frequency. Given the large amount of literature detailing their broad scope of applicability and that formal inferential methods for partially identifiable parameters are now available, hopefully these approaches will be employed with greater frequency in substantive settings in the future.

The sensitivity analyses described throughout this paper focus on departures from untestable assumptions which identify treatment effects. Other types of sensitivity analyses might be considered as well, e.g., to assess how robust inferences are to various analytical decisions

that are invariably made in data analysis. Rosenbaum (2002, §11.9) refers to such assessment as "stability analysis," in contrast to the types of sensitivity analyses discussed above. See Rosenbaum (1999, 2002) and Morgan and Winship (2007, §6.2) for further discussion regarding various types of sensitivity analyses beyond the type considered here.

Figure 2.1: Graphical depiction of the bounds and sensitivity analysis model described in Sections 2.3.3 – 2.3.4. The solid thin line with negative slope represents a set of joint distribution functions of $(Z,\ S(1),\ S(0),\ Y(1),\ Y(0))$ that all give rise to the same distribution of the observable random variables (Z, S, Y). The four dotted curves depict the log odds ratio selection model for $\gamma = 0, 1, 2, 4$. The $\gamma = 0$ model is equivalent to the no selection model. Each selection model identifies exactly one pair of expectations from this set, rendering the principal effect (2.10) identifiable. The thick black lines on the edge of the unit square correspond to the lower bound of the principal effect.

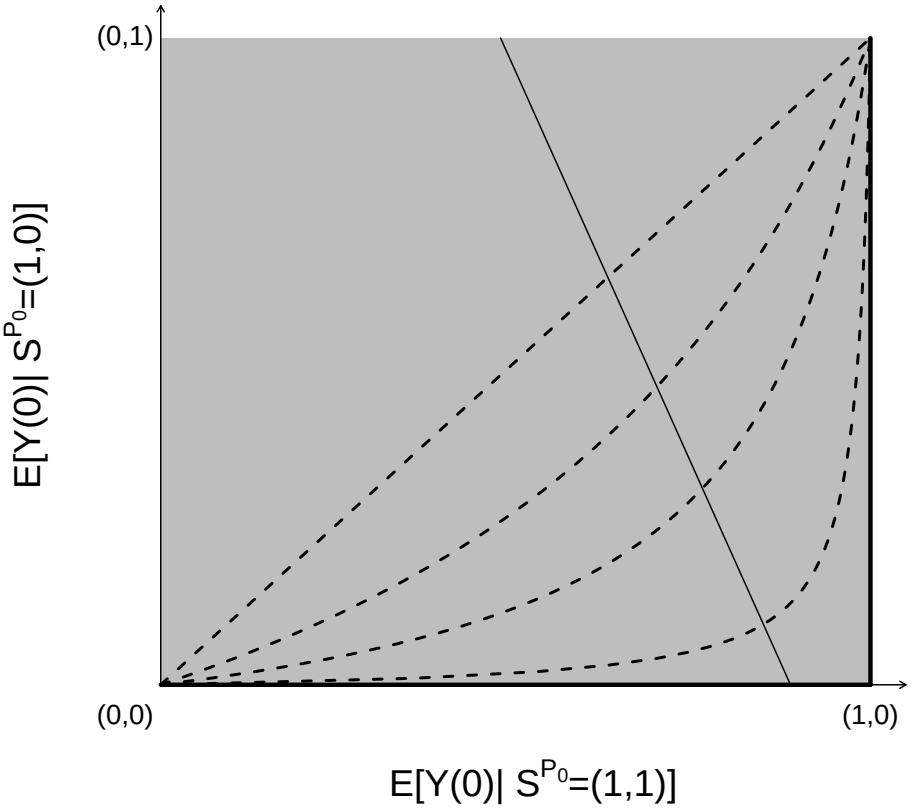
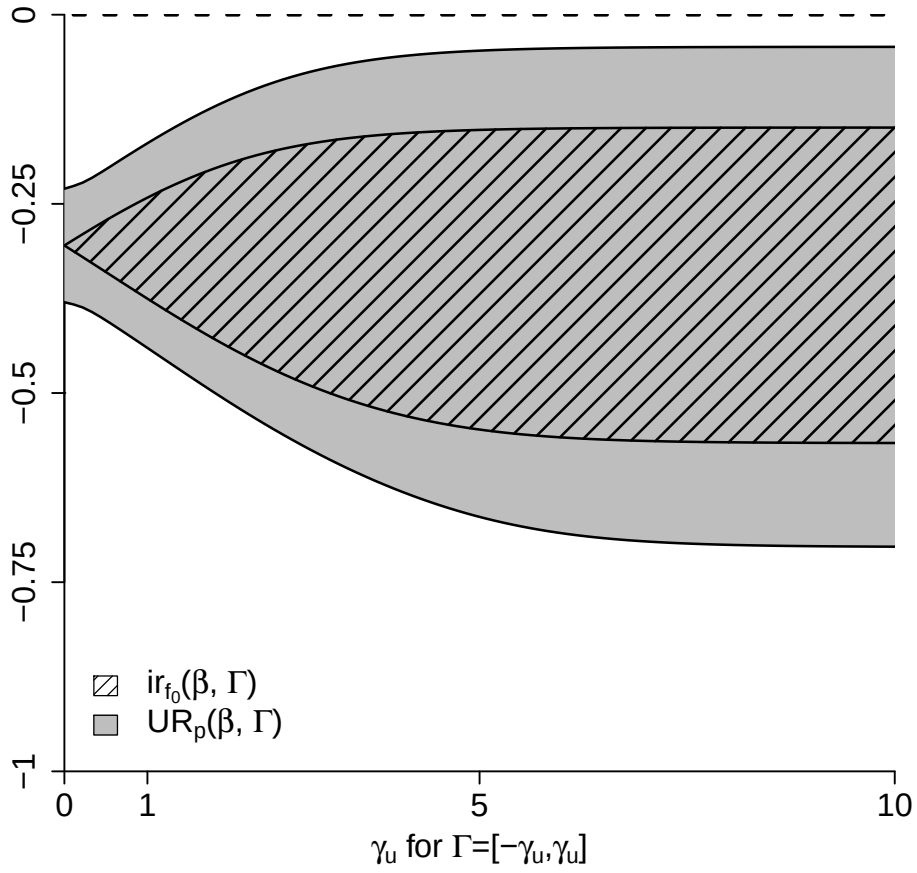


Figure 2.2: Estimated ignorance regions $\text{ir}_{f_0}(\beta, \Gamma)$ and 95% pointwise uncertainty regions $\text{UR}_p(\beta, \Gamma)$ for the pertussis vaccine example in Section 2.7.3. The principal effect (2.10) is denoted β and $\Gamma = [-\gamma_u, \gamma_u]$ for γ_u along the horizontal axis. The curve given by the lower boundary of the area with black slanted lines corresponds to $\hat{\beta}_l$, the minimum of the estimated ignorance regions, and the upper bound of the area with black slanted lines corresponds to $\hat{\beta}_u$, the maximum of the estimated ignorance region. The curve given by the lower (upper) boundary of the gray shaded area corresponds to the minimum (maximum) of the 95% pointwise uncertainty region.



CHAPTER 3: NONPARAMETRIC INSTRUMENTAL VARIABLE ANALYSIS OF COMPETING RISKS DATA

3.1 Introduction

In both randomized and non-randomized studies, researchers may seek to assess the causal effect of a treatment on the time to some event of interest, such as the effect of infant or maternal antiretroviral (ARV) therapy on the time to death or HIV infection of a breastfeeding infant of an HIV+ woman. However, such treatment effects are not identifiable without making empirically untestable assumptions about the relationship between the various causes of the event, the time to the event and the treatment allocation mechanism. With the exception of randomized clinical trials with perfect compliance to treatment assignment, the treatment allocation mechanism is unknown, which may confound standard "as treated" analyses. Such unmeasured confounding may bias treatment effect estimates in both observational studies and randomized studies with non-compliance to assigned treatment.

In some circumstances there may exist variables that are not related to the outcome except through their effect on treatment allocation. Such variables may be used to provide partial or point identification of treatment effects without knowledge of the treatment allocation mechanism (Manski, 1990; Imbens and Angrist, 1994; Angrist et al., 1996). These variables are referred to as instrumental variables and examples include treatment assignment in randomized clinical trials with non-compliance (Imbens and Angrist, 1994; Angrist et al., 1996), the calendar time for the approval of a new treatment by a regulatory agency (Martens et al., 2006; Cain et al., 2009), physician treatment prescribing preference (Brookhart and Schneeweiss, 2007), or randomized encouragement to take treatment (Martens et al., 2006). If the effect of the instrumental variable on treatment allocation is monotonic, then the instrumental variable may be used to identify treatment effects within the subpopulation whose

treatment is determined by the instrumental variable (Imbens and Angrist, 1994; Angrist et al., 1996; Hernán and Robins, 2006).

Instrumental variables have been used to identify treatment effects describing differences in survival analysis functions between treatment arms within the subpopulation described above. Inferential methods for censored data have been developed using both nonparametric methods (Baker and Lindeman, 1994; Baker, 1998; Abbring and van den Berg, 2005; Nie et al., 2011) as well as parameteric and semiparametric modelling techniques (Robins and Tsiatis, 1991; Loeys and Goetghebeur, 2003; Cuzick et al., 2007). In this paper, we consider competing risks data with multiple failure types and decompose the overall causal effect of treatment on the survival probability at a fixed time point into the sum of its causal effects on the various cause specific subdistributions. If an instrumental variable does not share common causes with either the time to the event or type of event experienced, it may be used to identify the cause-specific causal effect based on the difference in the cause-specific cumulative incidence function between treatment arms. Inferences may be obtained using nonparametric estimates of the cumulative incidence functions, analogously to the overall causal effect estimator which may entail nonparametric estimators of the survival functions.

Typically, in survival analysis, one performs an intent to treat test for differences in treatment specific survival functions using the log-rank statistic. This nonparametric test provides a global assessment of differences in survival functions over time. To our knowledge, the existing literature on nonparametric analyses of censored data with instrumental variables does not address testing for global treatment differences, providing inferential methods only at fixed time points (Baker and Lindeman, 1994; Baker, 1998; Abbring and van den Berg, 2005; Nie et al., 2011). In this paper we develop test statistics for differences in overall survival which are integrated weighted differences of the estimated causal effects over time, where the weight function may be chosen to emphasize time points of greatest interest. The tests are easily implemented using a straightforward variance estimator and are theoretically justified. The proposed statistics are extended to the competing risks setting, where they are constructed from nonparametric estimators of the differences in the treatment specific

cumulative incidence functions.

As an example of a setting where such methods may be applicable, consider the Breastfeeding, Antiretrovirals, and Nutrition (BAN) randomized clinical trial undertaken in Lilongwe, Malawi between 2004 and 2010 (Chasela et al., 2010). In this study 2369 HIV-infected breastfeeding mothers and their uninfected newborn babies were randomly assigned to one of three treatment regimens: maternal antiretrovial (ARV) therapy ($n = 849$); daily infant nevirapine (NVP) therapy ($n = 852$); or control ($n = 668$). The aim was to assess the effect of these treatment regimens on reducing mother to child transmission of HIV. Two challenges in the analysis of data from such trials are (i) not all participants comply to their randomized treatment regimen assignment and (ii) death (prior to HIV infection) is a competing risk for HIV infection. The randomized treatment assignment provides an instrumental variable that allows for estimation of treatment effects amongst those who would comply to whichever treatment they were assigned. In the BAN study treatment regimen adherence was measured via surveys administered to the mothers, allowing for estimation of such effects (assuming accurate self-report).

The organization of the paper is as follows. In Section 3.2 notation, assumptions, causal estimands and estimators are given, both for the overall and cause-specific causal effects. Section 3.3 describes nonparametric inferences using the estimators for both the standard survival set-up as well as in the presence of competing risks, and gives details of nonparametric test statistics for a global assessment of the causal effects over time. Section 3.4 presents the results of a simulation study examining the finite sample performance of these estimators and tests. Section 3.5 applies the methods derived in Section 3.3 to the BAN study. Section 3.6 concludes with a discussion.

3.2 Preliminaries

3.2.1 Notation

An instrumental variable that is often used to estimate causal effects is randomized treatment assignment in a clinical trial. Let R be an instrumental variable given by randomized assignment where $R = 0$ indicates assignment to control and $R = 1$ indicates assignment to treatment (e.g. maternal or infant ARV therapy in the BAN study). Though treatment effects of R on the outcome of interest may be identifiable if R is randomly assigned, treatment effects due to the actual treatment taken (which may differ from R) are typically the target of inference. Let Z be the actual treatment taken, where $Z = 0$ denotes treatment not taken, $Z = 1$ denotes that treatment was taken. Define potential treatment outcomes $Z(r)$ under randomized treatment assignment $r = 0, 1$; specifically let $Z(r) = 0$ indicate that the subject would not take treatment under randomized assignment r and $Z(r) = 1$ indicates that the subject would take treatment under randomized assignment r. As in Imbens and Angrist (1994) and Angrist et al. (1996), define principal strata based on the vector of the treatment potential outcomes $Z^{P_0} = (Z(0), Z(1))$ where $Z^{P_0} = (0, 1)$ are compliers (i.e. they only take treatment if they were assigned to do so), $Z^{P_0} = (1, 1)$ are the always treated, $Z^{P_0} = (0, 0)$ are the never treated, and $Z^{P_0} = (1, 0)$ are defiers (i.e. they would only take treatment when not assigned to do so).

Suppose we are interested in time to event outcomes that may be subject to competing risks. Let $T(r, z)$ be the potential first failure times under treatment assignment z and randomized assignment r and $\Delta(r, z)$ the potential event type or cause indicators that may take on values $1, \ldots, J$. Let T be the observed time to the first event for event types $j = 1, \ldots, J$, C be the censoring time and X the minimum of T and C (which is the observed follow up time). Let $\Delta = jI[T \leq C])$ be the observed event indicator where $\Delta = 0$ indicates that the subject was lost to follow up before the event was experienced. Suppose we observe n i.i.d copies of $\{X_i, R_i, Z_i, \Delta_i\}$.

3.2.2 Assumptions

Assumption 3.1. *Stable unit treatment value assumption (Rubin, 1978, SUTVA): if* $R = r$ *and* $Z = z$ *then* $Z(R) = Z(r)$ *and* $T(R, Z) = T(r, z)$ *and* $\Delta(R, Z) = \Delta(r, z)$ *for* $r, z = 0, 1$.

Assumption 3.2. *Independent instrument:* $R \amalg \{T(r, z), Z(r)$ *for* $r, z = 0, 1\}$.

Assumption 3.3. *Exclusion restriction:* $T(0, z) = T(1, z)$ *and* $\Delta(0, z) = \Delta(1, z)$ *for* $z = 0, 1$

Assumption 3.4. *Nonzero causal effect of* R *on* Z*:* $E[Z(1) - Z(0)] \neq 0$.

Assumption 3.5. *Monotonicity (Imbens and Angrist, 1994):* $Z(1) \geq Z(0)$.

Assumption 3.6. *Independent censoring:* $\{T, \Delta\} \amalg C|R$.

Assumption 3.1 is a standard assumption made in order to estimate causal effects defined using potential outcomes. Assumptions 3.2–3.4 qualify R as an instrumental variable and are the same assumptions found in Imbens and Angrist (1994) and Angrist et al. (1996). When the instrumental variable is randomly assigned, Assumption 3.2 will typically be considered plausible. Assumption 3.3 means that the potential outcomes only depend on z such that we may write $T(z) = T(r, z)$ and $\Delta(z) = \Delta(r, z)$. Assumption 3.5 implies that the defiers principal strata $Z^{P_0} = (1, 0)$ is empty (i.e., there is no subject that would take treatment only when not assigned to do so). Assumption 3.6 is made in order to estimate the all cause survival function and subdistribution functions in presence of right censoring.

3.2.3 Causal estimands

We are interested in causal effects describing differences between the survival curves of the treated versus the nontreated within the subpopulation defined by $Z^{P_0} = (0, 1)$. This is sometimes referred to as a local average treatment effect and is defined as

$$\delta(t) = \Pr[T(1) > t | Z^{P_0} = (0, 1)] - \Pr[T(0) > t | Z^{P_0} = (0, 1)]. \tag{3.1}$$

Under Assumptions 3.1–3.5, (3.1) is equivalent to

$$\delta(t) = \frac{\Pr[T > t|R = 1] - \Pr[T > t|R = 0]}{\Pr[Z = 1|R = 1] - \Pr[Z = 1|R = 0]} = \frac{S_1(t) - S_0(t)}{p_1 - p_0} = \frac{dS(t)}{dp} \tag{3.2}$$

where $S_r(t) = \Pr[T > t|R = r]$ is the survival function given $R = r$ and $p_r = \Pr[Z = 1|R = r]$. In absence of right censoring, (3.2) is equivalent to the standard instrumental variables estimand of Imbens and Angrist (1994) and Angrist et al. (1996). Under Assumptions 3.2 and 3.6, a consistent estimator $\widehat{\delta}(t)$ is found by plugging in the Kaplan Meier estimator of the survival functions $\widehat{S}_r(t) = \widehat{\Pr}[T > t|R = r]$ at time t and conditional on $R = r$ as well as a consistent estimator of each p_r such as an empirical sample mean of Z given $R = r$ (Baker, 1998; Abbring and van den Berg, 2005; Nie et al., 2011). This will be called the instrumental variables (IV) estimator of $\delta(t)$.

The local average treatment effect maybe further broken down into cause specific local average treatment effects describing differences in the subdistribution functions for specific cause j when there are competing risks for the failure time T. Namely, a local average treatment effect for cause j can defined as

$$\delta^j(t) = \Pr[T(0) \leq t,\, \Delta(0) = j|Z^{P_0} = (0, 1)] - \Pr[T(1) \leq t, \Delta(1) = j|Z^{P_0} = (0, 1)]. \tag{3.3}$$

It follows that $\delta(t) = \sum_{j=1}^{J} \delta^j(t)$, i.e., the local average treatment effect can be decomposed into the sum of cause-specific effects. Note the local average treatment effects can be zero while some of the cause specific effects are nonzero, e.g., if there are $\sum_{j \neq j'}^{J} \delta^{j'}(t) = -\delta^j(t)$ then this would occur. In context of the BAN study this could occur if infant (or maternal) ARV resulted in a reduced proportion of infants being infected with HIV, but also increased the proportion of infants dying (perhaps due to drug side effects) such that the proportions dying or becoming infected are the same in the treated versus the control arms.

In order to arrive at an expression of (3.3) that is identifiable from observable data, Assumption 3.2 will be replaced by the stronger condition given below.

Assumption 3.7. *Jointly independent instrument:* $R \amalg \{T(r, z), Z(r), \Delta(z)$ *for* $r, z = 0, 1\}$.

Under Assumptions 3.1–3.5 and 3.7, (3.3) is equivalent to the following

$$\delta^j(t) = \frac{\Pr[T \leq t,\, \Delta = j|R=0] - \Pr[T \leq t,\, \Delta = j|R=1]}{\Pr[Z=1|R=1] - \Pr[Z=1|R=0]} = \frac{F_0^j(t) - F_1^j(t)}{p_1 - p_0} = \frac{dF^j(t)}{dp} \quad (3.4)$$

where $F_r^j(t) = \Pr[T \leq t,\, \Delta = j|R=r]$ is the subdistribution (or cumulative incidence) function for cause j given $R = r$. Under Assumptions 3.6-3.7, a consistent estimator $\widehat{\delta^j}(t)$ is found by plugging in the Aalen and Johansen (1978) estimator of the subdistribution function $\widehat{F}_r^j(t) = \widehat{\Pr}[T \leq t,\, \Delta = j|R=r]$ for cause j at time t and conditional on $R = r$ and consistent estimates of each p_r. This will be called the IV estimator of $\delta^j(t)$. As with the estimands in (3.2) and (3.4), the estimator of the all cause local average treatment effect $\widehat{\delta}(t)$ equals the sum of the estimators of the cause specific local average treatment effects $\sum_{j=1}^J \widehat{\delta^j}(t)$.

3.3 Asymptotic Distributional Results

3.3.1 Pointwise Confidence Intervals

Here asymptotic pointwise confidence intervals for the local average treatment effect and the cause specific local average treatment effect are derived for some $t \in (0, \tau)$ where τ is the maximum follow up time. To present our estimation procedure, some additional notation is needed. Define $Y^i(t) = I(X_i \geq t)$ to be the unconditional at risk process, $Y_r^i = I(X_i \geq t, R_i = r)$ the conditional at risk process for randomized assignment r, and $Y_{rz}^i = I(X_i \geq t, R_i = r, Z_i = z)$ the conditional at risk process for randomized assignment r and treatment z. Define $N^{ji}(t) = I(X_i < t, \Delta_i = j)$ to be counting processes for the number of failures of type j up to time t. Let $N_r^{ji}(t) = I(X_i < t, \Delta_i = j, R_i = r)$ be the number of failures of type j up to time t for randomized treatment assignment r, and let $N_{rz}^{ji}(t) = I(X_i < t, \Delta_i = j, R_i = r, Z_i = z)$ be the number of failures of type j up to time t for randomized treatment assignment r and treatment z. Let $N^i(t) = \sum_{j=1}^J N^{ji}(t)$, $N_r^i(t) = \sum_{j=1}^J N_r^{ji}(t)$, $N_{rz}^i(t) = \sum_{j=1}^J N_{rz}^{ji}(t)$ be the corresponding total failures. Throughout, assume time is continuous such that $N^i(t)$ and $N^{i'}(t)$ do not jump at the same time for any $i \neq i' = 1 \ldots n$. Let the all cause hazard function be denoted by $\lambda(t) = \lim_{dt \to 0} \Pr[T \in (t, t+dt) |\, T > t]/dt$. Similarly, let $\lambda_r(t) = \lim_{dt \to 0} \Pr[T \in$

$(t, t+dt)|R=r,\, T>t]/dt$ and $\lambda_{rz}(t) = \lim_{dt\to 0} \Pr[T \in (t, t+dt),\, Z=z|\ R=r,\, T>t]/dt$. Let the cause specific hazard functions be $\lambda^j(t) = \lim_{dt\to 0} \Pr[T \in (t, t+dt), \Delta = j \,|\, T>t]/dt$. Similarly, let $\lambda_r^j(t) = \lim_{dt\to 0} \Pr[T \in (t, t+dt), \Delta = j | R=r,\, T>t]/dt$ and $\lambda_{rz}^j(t) = \lim_{dt\to 0} \Pr[T \in (t, t+dt), \Delta = j,\, Z=z|\ R=r,\, T>t]/dt$. Here we will consider the sequences of counting processes $N^j(t) = \sum_{i=1}^n N^{ji}(t)$, $N(t) = \sum_{j=1}^J N^j(t)$, $N_r^j(t) = \sum_{i=1}^n N_r^{ji}(t)$, $N_r(t) = \sum_{j=1}^J N_r^j(t)$, $N_{rz}^j(t) = \sum_{i=1}^n N_{rz}^{ji}(t)$, $N_{rz}(t) = \sum_{j=1}^J N_{rz}^j(t)$, $Y(t) = \sum_{i=1}^n Y^i(t)$, $Y_r(t) = \sum_{i=1}^n Y_r^i(t)$ and $Y_{rz}(t) = \sum_{i=1}^n Y_{rz}^i(t)$. Let $n_r = \sum_{i=1}^n I[R_i = r]$.

Proposition 3.1. *Assume that $n_r/n \to q_r > 0$ as $n \to \infty$ for $r = 0, 1$ and let $y_r(t) = \Pr[X \geq t \mid R = r]$ and $y_{rz}(t) = \Pr[X \geq t,\, Z = z \mid R = r]$. Assume that $y_r(t),\, y_{rz}(t) > 0$. Then*

$$\sqrt{n}\left\{\widehat{\delta}(t) - \delta(t)\right\} \xrightarrow{d} N(0, \sigma_\delta^2(t)) \text{ and } \sqrt{n}\left\{\widehat{\delta}^j(t) - \delta^j(t)\right\} \xrightarrow{d} N(0, \sigma_\delta^2(t,j)) \text{ as } n \to \infty$$

$$\begin{aligned}
\text{where } \sigma_\delta^2(t) = &\quad dp^{-2}\left[var\{\widehat{dS}(t)\} - 2\delta(t)cov\{\widehat{dS}(t), \widehat{dp}\} + \delta(t)^2 var(\widehat{dp})\right], \\
\sigma_\delta^2(t,j) = &\quad dp^{-2}\left[var\{\widehat{dF}^j(t)\} - 2\delta^j(t)cov\{\widehat{dF}^j(t), \widehat{dp}\} + \delta^j(t)^2 var(\widehat{dp})\right], \\
var\{\widehat{dS}(t)\} = &\sum_r \sigma_r^2 \text{ for } \sigma_r^2(t) = S_r(t)^2 \int_0^t \frac{\lambda_r(u)}{y_r(u)} du,\ var(\widehat{dp}) = \sum_r var(\hat{p}_r) \\
cov\{\widehat{dS}(t), \widehat{dp}\} = &\sum_r \sigma_{r1} \text{ for } \sigma_{rz}(t) = -S_r(t) \int_0^t \frac{y_{rz}(u)}{y_r(u)} \left\{\lambda_{rz}(u) - \lambda_r(u)\right\} du, \\
var\{\widehat{dF}^j(t)\} = &\sum_r \sigma_r^2(t,j) \text{ for } \sigma_r^2(t,j) = \int_0^t S_r(t)^2 \frac{\lambda_r^j(u)}{y_r(u)} du - 2\int_0^t S_r(u) \frac{\lambda_r^j(u)}{y_r(u)} \\
&\qquad \times \left\{\int_u^t S_r(s)\lambda_r^j(s) ds\right\} du + \int_0^t \frac{\lambda_r(u)}{y_r(u)} \left\{\int_u^t S_r(s)\lambda_r^j(s) ds\right\}^2 du \text{ and} \\
cov\{\widehat{dF}^j(t), \widehat{dp}\} = &\sum_r \sigma_{r1}(t,j) \text{ for } \sigma_{rz}(t,j) = \left[\int_0^t S_r(u) \frac{y_{rz}(u)}{y_r(u)} \left\{\lambda_{rz}^j(u) - \lambda_r^j(u)\right\} du \right. \\
&\qquad \left. - \int_0^t \frac{y_{rz}(u)}{y_r(u)} \left\{\lambda_{rz}(u) - \lambda_r(u)\right\} \int_u^t S_r(s)\lambda_r^j(s) ds\, du\right].
\end{aligned}$$

Consistent variance estimators $\widehat{\sigma}_\delta^2(t)$ and $\widehat{\sigma}_\delta^2(t,j)$ can be obtained by plugging in consistent estimators of $\delta(t)$, $\delta^j(t)$, $\sigma_r^2(t)$, $\sigma_{rz}(t)$, $\sigma_r^2(t,j)$, $\sigma_{rz}(t,j)$, p_1 and p_0 for $r = 0, 1$, and $z = 1$ in

the expressions above; specifically

$$\widehat{\sigma}_r^2(t) = \widehat{S_r}(t)^2 \int_0^t \left\{ \frac{n_r}{Y_r(u)-1} \right\} \frac{dN_r(u)}{Y_r(u)}, \quad \widehat{\sigma}_{rz}(t) = \widehat{S_r}(t) \int_0^t \left\{ \frac{Y_{rz}(u)dN_r(u)}{(Y_r(u))^2} - \frac{dN_{rz}(u)}{Y_r(u)} \right\},$$

$$\widehat{\sigma}_r^2(t,j) = \int_0^t \widehat{S_r}(u)^2 \left\{ \frac{n_r\{Y_r(u)-1\}}{Y_r(u)} \right\} \frac{dN_r^j(u)}{\{Y_r(u)\}^2} - 2\int_0^t \{\widehat{F}_r^j(t) - \widehat{F}_r^j(u)\}\widehat{S}_r(u)\frac{dN_r^j(u)}{Y_r(u)}$$
$$+ \int_0^t \{\widehat{F}_r^j(t) - \widehat{F}_r^j(u)\}^2 \frac{dN_r^j(u)}{Y_r(u)\{Y_r(u)-1\}} \text{ and}$$

$$\widehat{\sigma}_{rz}(t,j) = \int_0^t \widehat{S_r}(t)\frac{Y_{rz}(u)dN_r^j(u)}{\{Y_r(u)\}^2} - \int_0^t \widehat{S_r}(t)\frac{dN_{rz}^j(u)}{Y_r(u)} - \int_0^t \{\widehat{F}_r^j(t) - \widehat{F}_r^j(u)\}\frac{Y_{rz}(u)dN_r(u)}{\{Y_r(u)\}^2}$$
$$+ \int_0^t \{\widehat{F}_r^j(t) - \widehat{F}_r^j(u)\}\frac{dN_{rz}(u)}{Y_r(u)}.$$

The estimator $\widehat{\sigma}_r^2(t)$ is the usual Greenwood estimator of the variance of $\hat{S}_r(t)$. The estimator $\widehat{\sigma}_r^2(t,j)$ is the estimator of the variance of $\widehat{F}_r^j(t)$ proposed in Gaynor et al. (1993) (which most accurately estimates the true variance of $\widehat{F}_r^j(t)$ when compared to several competing estimators according to the simulation study in Braun and Yuan, 2007). Using Proposition 3.1, a $100(1-\alpha)\%$ confidence interval for $\delta(t)$ is given by $\widehat{\delta}(t) \pm z_{\alpha/2}\widehat{\sigma}_\delta(t)/\sqrt{n}$ and a $100(1-\alpha)\%$ confidence interval for $\delta^j(t)$ is given by $\widehat{\delta}^j(t) \pm z_{\alpha/2}\widehat{\sigma}_\delta(t,j)/\sqrt{n}$ where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal variate. A proof of Proposition 3.1 is contained in Appendix A.1.

3.3.2 Hypothesis Testing

To conduct tests of any difference between the two treatment groups in the survival curves and subdistribution curves for cause j, consider testing the following hypotheses

$$H_0 : \delta_w(t_0) = \int_0^{t_0} w(u)\delta(u)du = 0 \text{ and } H_0^j : \delta_w^j(t_0) = \int_0^{t_0} w(u)\delta^j(u)du = 0$$

for $t_0 \in (0,\tau)$ and where w is a user defined weight function.

Proposition 3.2. *Suppose there exists a non-negative function W such that for some $t \in$*

$[0, t_0)$, *assume that*

$$\sup_{t\in[0,t_0]} \left|\widehat{W}(t) - w(t)\right| \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Then for the null hypotheses H_0 : *and* H_0^j : *we have that*

$$\text{under } H_0,\ \sqrt{n}\left\{\widehat{\delta}_w(t_0) - \delta_w(t_0)\right\} du \xrightarrow{d} N(0, \sigma_w^2(t_0)) \text{ as } n \to \infty \text{ and}$$

$$\text{under } H_0^j,\ \sqrt{n}\left\{\widehat{\delta}_w^j(u) - \delta_w^j(u)\right\} du \xrightarrow{d} N(0, \sigma_w^2(t_0, j)) \text{ as } n \to \infty$$

$$\text{where } \widehat{\delta}_w(t_0) = \int_0^{t_0} \widehat{W}(u)\widehat{\delta}(u)du,\ \widehat{\delta}_w^j(t_0) = \int_0^{t_0} \widehat{W}(u)\widehat{\delta}^j(u)du,$$

$$\sigma_w^2(t_0) = dp^{-2}\left[\int_0^{t_0}\left\{\int_t^{t_0} w(u)S(u)du\right\}^2 \left\{y_0(t)^{-1} + y_1(t)^{-1}\right\}\lambda(t)dt\right] \text{ and}$$

$$\sigma_w^2(t_0, j) = dp^{-2}\left[\int_0^{t_0}\left\{\int_t^{t_0} w(u)du\right\}^2 \left\{y_0(t)^{-1} + y_1(t)^{-1}\right\} d\sigma^2(t, j)\right].$$

Further, $\sigma_w^2(t_0)$ *and* $\sigma_w^2(t_0, j)$ *may be consistently estimated by*

$$\widehat{\sigma}_w^2(t_0) = \widehat{dp}^{-2}\sum_r \int_0^{t_0}\left\{\int_t^{t_0} \widehat{W}(u)\widehat{S}_r(u)du\right\}^2 \left\{\frac{n_r}{(Y_r(u)-1)}\right\}\frac{dN_r(u)}{Y_r(u)} \text{ and}$$

$$\widehat{\sigma}_w^2(t_0, j) = \widehat{dp}^{-2}\sum_r \int_0^{t_0}\left\{\int_t^{t_0} \widehat{W}(u)du\right\}^2 d\widehat{\sigma}_r^2(u, j).$$

Weighted instrumental variables (WIV) tests with rejection regions defined by

$$Q = \left\{\widehat{\delta}_w(t_0) : \left|\sqrt{n}\widehat{\delta}_w(t_0)/\widehat{\sigma}_w(t_0)\right| > z_{1-\alpha/2}\right\} \text{ and}$$

$$Q^j = \left\{\widehat{\delta}_w^j(t_0) : \left|/\sqrt{n}\widehat{\delta}_w^j(t_0)/\widehat{\sigma}_w(t_0, j)\right| > z_{1-\alpha/2}\right\}$$

provide unbiased 2-sided size α *tests of* $H_0 : \delta_w(t_0) = 0$ *and* $H_0^j : \delta_w^j(t_0) = 0$.

Rejection of the WIV test for H_0 indicates that the effect of treatment on the all-cause survival experience within the $Z^{P_0} = (0, 1)$ principal strata is nonzero.

Rejection of the WIV test for H_0^j indicates that the effect of treatment on the cumulative

incidence of the event due to cause j is nonzero. Similar to the estimands in (3.2) and (3.4), $\delta_w(t) = \sum_{j=1}^{J} \delta_w^j(t)$. Again, the effect of treatment on one cause j may cancel with the effect of treatment on another cause j' such that the hypothesis of no treatment effect on the all cause survival H_0 is null, but the cause specific hypotheses of no treatment effect H_0^j and $H_0^{j'}$ are not null. Therefore the availability of unbiased tests of H_0^j for $j = 1, \ldots, J$ allows for testing of treatment effects on cause specific subdistributions that may have been missed if only a test of a treatment effect on the all cause survival H_0 were conducted. A proof of Proposition 3.2 is contained in Appendix A.2.

3.4 Simulation Study

Simulations were conducted under Asssumptions 3.1–3.7. For each simulated data set n principal strata vectors Z^{P_0} were simulated using a multinomial random number generator with parameter $\boldsymbol{\theta} = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11})$ with $\theta_{ij} = \Pr[Z^{P_0} = (i, j)]$; ($\theta_{10} = 0$ under Assumption 3.5). The parameter θ_{01} is the proportion of the population that are in the compliers principal strata and provides a measure of the strength of the instrument in determining treatment allocation. The randomized treatment assignment R was simulated by randomly permuting a vector of size n containing 0 $p_0 n$ times and 1 for the $p_1 n$ remaining entries. The random variable Z was determined based on R and Z^{P_0}. Censoring times C were generated using a uniform random number generator on the interval (C_R, $C_R + \Delta C_R$). The time to the first event $T(Z) = T$ was simulated by sampling from the distribution defined by an overall hazard of $\sum_{j=1}^{J} \lambda_{rz}^j(t)$ where each $\lambda_{rz}^j(t)$ is a Weibull hazard of the form $\kappa\gamma(\gamma t)^{\kappa-1}$ for various scenarios as detailed in Table 3.1. The event indicator Δ was simulated by sampling from a multinomial random variable with $\Pr[\Delta = j|T] = \lambda_{rz}^j(T) / \sum_{j=1}^{J} \lambda_{rz}^j(T)$ for $j = 1, 2$. If the subject was censored, Δ was set to 0. All results are based on 5,000 Monte Carlo simulations, $p_0 = 0.5$, $C_0 = 4$, $\Delta C_0 = 6$, $C_1 = 3$, $\Delta C_1 = 3$, $w(t) = 1$ and $t_0 = \min\{\max_i(X_i|R_i = 0), \max_i(X_i|R_i = 1)\}$.

Naive "as treated" analysis about (3.1) and (3.3) might entail computing the estimators

$$\tilde{\delta}(t) = \widehat{S}_{11}(t) - \widehat{S}_{00}(t) \text{ and } \tilde{\delta}^j(t) = \widehat{F}^j_{00}(t) - \widehat{F}^j_{11}(t) \tag{3.5}$$

where $\widehat{S}_{rz}(t)$ and $\widehat{F}^j_{rz}(t)$ are the Kaplan Meier estimator of the survival function and the Aalen Johansen estimator of the subdistribution function conditional on $R = r$ and $Z = z$. Pointwise confidence intervals for (3.1) and (3.3) might be computed by appealing to asymptotic normality results (Andersen et al., 1995) for $\widehat{S}_{rz}(t)$ and $\widehat{F}_{rz}(t)$. In this "as treated" analysis, testing the hypotheses H_0 and H^j_0 might be accomplished using weighted Kaplan Meier (WKM) tests as in Pepe and Fleming (1989). The coverage of the pointwise confidence intervals for $\delta(t)$ and $\delta^j(t)$ and power of the WIV tests for H_0 and H^j_0 in Propositions 3.1 and 3.2 are compared to the coverage of pointwise confidence intervals and the power of WKM tests in this naive analysis.

Nonproportional hazards in the treated versus the control amongst the complier principal strata are assumed in all scenarios. Scenario 1 describes a situation in which one cause ($j = 2$) exhibits a causal treatment effect in the complier principal strata. In this scenario, the power to reject H^2_0 is similar to H_0 and the power of H^1_0 is small (though note that this scenario is not null, i.e. $H^1_0 : \delta^1_w(t_0) \neq 0$). Scenario 2 describes a situation in which both causes exhibit causal treatment effects, but these effects cancel each other out such that $\delta_w(t_0) = 0$ (as described in Section 3.2.3 and at the end Section 3.3.2). The power to reject H_0 in Scenario 2 reflects that this test is consistent and the type I error is controlled. These opposing causal effects for $j = 1$ and 2 are roughly the same magnitude as $\delta^2_w(t_0)$ in Scenario 1, and the power to reject both H^1_0 and H^2_0 in Scenario 2 is similar to the power to reject H^2_0 in Scenario 1. Scenario 3 describes a situation in which both causes exhibit a causal treatment effect that are the same sign and magnitude. As would be expected, the power to reject H_0 in this situation is higher than that of H^1_0 or H^2_0, which are roughly the same. Scenario 4 describes a situation in which there are no causal treatment effects in the complier principal strata for cause 1 or 2 such that $\delta_w(t_0) = \delta^1_w(t_0) = \delta^2_w(t_0) = 0$. Again, as expected the results here demonstrate that the tests of H_0, H^1_0 and H^2_0 are consistent. In all scenarios the strength of the instrument (as

measured by θ_{01} – the proportion of the population who are compliers) has large effects on the power of the test, with increasing instrument strength yielding increased power.

The power for the corresponding naive weighted Kaplan Meier tests are higher, however, these tests are not unbiased. The estimated power is greater than 5% in the scenarios where the treatment effect is null (Scenario 4, and Scenario 2 for $j = (\text{all})$), meaning that these tests have inflated type I error and therefore should not be used to test for the local average treatment effects in (3.1) and (3.3).

Table 3.2 shows that the IV estimators $\widehat{\delta}(t)$ and $\widehat{\delta}^j(t)$ are unbiased and that the variance estimators accurately estimate the true variance (as indicated by the ratio of the average estimated variance and the empirical standard error). The coverage of the IV pointwise confidence intervals exhibit the ideal 0.95 in almost all scenarios (though there is slight over coverage in Scenario 2 for $\delta^1(t)$ where the treatment effect in the compliers principal strata has the opposite sign of that of the difference between the always treated and never treated principal strata). On the other hand, the naive "as treated" estimators $\tilde{\delta}(t)$ and $\tilde{\delta}^j(t)$ have higher bias and the coverage for the corresponding confidence intervals is poor in several scenarios (e.g., see Scenario 2, $j = 1$ or Scenario 4, for all j). The power to reject $H_0^j(t) : \delta^j(t) = 0$ based on the IV pointwise confidence intervals gives similar results as what was seen in Table 3.1, particularly for $t = 5$. These tests are again unbiased as indicated by the null scenarios yielding estimated power of approximately 5%. However, testing $H_0^j(t)$ using a naive analysis again results in inflated type I error.

3.5 Application to the BAN Study

In this section the methods devleoped in Section 3.3 are employed to compare cumulative incidence of HIV or death in the infant NVP arm and the maternal ARV arm to the control group in the BAN study. Treatment Z is the actual treatment taken based on the randomized assignment R and the treatment compliance surveys taken in the weeks following randomization. A subject was considered noncompliant (i.e., $Z = 0$) if any pills were reported as missed on the first completed treatment compliance survey. In the maternal ARV arm, 12%

of subjects met this criteria and in the infant NVP arm, 5% met this criteria. It was assumed that no patients in the control arm took either the infant NVP or maternal ARV treatment regimens which would imply that Assumption 3.5 holds.

The nonparametric IV and naive "as treated" estimates along with corresponding 95% confidence intervals of $\delta^j(t)$ are given in Table 3.3 for time to HIV infection ($j = 1$) or death ($j = 2$). Figure 3.1 depicts IV estimates of the all cause cumulative incidence functions partitioned by cumulative incidence of HIV and death for each treatment arm as well as the results of the WIV tests for H_0 and H_0^j. Table 3.3 and Figure 3.1 show that infant NVP decreases the probability of both infant HIV infection and death and this result is statistically significant for the composite endpoint and the HIV infection endpoint. The estimated effect of maternal ARV is positive for cumulative incidence of HIV, death and the composite endpoint, however none of these effects are significant based on the pointwise IV confidence intervals.

Table 3.3 and Figure 3.1 also demonstrate that the IV pointwise confidence intervals and WIV tests give qualitatively similar results to a naive analysis adjusting for compliance (as described in Section 3.5) when comparing the infant NVP arm to control. This might be expected because the proportion that were compliant in the infant NVP arm was quite high. However, different conclusions are reached by IV based and naive analyses when comparing the maternal ARV arm to control for the HIV infection endpoint. Specifically, a significant positive effect of maternal ARV versus control is found when using the WIV test ($\left|\sqrt{n}\widehat{\delta}_w^2(t_0)/\widehat{\sigma}_w(t_0)\right| = 1.96$, p-value 0.05) whereas the naive WKM test does not reject the null hypothesis H_0^1 of no treatment effect on cumulative incidence of HIV (Z score 1.67, p-value 0.09). Also, as seen in Table 3.3, the IV based estimates of the difference in cumulative incidence of HIV between maternal ARV and control are roughly 20-30% greater than the naive estimates. Additionally, at 18 weeks a naive confidence interval for $\delta(t)$ for maternal ARV versus control excludes 0 indicating a significant positive effect of maternal ARV on time to death or HIV infection, but the IV confidence interval does not exclude 0 and therefore does not indicate a significant positive effect.

3.6 Discussion

In this paper large sample properties are derived of nonparametric IV estimators and global hypothesis test statistics for the local average treatment effects (3.1) and (3.3) for right censored data in the presence of competing risks. Without competing risks, our proposed test enables a test for the causal effect on survival at all time points, which contrasts with previous work on fixed time points. The methods are valid under a wide range of relationships between the competing causes for the event, for local average treatment effects with nonproportional hazards and for various scenarios regarding the similarity of the distributions of the time to the event in the various principal strata. As demonstrated in Table 3.1, weaker instrumental variables (i.e. smaller proportions of the population whose treatment is determined by the instrument) yield less powerful WIV global tests of no treatment effect, but the WIV tests remain consistent for both a stronger and weaker instrumental variable. Also as evidenced by Table 3.1, naive treatment comparisons similar to those described in Section 3.4 do not yield valid results for the hypothesis test of no local average treatment effect on the all cause survival experience or on the cumulative incidence of some specific cause.

As demonstrated by Section 3.5, the use of such naive analyses may result in different conclusions being drawn about a treatment effect, which may impact important clinical or policy decisions. In the BAN study rates of non-compliance were low, and application of the IV methods here demonstrate that even when there is a low rate of non-compliance, the IV pointwise confidence intervals and WIV tests may yield different results than a naive analysis, highlighting the importance of using these methods for randomized studies with non-compliance to treatment assignment and a competing risks outcome. In a study with a higher rate of non-compliance the amount of discordance between the two analyses will likely increase.

Though the results here are valid for any instrumental variable meeting Assumptions 3.1–3.5 and 3.7, finding an instrumental variable that is unrelated to the outcome may be difficult. Relaxing Assumption 3.7 such that the instrumental variable R is independent of the outcome conditional on some set of covariates might allow for more candidate instrumental variables to

choose from. Additionally, treatment effects adjusted for covariates might also be desired. In this paper compliance is simplified to an all or nothing binary measure; however, in many real world applications compliance may be more complicated with some subjects being partially compliant. Thus results that allow for a more general form of the either the instrumental variable R or the treatment received Z may also be useful. The use of multiple weaker instrumental variables to identify local average treatment effects might also have utility in many real world applications (Hahn et al., 2004; Hausman et al., 2012).

Table 3.1: Simulation scenarios for $T(Z)$ for Z^{P_0} in presence of competing risks ($J = 2$) and power of a size $\alpha = 0.05$ WIV test of $H_0 : \delta_w^j(t_0) = 0$ and the naive WKM test discussed in Section 3.4 for $n = 300, 1000, 2000$. Results are based on $\boldsymbol{\theta} = ([1-\theta_{01}]/2, \theta_{01}, 0, [1-\theta_{01}]/2)$ for various θ_{01}. The hazard for each j within each Z^{P_0} has Weibull hazard of the form $\kappa\gamma(\gamma t)^{\kappa-1}$ for parameters (γ, κ). For $Z^{P_0} = (1, 1)$, $(\gamma, \kappa) = (0.10, 1)$ for $j = 1, 2$ and for $Z^{P_0} = (0, 0)$, $(\gamma, \kappa) = (0.16, 1)$ for $j = 1, 2$.

			(γ,κ) for $Z^{P_0}=(0,1)$		Power H_0^j: $\delta_w^j(t_0)$=0					
					WIV			Naive WKM		
					n=			n=		
Scen-ario	θ_{01}	j	z=0	z=1	300	1000	2000	300	1000	2000
1	0.6	1	(0.12,1.2)	(0.12,1.2)	5	8	14	83	98	14
		2	(0.24,1.2)	(0.12,1.2)	63	95	99	90	100	100
		(all)			56	97	100	89	100	100
	0.3	1	(0.12,1.2)	(0.12,1.2)	4	4	4	59	93	98
		2	(0.24,1.2)	(0.12,1.2)	20	51	80	20	51	80
		(all)			19	50	78	79	99	100
2	0.6	1	(0.1,1.2)	(0.2,1.2)	59	96	99	87	99	100
		2	(0.3,1.2)	(0.2,1.2)	65	97	99	89	100	100
		(all)			6	6	6	17	45	74
	0.3	1	(0.1,1.2)	(0.2,1.2)	18	45	74	18	45	74
		2	(0.3,1.2)	(0.2,1.2)	20	50	76	20	50	76
		(all)			6	5	5	34	83	99
3	0.6	1	(0.19,1.2)	(0.12,1.2)	14	35	61	30	69	90
		2	(0.19,1.2)	(0.12,1.2)	15	35	61	15	35	61
		(all)			62	98	100	91	100	100
	0.3	1	(0.19,1.2)	(0.12,1.2)	6	10	17	13	20	29
		2	(0.19,1.2)	(0.12,1.2)	8	10	17	17	19	31
		(all)			21	55	84	42	94	100
4	0.6	1	(0.2,1.2)	(0.2,1.2)	4	3	2	6	9	10
		2	(0.2,1.2)	(0.2,1.2)	5	3	3	6	10	12
		(all)			6	5	5	13	32	58
	0.3	1	(0.2,1.2)	(0.2,1.2)	4	3	3	5	7	9
		2	(0.2,1.2)	(0.2,1.2)	5	3	2	5	7	9
		(all)			5	5	5	5	17	31

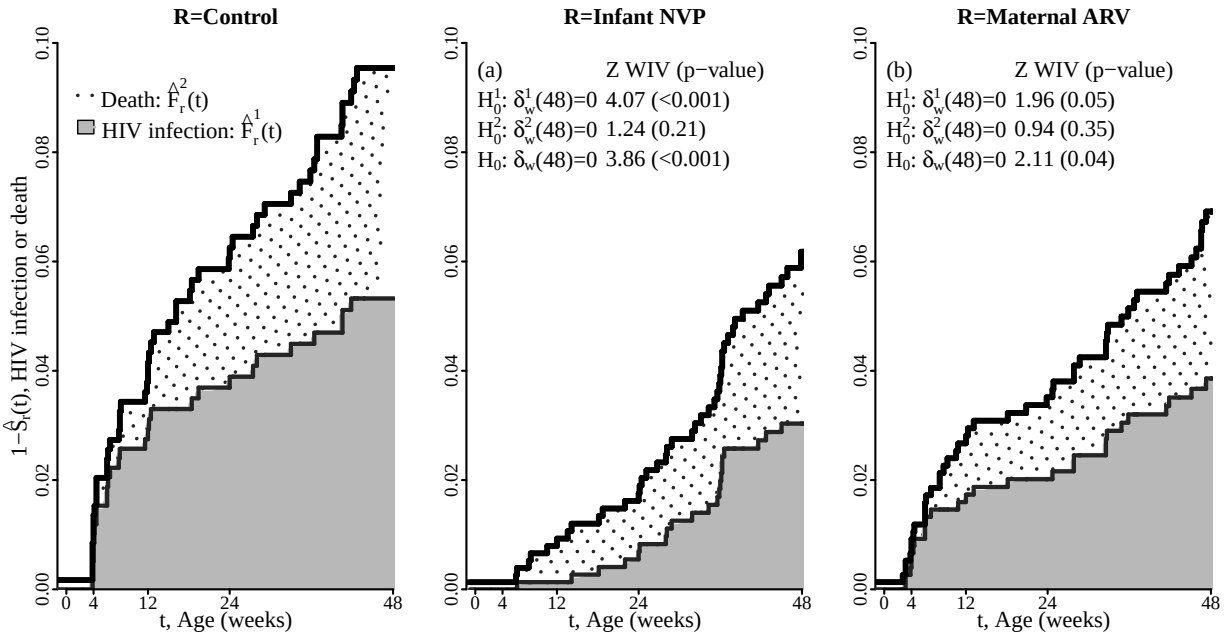
Table 3.2: Simulation results: bias ($\times$ 100), empirical standard error (ESE) ($\times$ 100), the ratio of the average estimated standard error and the empirical standard error (ESE Ratio, %), coverage of pointwise 95% confidence intervals for $\delta^j(t)$ and the percent power to reject $H_0^j(t) : \delta^j(t) = 0$ (%) based on (i) the IV estimators and pointwise confidence intervals and (ii) the naive estimators and confidence intervals for simulation Scenarios 1–4 as described in Table 3.1 for $\theta_{01} = 0.6$ and $n = 1000$.

Scen-			Bias		ESE		ESE Ratio		Coverage		Power	
ario	j	t	$\widehat{\delta}^j(t)$	$\tilde{\delta}^j(t)$	$\widehat{\delta}^j(t)$	$\tilde{\delta}^j(t)$	$\widehat{\delta}^j(t)$	$\tilde{\delta}^j(t)$	$\widehat{\delta}^j(t)$	$\tilde{\delta}^j(t)$	$\widehat{\delta}^j(t)$	$\tilde{\delta}^j(t)$
1	1	3	-0.2	2.9	4.4	3.0	104	99	96	84	11	6
		5	-0.2	4.1	5.2	3.4	105	101	96	79	31	21
	2	3	0.2	-3.3	5.0	3.6	94	90	93	82	99	100
		5	0.2	-4.3	5.6	3.7	95	98	93	79	100	100
	(all)	3	0.0	-0.4	5.3	3.5	100	101	95	95	93	100
		5	0.0	-0.3	5.2	3.4	100	101	95	95	92	100
2	1	3	-0.1	7.0	4.6	3.1	110	97	97	36	98	97
		5	-0.1	8.5	5.3	3.5	112	99	97	30	99	100
	2	3	0.1	-2.8	5.2	3.7	94	92	93	87	97	100
		5	0.1	-4.2	5.6	3.7	93	99	93	79	99	100
	(all)	3	0.0	4.1	5.2	3.4	100	102	95	78	6	22
		5	0.0	4.2	4.8	3.0	100	101	95	72	5	28
3	1	3	-0.3	-0.5	4.8	3.3	97	95	94	94	54	84
		5	-0.1	-0.3	5.5	3.7	97	98	94	94	41	70
	2	3	0.0	-0.4	4.8	3.3	96	96	94	94	57	85
		5	0.1	-0.2	5.6	3.7	96	99	94	95	43	70
	(all)	3	-0.2	-0.9	5.1	3.4	103	103	96	95	96	100
		5	0.0	-0.6	5.1	3.4	102	101	96	95	95	100
4	1	3	0.1	2.3	5.0	3.5	98	93	95	88	5	12
		5	0.0	2.3	5.5	3.7	100	99	95	90	5	10
	2	3	-0.1	2.2	5.0	3.5	99	93	94	89	6	11
		5	-0.1	2.2	5.6	3.8	98	97	95	90	5	10
	(all)	3	0.1	4.4	5.3	3.5	98	98	94	74	6	26
		5	0.0	4.4	4.9	3.1	98	99	95	69	5	31

Table 3.3: Results for the BAN study: IV $\hat{\delta}^j(t)$ and naive $\tilde{\delta}^j(t)$ estimates ($\times 100$) and corresponding 95% confidence intervals for (a) infant NVP versus control and (b) maternal ARV versus control and for endpoints of infant HIV infection ($j = 1$), death ($j = 2$) and HIV infection or death (all j).

	Treatment Comparison			
	(a) Infant NVP vs control		(b) Maternal ARV vs control	
Endpoint				
t	$\widehat{\delta}^j(t)$ (95% CI)	$\tilde{\delta}^j(t)$ (95% CI)	$\widehat{\delta}^j(t)$ (95% CI)	$\tilde{\delta}^j(t)$ (95% CI)
HIV infection ($j = 1$)				
6 weeks	1.60 (0.56, 2.64)	1.53 (0.54, 2.52)	0.69 (-0.68, 2.07)	0.60 (-0.63, 1.84)
18 weeks	3.31 (1.76, 4.86)	3.16 (1.67, 4.65)	1.79 (-0.19, 3.76)	1.59 (-0.19, 3.36)
28 weeks	3.41 (1.56, 5.25)	3.36 (1.60, 5.12)	2.20 (-0.01, 4.42)	1.88 (-0.13, 3.88)
48 weeks	2.55 (0.19, 4.92)	2.59 (0.32, 4.85)	2.07 (-0.58, 4.72)	1.71 (-0.70, 4.12)
Death ($j = 2$)				
6 weeks	0.40 (-0.26, 1.05)	0.37 (-0.26, 1.01)	0.28 (-0.49, 1.06)	0.36 (-0.29, 1.01)
18 weeks	0.69 (-0.61, 1.98)	0.61 (-0.66, 1.87)	0.43 (-1.05, 1.91)	0.80 (-0.44, 2.04)
28 weeks	1.12 (-0.55, 2.78)	0.98 (-0.64, 2.60)	1.05 (-0.80, 2.89)	1.44 (-0.13, 3.00)
48 weeks	1.55 (-0.59, 3.68)	1.34 (-0.74, 3.42)	2.01 (-0.29, 4.30)	2.35 (0.38, 4.32)
HIV infection or death (all)				
6 weeks	2.00 (0.77, 3.22)	1.90 (0.73, 3.08)	0.98 (-0.59. 2.55)	0.96 (-0.43, 2.35)
18 weeks	4.00 (2.00, 6.00)	3.76 (1.83, 5.70)	2.22 (-0.21. 4.65)	2.39 (0.24, 4.53)
28 weeks	4.52 (2.07, 6.97)	4.34 (1.98, 6.70)	3.25 (0.43. 6.07)	3.31 (0.80, 5.82)
48 weeks	4.10 (0.98, 7.22)	3.93 (0.91, 6.94)	4.08 (0.68. 7.47)	4.06 (1.01, 7.11)

Figure 3.1: Application to the BAN study: cumulative incidence estimates partitioned by cause and results of the hypothesis tests of no treatment effect on cumulative incidence of HIV, H_0^1: $\delta_w^1(t_0) = 0$; no treatment effect on death, H_0^1: $\delta_w^2(t_0) = 0$; and no effect of treatment on death or cumulative incidence of HIV, H_0: $\delta_w(t_0) = 0$ based on the WIV tests in Proposition 3.2 for (a) infant NVP versus control and (b) maternal ARV versus control.



CHAPTER 4: IDENTIFICATION OF TREATMENT EFFECTS WITH INTERFERENCE USING INSTRUMENTAL VARIABLES

4.1 Introduction

When studying causal effects of a treatment or exposure, sometimes the treatment or exposure received by one individual may affect the outcomes of other individuals under study. This is referred to as interference and is most frequently encountered in settings in which outcomes are largely dependent on social happenings. Some well known examples of settings where this might occur include the study of infectious diseases and vaccination, educational interventions, and effects of housing voucher programmes. Until recently, most causal inference research has operated under the assumption that there is no interference between units (Cox, 1958), which is part of the assumption commonly known as the stable unit treatment value assumption or SUTVA (Rubin, 1980). In the aforementioned settings, this assumption is undoubtedly violated. Moreover, effects due to interference between units are often a target of inference useful in determining important social and public health policies, where policy makers must consider the totality of an effect of a treatment or exposure, not just the effect it has at the individual level.

Though most causal inference operates under the assumption of no interference, Rubin (1980) noted that the potential outcomes framework could be extended to accommodate interference between units. Drawing inference about treatment effects in the presence of interference has since become an active area of research, especially in the last decade (Halloran and Struchiner, 1995; Hong and Raudenbush, 2006; Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; Aronow and Samii, 2011; Tchetgen Tchetgen and Vanderweele, 2012; Bowers et al., 2012). Though the advancements made account for interference and define new causal effects describing the level of interference, many of the results obtained assume a two

stage randomized experiment (randomizing both individuals as well as groups of individuals to different treatment allocation strategies). Two stage randomized experiments are sometimes difficult, impractical or unethical to implement in practice, thus adaptations to these methods for use with observational data or for randomized designs not necessarily having randomization at both the group and the individual level may have great utility. Hong and Raudenbush (2006) proposed estimators that stratify by the z-score of the propensity model to estimate interference effects in observational data. Tchetgen Tchetgen and Vanderweele (2012) proposed inverse probability of treatment weighted estimators that Perez-Heydrich et al. (2014) demonstrate consistently estimate causal interference effects in observational settings. Both the stratified and inverse probability weighted approaches rely on the assumption that there is no unmeasured confounding with regard to treatment selection. If such an assumption is to be avoided, many of the causal interference effects are in general not identifiable if treatment assignment is not randomized at either the group or individual level. Manski (2013) delineates partial identification results under general interference and describes various classes of bounding assumptions which shrink the identification regions obtained. These results lay a foundation for identification of causal effects under interference.

In this article, we assume individuals can be partitioned into groups such that there may be interference between individuals in the same groups but there is no interference between individuals in different groups. Under this assumption, identification results are derived for direct, indirect, and total effects of treatment in the presence of interference under various assumptions. These results may be used in observational settings where there exist variables that are only associated with the outcome through their effect on treatment allocation, which are commonly known as instrumental variables. These bounds might also be applicable to randomized studies where an instrumental variable may be available (e.g. a randomized encouragement design). We also derive consistent estimators of the derived bounds and estimands. The remainder of this article is organized as follows. In Section 4.2, the notation and a few key assumptions are introduced. Section 4.3 defines the direct, indirect and total effects of treatment. In Section 4.4, bounds for these three causal effects are derived under varying sets of assumptions. Section 4.5 discusses estimation of the bounds and Section 4.7

presents the motivating example on rotavirus vaccination amongst U.S. infants and applies the obtained results. Section 4.8 concludes with a discussion.

4.2 Notation, Potential Outcomes and Assumptions

Suppose we have a random sample (from some super population) of N groups of individuals each containing n_i individuals for $i = 1, \ldots, N$. Let Z_{ij} denote the treatment selected (or received) by individual j in group i where $Z_{ij} = 1$ indicates treatment selected and 0 indicates treatment not selected. Let $\overline{Z}_{i,-j}$ be the proportion of individuals in group i other than individual j that are treated and $\overline{Z}_i$ the proportion of individuals treated in group i.

At the individual level we are interested some binary outcome Y_{ij}. For instance, Y_{ij} may indicate whether or not an infection of interest occurred where $Y_{ij} = 1$ indicates that the individual becomes infected and $Y_{ij} = 0$ otherwise. Let the individual potential outcomes be denoted $Y_{ij}(z_{ij}, \overline{z}_{i,-j}, r_{ij})$ for $z_{ij} = 0, 1$, $\overline{z}_{i,-j} \in [0, 1]$ and $r_{ij} \in \mathcal{R}$, allowing for different potential outcomes for each individual treatment choice z_{ij}, each value of the proportion of other individuals in the group treated $\overline{z}_{i,-j}$ and each value of the instrumental variable r_{ij} (further discussed below). The potential outcomes are assumed to remain constant with changes in the treatment status of members of other groups; this assumption has been referred to as partial interference (Sobel, 2006) or constant treatment response (Manski, 2013). This assumption is reasonable if interaction between members of different groups is minimal or nonexistent and will be made throughout the remainder of this paper. The potential outcomes are also assumed to remain constant regardless of which specific members of the individual's group are treated. This has been referred to as stratified interference (Hudgens and Halloran, 2008). Let $\mathcal{Z}_{ij}$ be the set of all possible realizations of the vector $(Z_{ij}, \overline{Z}_{i,-j}, R_{ij})$ and denote the set of all possible potential outcomes for individual j in group i as $\mathcal{Y}_{ij}(\mathcal{Z}_{ij})$. Finally, let V_i be some set of measured group level covariates that might confound the relationship between $(Z_{ij}, \overline{Z}_{i,-j})$ and $\mathcal{Y}_{ij}(\mathcal{Z}_{ij})$.

Assumptions 4.1–4.2 introduced below for $j = 1, \ldots, n_i$ and $i = 1, \ldots, N$ are assumed throughout the rest of the paper.

Assumption 4.1. *Causal consistency: if* $Z_{ij} = z_{ij}$, $\overline{Z}_{i,-j} = \overline{z}_{i,-j}$ *and* $R_{ij} = r_{ij}$, *then* $Y_{ij}(z_{ij}, \overline{z}_{i,-j}, r_{ij}) = Y_{ij}$ *and* $Z_{ij}(r_{ij}) = Z_{ij}$ *for* $z_{ij} = 0, 1$, $r_{ij} \in \mathcal{R}$ *and for all* $\overline{z}_{i,-j}$.

Assumption 4.1 connects the observed outcomes to the potential outcomes. It states that the observed outcome is equal to the potential outcome under the observed treatment Z_{ij}, observed proportion of other group members treated $\overline{Z}_{i,-j}$, and instrumental variable R_{ij}.

Assumption 4.2. *Positivity:* $dF_{\overline{Z}_{i,-j}, Z_{ij}}(z, \overline{z}_{i,-j}) > 0$ *for* $z = 0, 1$ *and all* $\overline{z}_{i,-j}$.

Here $F_{\overline{Z}_{i,-j}, Z_{ij}}(z, \overline{z}_{i,-j})$ denotes the joint distribution of $\overline{Z}_{i,-j}, Z_{ij}$ at $(z, \overline{z}_{i,-j})$ (in general let $F_{A,B}(a, b)$ denote the joint distribution of the random variables A and B at (a, b)). Assumption 4.2 states that every combination of Z_{ij} and $\overline{Z}_{i,-j}$ is observed as the number of groups goes to infinity.

In many circumstances an instrumental variable will be available and may provide a means for arriving at tighter bounds on the causal effects defined in Section 4.3 below. Suppose that R_{ij} is some instrumental variable taking finitely many values in some set $\mathcal{R}$. Assumptions 4.3–4.6 below are analogous to the assumptions made in Imbens and Angrist (1994) and Angrist et al. (1996) to estimate average treatment effects when an instrumental variable is available (in the absence of interference). A variable R_{ij} will qualify as an instrumental variable if it meets Assumptions 4.3–4.5 below. Here assume that each individual in each group has potential treatment outcomes $Z_{ij}(r_{ij})$ for each level of this instrumental variable $r_{ij} \in \mathcal{R}$.

Assumption 4.3. *Nonzero causal effect of* R_{ij} *on* Z_{ij}*:* $E[Z_{ij}(r) - Z_{ij}(r')] \neq 0$ *for* $r \neq r'$.

Assumption 4.4. *Exclusion restriction :* $Y_{ij}(z_{ij}, \overline{z}_{i,-j}, r_{ij}) = Y_{ij}(z_{ij}, \overline{z}_{i,-j}, r'_{ij})$ *for* $z_{ij} = 0, 1$, $r_{ij}, r'_{ij} \in \mathcal{R}$ *and all* $\overline{z}_{i,-j}$

Assumption 4.3 states that the instrumental variable has some effect on treatment selection and Assumption 4.4 states that the instrumental variable has no effect on the potential outcomes for Y_{ij} for fixed z_{ij} and $\overline{z}_{i,-j}$. Assumptions 4.3 and 4.4 together imply that R_{ij} only affects the potential outcomes $\mathcal{Y}(\mathcal{Z}_{ij})$ through its effect on treatment selection.

Assumption 4.5. *Independent instrument:* $R_{ij} \amalg \{\mathcal{Y}_{ij}(\mathcal{Z}_{ij}), Z_{ij}(r_{ij})\}$ *for* $r_{ij} = 0, 1\}$

Assumption 4.5 states that the distribution of the observed value of instrumental variable does not depend on the potential outcomes $\mathcal{Y}(\mathcal{Z}_{ij})$. This assumption might be considered valid for variables R_{ij} such as calendar time, randomized encouragement to take treatment, or randomized treatment assignment.

Assumption 4.6. *Monotonicity of r_{ij} on Z_{ij}: $Z_{ij}(r_{ij}) \geq Z_{ij}(r'_{ij})$ for $r_{ij} > r'_{ij} \in \mathcal{R}$*

Under Assumption 4.6 there are no individuals that would be treated under smaller values of r but not under larger values of r. For a variable corresponding to calendar time of enrollment in a study, this assumption indicates that there are no individuals who would get treated if they enrolled earlier, but not if they enrolled later.

When there is no interference, an ideal instrumental variable to estimate the average treatment effect is given by the individual randomized treatment assignment (assuming no noncompliance to treatment assignment). As mentioned in the Introduction, when there is interference the gold standard for estimating causal effects is achieved by randomly assigning groups to different treatment allocation programmes p, meaning that $\overline{Z}_i$ is randomly assigned, and then randomly assigning Z_{ij} based on $\overline{Z}_i$. Thus an ideal instrumental variable is given by $R_{ij} = (R_i^p, R_{ij}^z)$ where R_i^p is the group treatment allocation strategy assignment and R_{ij}^z is the individual level treatment assignment based on R_i^p (again assuming no noncompliance to treatment assignment).

4.3 Causal Estimands

Often it is of interest in public health to draw inference about the relative effectiveness of different group wide treatment allocation programmes. For example, policy makers might be interested in the effect of vaccinating 90% of school aged children compared to vaccinating a smaller percentage on the incidence of some childhood disease. Consider two treatment allocation programmes $p = 0, 1$ where the proportion of individuals in a group that are treated under programme p follows some distribution indexed by parameter α_p denoted $F_{\overline{Z}}(\overline{z}|\alpha_p)$. Causal effects in the presence of interference can be defined as contrasts between average

potential outcomes under a particular treatment programme and individual treatment status (Sobel, 2006; Hudgens and Halloran, 2008; Tchetgen Tchetgen and Vanderweele, 2012). Specifically, for individual j in group i, define the individual average potential outcome under α_p as

$$\overline{Y}_{ij}(z, \alpha_p, r) = \int_0^1 Y_{ij}(z, \overline{z}_{i,-j}, r) dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{z}_{i,-j}|z; \alpha_p). \tag{4.1}$$

where $F_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{z}_{i,-j}|z)$ is the conditional distribution of $\overline{Z}_{i,-j}$ given $Z_{ij} = z$ (in general let $F_{A|B}(a|b)$ denote the distribution of the random variable A at a given $B = b$). Under Assumption 4.4, $\overline{Y}_{ij}(z, \alpha_p, r) = \overline{Y}_{ij}(z, \alpha_p, r')$ for $r, r' \in \mathcal{R}$ which we denote by $\overline{Y}_{ij}(z, \alpha_p)$. Define $E[\overline{Y}_{ij}(z; \alpha_p)]$ to be the mean individual average potential outcome (in the super-population) under individual treatment status z and group treatment allocation programme p.

As delineated by Halloran and Struchiner (1995), several effects may be of interest when studying interference. Direct effects study the effect of treatment ($z = 1$ compared to $z = 0$) while holding the treatment allocation strategy fixed. Indirect effects compare the effect of different treatment programmes ($p = 1$ compared to $p = 0$) while holding the treatment constant. These are also referred to as spillover effects (Sobel, 2006; Tchetgen Tchetgen and Vanderweele, 2012). For the purposes here, we are only interested in indirect effects for the untreated $z = 0$. T otal effects compare the effects of treatment ($z = 1$ compared to $z = 0$) while also comparing treatment allocation strategies. Formally, define

$$\begin{aligned}
\text{DE}(\alpha_0) &= E[\overline{Y}_{ij}(0; \alpha_0)] - E[\overline{Y}_{ij}(1; \alpha_0)] \\
\text{IE}(\alpha_0, \alpha_1) &= E[\overline{Y}_{ij}(0; \alpha_0)] - E[\overline{Y}_{ij}(0; \alpha_1)] \\
\text{TE}(\alpha_0, \alpha_1) &= E[\overline{Y}_{ij}(0; \alpha_0)] - E[\overline{Y}_{ij}(1; \alpha_1)]
\end{aligned} \tag{4.2}$$

to be the direct, indirect and total effects. Assumptions 4.3–4.9 introduced above and below are used to bound or identify the causal effects defined in (4.2).

Assumption 4.7. *No unmeasured confounding for other group members treatment:* $\mathcal{Y}_{ij}(\mathcal{Z}_{ij})$II $\overline{Z}_{i,-j}|\{V_i, R_{ij}\}$

Assumption 4.7 will be valid if all variables that confound the relationship between $\overline{Z}_{i,-j}$ and $\mathcal{Y}_{ij}(\mathcal{Z}_{ij})$ are measured in V_i and R_{ij}. Because $\overline{Z}_{i,-j}$ pertains to the treatment selections of other individuals, this assumption might be considered valid for a rich enough set of covariates V_i, perhaps containing only group level demographic information.

Assumption 4.8. *No unmeasured confounding for individual treatment:* $\mathcal{Y}_{ij}(\mathcal{Z}_{ij})$ II $Z_{ij}|$ $\{V_i, R_{ij}, \overline{Z}_{i,-j}\}$

Assumption 4.8 will be valid if all variables that confound the relationship between Z_{ij} and the potential outcomes $\mathcal{Y}_{ij}(\mathcal{Z}_{ij})$ are measured in V_i and R_{ij} for $\overline{Z}_{i,-j}$. However, there may exist unmeasured individual level factors that confound this relationship, meaning that Assumption 4.8 might be considered less plausible than Assumption 4.7 for V_i containing only group level and/or demographic covariates (as will be the case with the rotavirus vaccine data examined in Section 4.7).

Assumption 4.9. *Constant direct effect across* R_{ij}*:* $E[\overline{Y}_{ij}(0, \alpha_p, r) - \overline{Y}_{ij}(1, \alpha_p, r)|Z_{ij} = z, R_{ij} = r] = E[\overline{Y}_{ij}(0, \alpha_0, r) - \overline{Y}_{ij}(1, \alpha_0, r)|Z_{ij} = z, R_{ij} = r']$ *for all* $r \neq r'$ *and* $p = 0, 1$.

Assumption 4.9 will be valid if the direct effect remains constant amongst those individuals who selected treatment z across the strata defined by the instrumental variable. Assumptions similar to Assumption 4.9 are considered in Hernán and Robins (2006) are used to identify the average treatment effect using an instrumental variable in the absence of interference.

4.4 Identification Results

Bounds for $\text{DE}(\alpha_0)$, $\text{IE}(\alpha_0, \alpha_1)$ and $\text{TE}(\alpha_0, \alpha_1)$ under some subset of Assumptions 4.1–4.9 can be found by formulating optimization problems maximizing and minimizing $E[\overline{Y}_{ij}(z; \alpha_p)]$ for $z, p = 0, 1$ subject to the constraints imposed by the subset of assumptions.

Result 4.1. *Under Assumptions 4.1–4.2 and 4.7–4.8, the upper and lower bounds for* $E[\overline{Y}_{ij}(z;\alpha_p)]$ *are both given by*

$$\mu^{IPW}(z,\alpha_p) = \{dF_{Z_{ij}|V_i}(Z_{ij}|V_i,R_{ij})\}^{-1} Y_{ij}^w(z,\alpha_p;V_i) \tag{4.3}$$

where

$$Y_{ij}^w(z,\alpha_p;V_i,R_{ij}) = \frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{Z}_{i,-j}|Z_{ij};\alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i,R_{ij})} I(Z_{ij}=z)Y_{ij}$$

and thus $E[\overline{Y}_{ij}(z;\alpha_p)]$ *is identified as are* $DE(\alpha_0)$, $IE(\alpha_0,\alpha_1)$ *and* $TE(\alpha_0,\alpha_1)$.

A proof of Result 4.1 is given in Appendix B.1.

Result 4.2. *Under Assumptions 4.1–4.7, a sharp lower bound for* $E[\overline{Y}_{ij}(z;\alpha_p)]$ *is given by*

$$\mu^{LB}(z,\alpha_p) = E\left[\max_{r\in\mathcal{R}}\left\{E\left[Y_{ij}^w(z,\alpha_p;V_i,r)\right]\right\}\right] \tag{4.4}$$

and a sharp upper bound is given by

$$\mu^{UB}(z,\alpha_p) = E\left[\min_{r\in\mathcal{R}}\left\{E\left[Y_{ij}^w(z,\alpha_p;V_i,r) + dF_{Z_{ij}|V_i,R_{ij}}(1-Z_{ij}|V_i,r)\right]\right\}\right] \tag{4.5}$$

The bounds $\mu^{LB}(z,\alpha_p)$, $\mu^{LB}(z,\alpha_p)$ reduce to the bounds in Manski (1990) if there is no interference (i.e. $n_i = 1$ for all subjects). The length of these bounds (upper minus lower) is at most $\min_{r\neq r'\in\mathcal{R}}\{\Pr[Z_{ij}=0|R_{ij}=r] + \Pr[Z_{ij}=1|R_{ij}=r']\}$, which would be the rate of noncompliance for R_{ij} given by randomized treatment assignment.

Result 4.3. *Under Assumptions 4.1–4.7 and 4.9 it follows that*

$$\begin{aligned}
TE^{IV}(\alpha_0,\alpha_1) =& (d_r E[Z_{ij}])^{-1}\left\{E\left[Y_{ij}^w(\alpha_0;V_i,R_{ij})|R_{ij}=r_l\right] - E\left[Y_{ij}^w(\alpha_1;V_i,R_{ij})|R_{ij}=r_u\right]\right\} \\
IE^{IV}(\alpha_0,\alpha_1) =& \left\{E\left[Y_{ij}^w(\alpha_0;V_i,R_{ij}) - Y_{ij}^w(\alpha_1;V_i,R_{ij})|R_{ij}=r_l\right]E[Z_{ij}|R_{ij}=r_u]\right. \\
& \left. -E\left[Y_{ij}^w(\alpha_0;V_i,R_{ij}) - Y_{ij}^w(\alpha_1;V_i,R_{ij})|R_{ij}=r_u\right]E[Z_{ij}|R_{ij}=r_l]\right\} \\
DE^{IV}(\alpha_0) =& TE^{IV}(\alpha_0,\alpha_1) - IE^{IV}(\alpha_0,\alpha_1)
\end{aligned} \tag{4.6}$$

where

$$d_r E[Z_{ij}] = E[Z_{ij}|R_{ij} = r_u] - E[Z_{ij}|R_{ij} = r_l]$$

$$Y_{ij}^w(\alpha_p; V_i, R_{ij}) = \frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{Z}_{i,-j}|Z_{ij}; \alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i, R_{ij})} Y_{ij}.$$

In absence of interference $\text{IE}^{IV}(\alpha_0, \alpha_1) = 0$ and $\text{TE}^{IV}(\alpha_0, \alpha_1) = \text{DE}^{IV}(\alpha_0)$ which is the instrumental variable estimand of Imbens and Angrist (1994).

4.5 Estimation

Under Assumptions 4.1–4.2 and 4.7–4.8 a consistent estimator of $\mu^{IPW}(z, \alpha_p)$ is

$$\widehat{\mu}^{IPW}(z, \alpha_p) = \frac{1}{\sum_{i=1}^N n_i} \sum_{i=1}^N \sum_{j=1}^{n_i} \{dF_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i, R_{ij})\}^{-1} \widehat{Y}_{ij}^w(z, \alpha_p; V_i, R_{ij}) \tag{4.7}$$

where

$$\widehat{Y}_{ij}^w(z, \alpha_p; V_i) = \frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{Z}_{i,-j}|Z_{ij}; \alpha_p)}{\widehat{dF}_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i, R_{ij})} I(Z_{ij} = z) Y_{ij}.$$

Here $\widehat{dF}_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i, R_{ij})$ and $\widehat{dF}_{\overline{Z}_i|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i, R_{ij})$ are found by fitting correctly specified parametric models for $\bar{Z}_i$ and Z_{ij}. The estimator $\widehat{\mu}^{IPW}(z, \alpha_p)$ is similar to the IPW estimator found in Tchetgen Tchetgen and Vanderweele (2012) under stratified interference and allowing for the a continuous distribution of $\overline{Z}_{i,-j}$. Under Assumptions 4.1–4.7, a consistent estimator of $\mu^{LB}(z, \alpha_p)$ is

$$\widehat{\mu}^{LB}(z, \alpha_p) = \int_{V_i \in \mathcal{V}} \max_{r \in \mathcal{R}} \left\{ \widehat{E} \left[\widehat{Y}_{ij}^w(z, \alpha_p; V_i, r) \right] \right\} \widehat{dF}_{V_i}(V_i)$$

where $\mathcal{V}$ is the set of all possible V_i and $\widehat{E}[\min_{r \in \mathcal{R}}\{\cdot\}]$ can be found by taking sample means within strata defined by $R_{ij} = r$ or by fitting a correctly specified parametric model when

R_{ij} may take on many possible values. Similarly, a consistent estimator of $\mu^{UB}(z, \alpha_p)$ is

$$\widehat{\mu}^{UB}(z, \alpha_p) = \int_{V_i \in \mathcal{V}} \min_{r \in \mathcal{R}} \left\{ \widehat{E}\left[\widehat{Y}_{ij}^{w}(z, \alpha_p; V_i, r)\right] + \widehat{dF}_{Z_{ij}|V_i, R_{ij}}(1 - Z_{ij}|V_i, r) \right\} \widehat{dF}_{V_i}(V_i). \quad (4.8)$$

A consistent estimator of $\text{TE}^{IV}(\alpha_0, \alpha_1)$, $\text{IE}^{IV}(\alpha_0, \alpha_1)$ and $\text{DE}^{IV}(\alpha_0)$ can be found by plugging in in consistent estimators for $E[Z_{ij}|R_{ij} = r]$ and $E\left[Y^{w}(\alpha_p)|R_{ij} = r\right]$, which both may be found by taking sample means of Z_{ij} and $\widehat{Y}_{ij}^{w}(\alpha_p; V_i, R_{ij})$ within strata defined by $R_{ij} = r$ or by fitting correctly specified parametric models for Z_{ij} and $Y_{ij}^{w}(\alpha_p; V_i, R_{ij}))$ if R_{ij} has many levels. Here

$$\widehat{Y}_{ij}^{w}(\alpha_p; V_i, R_{ij}) = \frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{Z}_{i,-j}|Z_{ij}; \alpha_p)}{\widehat{dF}_{\overline{Z}_{i,-j}|V_i, R_{ij}}(\overline{Z}_{i,-j}|V_i, R_{ij})} Y_{ij}.$$

4.6 Simulation Study

Simulations were conducted under various sets of Assumptions 4.1–4.9 for $n_i = 60$ individuals for all i in $N = 300$ groups. For each individual in each group a binary instrumental variable R_{ij} meeting meeting assumptions 4.3–4.5 was simulated using a Bernoulli random number generator. The mean coverage for each group $\overline{Z}_i$ was simulated using a beta random number generator with mean dependent on a binary group level covariate V_i. Individual treatment $Z_{ij}|\overline{Z}_i$ was simulated using a Bernoulli random number generator with mean $\overline{Z}_i$. For each individual the potential treatment outcome under the unobserved value of the instrument was also simulated yielding the vector $Z_{ij}^{P_0} = (Z_{ij}(0,\ Z_{ij}(1))$ for each individual in each group. A binary outcome $Y_{ij} = Y_{ij}(Z_{ij}, \overline{Z}_{i,-j})$ was simulated using a Bernoulli random number generator that depends on $\overline{Z}_{i,-j}$ and $Z_{ij}^{P_0}$.

Estimates of DE(0.25), IE(0.25, 0.75), and TE(0.25, 0.75) based on $\mu^{IPW}(z, \alpha)$ and $\mu^{IV}(z, \alpha)$ were computed and corresponding bootstrap confidence intervals (using 200 replicates found by resampling entire groups) were computed for each of 1000 simulated data sets and were used to estimate the bias, empirical standard error and coverage of the bootstrap confidence intervals. Estimated bounds based on $\mu^{LB}(z, \alpha)$ and $\mu^{UB}(z, \alpha)$ were computed as well as the length of these bounds. Strong uncertainty regions as found in Vansteelandt et al.

(2006) were computed by taking the union of bootstrap confidence intervals for the lower and upper bounds.

As demonstrated by Table 4.1, when all Assumptions 4.1–4.9 are met, estimators based on $\mu^{IPW}(z, \alpha)$ and $\mu^{IV}(z, \alpha)$ are consistent and the coverage of the corresponding confidence intervals is approximately equal to the nominal 95%. The bounds under this scenario may were informative (i.e., they exclude 0) for some of the simulated datasets, however inference using the strong uncertainty regions is quite conservative as evidenced by the estimated coverage of 100% for each of the causal estimands examined. When Assumptions 4.1–4.4 and 4.7–4.8 are met, but Assumption 4.9 is not met then the estimators based on $\mu^{IPW}(z, \alpha)$ remain consistent and the corresponding confidence intervals have approximately 95% coverage as expected based on the results in Section 4.4. However, estimators based on $\mu^{IV}(z, \alpha)$ are not consistent and the corresponding confidence intervals exhibit significant under coverage. Conversely, when Assumptions 4.1–4.7 and 4.9 are met, but 4.8 is not met then the estimators based on $\mu^{IV}(z, \alpha)$ are consistent and the coverage of the corresponding confidence intervals is approximately nominal as would also be expected given the results in Section 4.4. However, under this set of assumptions estimators based on $\mu^{IPW}(z, \alpha)$ are no longer consistent and the corresponding confidence intervals have lower than nominal coverage (though the coverage is not as poor as that of the $\mu^{IV}(z, \alpha)$ estimators when Assumption 4.9 is not met). The bounds in each of these two scenarios are again conservative with regards to the coverage of the 95% strong uncertainty regions, but again may be informative. When only Assumptions 4.1–4.4 and 4.7 are met then estimators based on both $\mu^{IPW}(z, \alpha)$ and $\mu^{IV}(z, \alpha)$ are not consistent and the corresponding 95% confidence intervals have poor coverage. However, the bounds still provide valid (albeit conservative) inference about the interference effects examined. These bounds were also informative for some of the simulated datasets.

4.7 Motivating Example: Rotavirus Vaccination in U.S. Infants

Panozzo et al. (2014) analyzed data from the MarketScan Research Databases (Thomson Truven Healthcare, Inc.) that contain information on (i) rotavirus vaccination and (ii) in-

patient and outpatient claims including diagnoses of acute gastroenteritis (AGE) and more specifically, rotavirus gastoenteritis (RGE) for privately insured US infants. Infants at least one outpatient claim and that were born between May 1, 2006 and April 30, 2010 were extracted from the databases.

To assess the efficacy of rotavirus vaccine (the treatment of interest here) on reducing rotavirus gastroenteritis (RGE) and/or acute gastroenteritis (AGE) hospitalization, Panozzo et al. (2014) considered this cohort of infants and compared them to a cohort of infants born between May 1, 2000 and April 30, 2005 when a rotavirus vaccine was not available. Indirect effects of rotavirus vaccination were computed by comparing the unvaccinated in the 2006–2010 cohort to this cohort, and direct effects were computed by comparing the vaccinated infants in the 2006–2010 cohort to the respective unvaccinated infants in the 2006–2010 cohorts. Such an analysis provides direct effect estimates for the observed treatment allocation laws and indirect effects comparing these observed allocation laws to an allocation law where no one receives treatment. They found that rotavirus vaccination had a direct effect of reducing rotavirus hospitalization by 87-92% and an indirect effect of reducing rotavirus vaccination by an additional 3-8%. However, this does not take into account geographic variation in vaccine coverage.

In order to assess the interference effects in (4.2) and to account for the geographic specific coverage rates, here we apply above derived results to the data considered by Panozzo et al. (2014). We consider infants in the 2000–2005 and 2006–2010 cohorts who had at least 9 months of contiguous health insurance enrollment, had data recorded on the county in which they received health care services, and were enrolled in a county where at least 19 other infants resided and were included in the database (in order to model the coverage, or $\overline{Z}_{i,-j}$ using assuming a continuous distribution). Table 4.2 describes these cohorts of infants in more detail as well as the groups, which are defined as all captured infants over 6 weeks of age being provided health care services in the same county. These groups will be used to define the direct, indirect and total effects of vaccination.

The first dose of rotavirus vaccination should occur by age 4 months and a potential

instrumental variable might be given by the calendar time in which the infant reaches 4 months of age. Specifically, let $R_{ij} = R^p_{ij} R^z_{ij}$ where R^p indicates whether or not the vaccine was available when the infant reached 4 months of age and R^z_{ij} indicates the number of years since the introduction of the rotavirus vaccine (Feb. 2006) that infant reached 4 months of age provided $R^p_i = 1$ (rounded to the nearest quarter year). The potential outcomes $Z_{ij}(r_{ij})$ may be interpreted as the vaccination choice of individual j in group i had the vaccine been available r_{ij} years before the infant reached 4 months of age. Figure 4.1 depicts the 3,499 US counties and their estimated rotavirus vaccine coverage in each year from 2006–2010 as indicated by the Marketscan research databases and demonstrates that calendar time R^z_{ij} appears to have a fairly large positive effect on vaccination choice indicating that calendar time may be a good instrument for Z_{ij} provided that it does not have an effect on the outcome Y_{ij}. Here the outcomes of interest will be an acute gastroenteritis (AGE) diagnosis, or a rotavirus gastroenteritis (RGE) diagnosis from either an inpatient or an outpatient file (meaning that diagnoses not resulting in hospitalization were included in this analysis).

The proportion of other group members vaccinated $\overline{Z}_{i,-j}$ was modeled using a mixed effects beta regression model with covariates V_i including the rural-urban continuum code of the county; high, medium or low unemployment in the county (in the year 2006); whether or not there was a state funded vaccination program and whether or not $> 25\%$ of adults completed a college education. As $\overline{Z}_{i,-j}$ has repeated measures over calendar time, a random intercept and calendar time slope for each county was also included. Individual level vaccination $Z_{ij}|V_i$ was modeled using logistic regression with these same covariates and a random intercept for the county. Table 4.2 gives the observed proportions of each of the levels of the covariates V_i stratified by year of 4 month birthday R_{ij} for the 936,410 infants in the 2000-2005 and 2006-2010 cohorts.

Estimated values of $\text{DE}(\alpha)$, $\text{IE}(0, \alpha)$ and $\text{TE}(0, \alpha)$ based on (4.3) and 4.6 and estimated bounds based on (4.4) and (4.5) can be found in Figure 4.2 for various α and for both the AGE and the RGE outcome. For both the AGE and the RGE outcomes, the estimated indirect effect based on (4.3) and (4.6) is positive and steadily increases as α increases and the direct

effect is positive for lower values of α and then turns negative as α increases (at lower values of α for the AGE outcome than the RGE outcome). The total effect is positive for all α and slightly increases with α. Bounds for DE(α) based on (4.4) and (4.5) do not include 0 for all $\alpha \leq 0.33$. This indicates that the assumption of no unmeasured confounding for individual vaccination choice was not necessary in order to determine the sign of the direct effect when vaccine is coverage is low (for the AGE outcome only). In contrast, bounds for IE$(0, \alpha)$ based on (4.4) and (4.5) do not include 0 for all $\alpha \geq 0.69$, indicating that the assumption of no unmeasured confounding for individual vaccination choice was not necessary in order to determine the sign of the indirect effects when there is high coverage (again for the AGE outcome only). Bounds for TE$(0, \alpha)$ for the AGE outcome exclude 0 for all α.

4.8 Discussion

The results obtained demonstrate that potentially informative bounds for interference effects may be obtained under a fairly reasonable set of assumptions using an instrumental variable. The length of the bounds will be related to the ability of the instrumental variable to predict treatment selection. For rare outcomes, such as the RGE outcome in the rotavirus example above, bounds based on (4.4) and (4.5) will not be informative unless the instrumental variable very near perfectly predicts treatment selection. Identification results under two different sets of assumptions is also given, providing a means for inference about the defined causal interference effects if the required set of assumptions are considered plausible. All of these results are corroborated by the simulation study in Section 4.6. Analysis of interference effects might proceed by comparing the results obtained using the bounds and the estimators based on $\mu^{IPW}(z, \alpha)$ and $\mu^{IV}(z, \alpha)$.

In the rotavirus data, the infants were subject to both administrative censoring and right censoring due to loss of health insurance enrollment. The results here could easily be extended to account for right censoring by considering weighted Kaplan Meier estimates of the survival curve or weighted Cox models or weighted accelerated failure time models. In addition, bounds without Assumption 4.6 could be found using the simplex algorithm similarly to

Balke and Pearl (1997).

Assumptions such as 4.7, 4.8 or 4.9 might not be considered valid in general, but mild departures from this assumption could be assessed in a sensitivity analysis. Sensitivity analyses examining departures from 4.9 may be conducted in a similar fashion to the sensitivity analyses discussed in Robins et al. (1999) in absence of interference. Similarly, mild departures from Assumptions 4.7 and 4.8 maybe assessed in a sensitivity analysis by positing the existence of some unmeasured confounding variable U_{ij} as in VanderWeele and Halloran (2014).

Table 4.1: Simulation study results: true effect (×100), bias (×100), empirical standard error (ESE, ×100), coverage of confidence intervals and strong uncertainty regions (%) for the estimators based on $\mu^{IPW}(z,\alpha)$, $\mu^{IV}(z,\alpha)$ and the lower (LB) and upper (UB) bounds in 4.4 and 4.5 as well as the length of the bounds.

							Coverage			
Assumptions			Bias		ESE		CI		UR_s	Length
met	Effect	Truth	IPW	IV	IPW	IV	IPW	IV		UB−LB
4.1–4.9	DE(0.25)	23	3.1	-0.2	2.4	2.4	94	97	100	37
	IE(0.25, 0.75)	24	2.9	2.3	4.1	4.2	95	96	100	37
	TE(0.25, 0.75)	47	3.2	2.6	3.3	3.6	94	97	100	32
4.1–4.8	DE(0.25)	23	3.5	12.2	3.5	6.2	94	34	100	49
	IE(0.25, 0.75)	24	2.9	-7.1	3.7	5.1	95	27	100	49
	TE(0.25, 0.75)	47	2.7	5.3	3.4	6.4	94	37	100	41
4.1–4.7 & 4.9	DE(0.25)	23	36.1	-0.3	6.5	3.2	79	93	100	50
	IE(0.25, 0.75)	24	25.2	0.8	7.0	3.6	24	93	100	53
	TE(0.25, 0.75)	47	28.3	0.6	8.4	3.1	78	94	100	42
4.1–4.7	DE(0.25)	24	29.1	-6.3	7.0	2.2	55	54	99	59
	IE(0.25, 0.75)	25	7.7	-6.4	7.3	2.9	78	21	99	64
	TE(0.25, 0.75)	49	21.4	-9.6	6.9	2.6	57	55	100	49

Figure 4.1: Map of the US counties estimated rotavirus vaccine coverage by study year as indicated by color. Deepening shades indicate higher vaccine coverage as indicated by the legend. Orange or red shaded counties indicate a metropolitan county (100,000 or more individuals) and blue shaded counties are nonmetropolitan (source: United States Department of Agriculture, Economic Research Service from the 2010 US census). Grey shaded areas indicate that no infants were enrolled in the study for that county and study year.

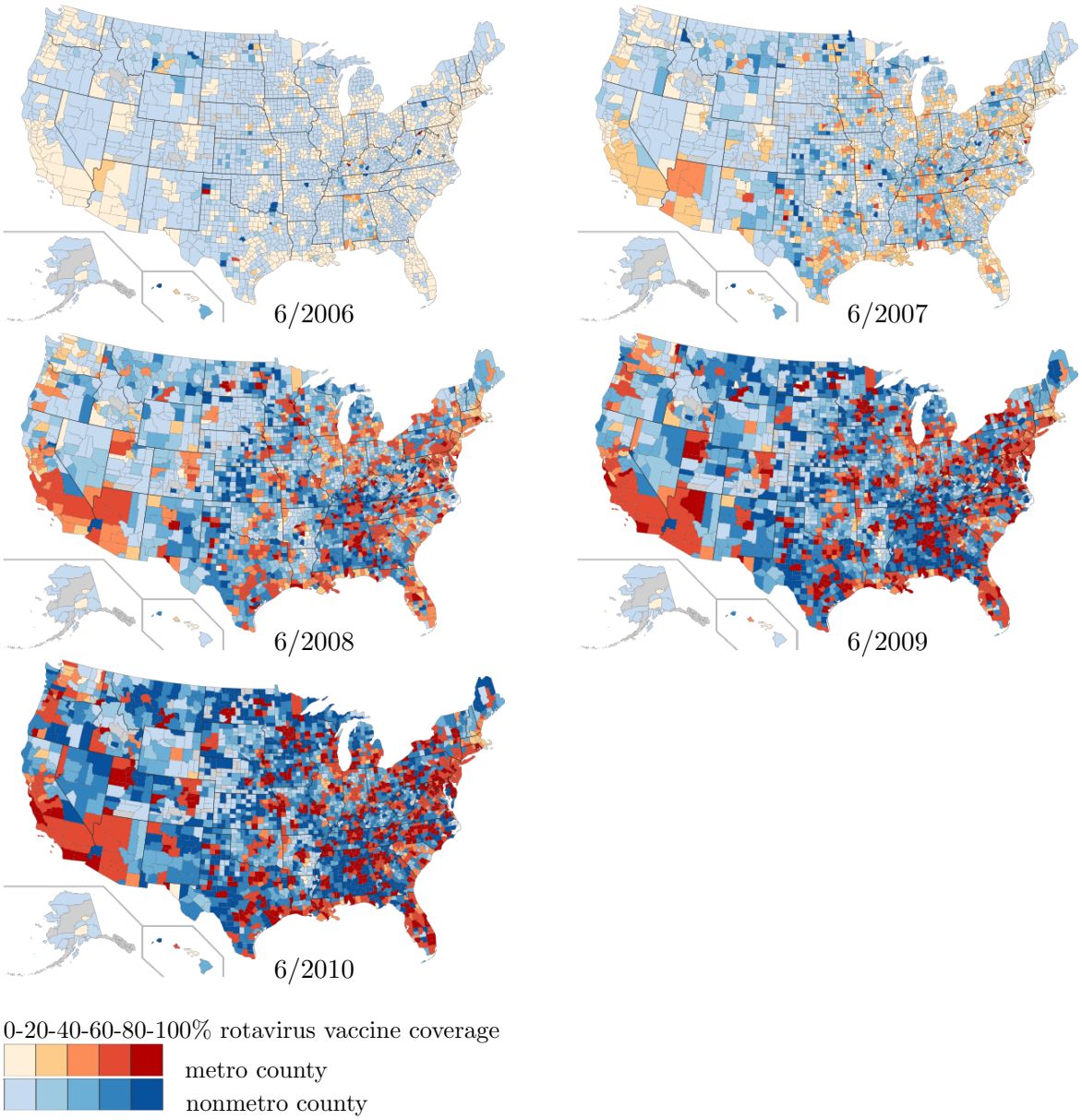
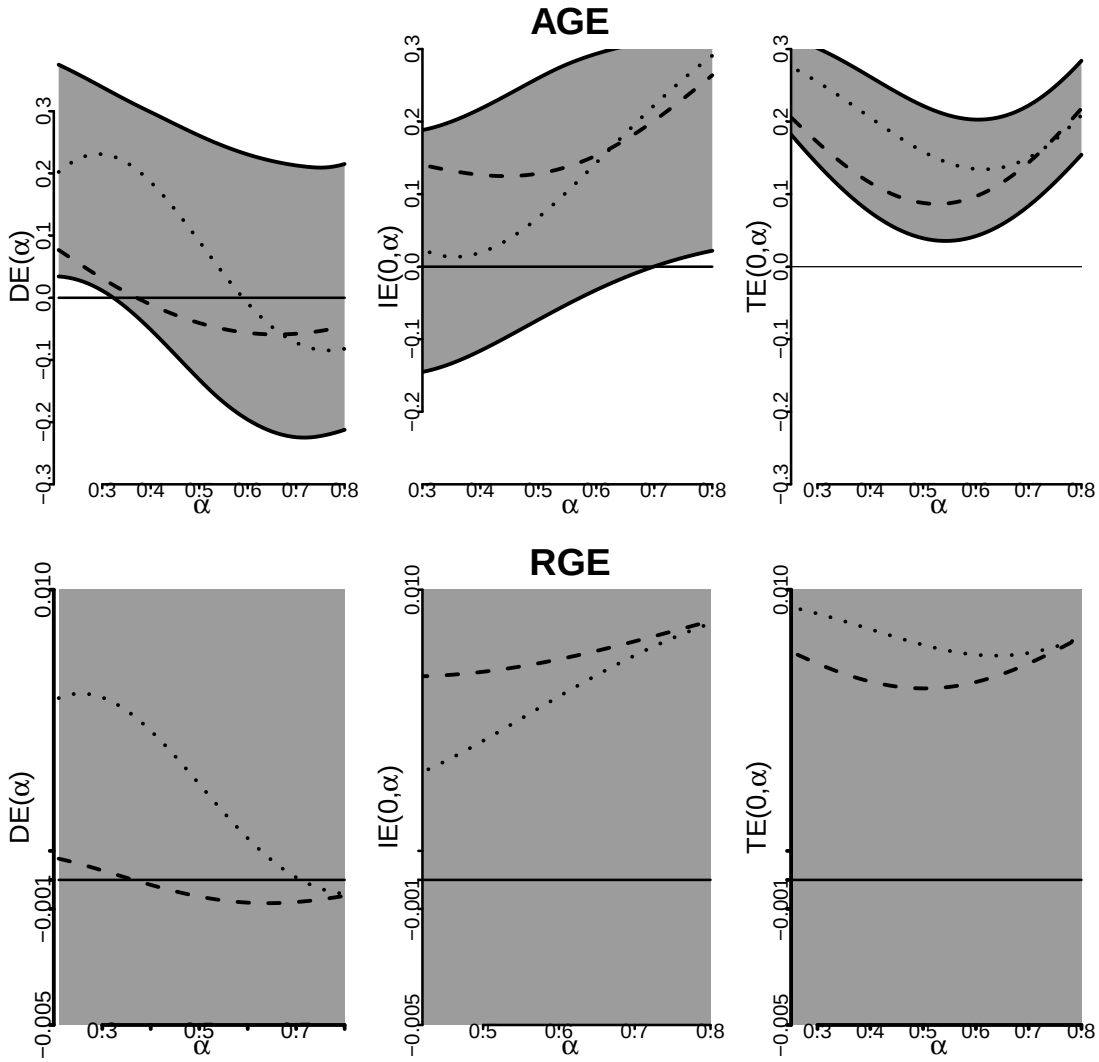


Table 4.2: Group level characteristics: rural-urban continuum code of the county; high, medium or low unemployment in the county (in the year 2006); whether or not there was a state funded vaccination program and whether or not > 25% of adults completed a college education (variables V_i) by enrollment year of the infants extracted from the MarketScan Research Databases and followed for rotavirus or acute gastroenteritis hospitalization (936,410 total infants).

	No vaccine, $R_{ij}^p = 0$ Years			Vaccine available, $R_{ij}^p = 1$ Years, R_{ij}^z			
	2000-01	2002-03	2004-05	2006	2007	2008	2009-10
Rural urban continuum code							
Metro ≥1000k	54.3%	51.3%	58.3%	60.0%	58.9%	61.0%	63.6%
Metro 250k 1000k	21.3%	23.0%	21.6%	20.6%	20.9%	20.4%	20.1%
Metro <250k	9.1%	9.8%	9.0%	9.7%	9.6%	8.8%	8.2%
Urban ≥ 20k:							
adjacent to metro	4.7%	4.5%	3.4%	3.6%	3.6%	3.1%	2.9%
not adjacent to metro	1.6%	2.2%	1.6%	1.9%	2.0%	1.8%	1.4%
Urban 2.5–19.999k:							
adjacent to metro	4.9%	5.0%	3.6%	2.7%	3.1%	3.0%	2.4%
not adjacent to metro	2.6%	2.8%	1.7%	1.3%	1.6%	1.6%	1.3%
Rural <2.5k:							
adjacent to metro	0.6%	0.6%	0.5%	0.1%	0.1%	0.2%	0.1%
not adjacent to metro	0.7%	0.7%	0.4%	0.1%	0.1%	0.1%	0.1%
Unemployment							
High (>6.5%)	19.2%	18.9%	26.1%	21.8%	22.2%	20.1%	19.8%
Medium (4-6.5%)	68.9%	67.8%	61.6%	68.5%	67.5%	66.6%	67.4%
Low (< 4%)	11.9%	14.3%	12.2%	9.7%	10.3%	13.2%	12.7%
State vaccination program	N/A	N/A	N/A	10.0%	8.8%	8.5%	9.5%
> 25% adults college educated	67.3%	68.7%	69.8%	68.2%	66.8%	69.5%	72.1%
Total	19.9k	123.4k	174.2k	105.7k	145.4k	184.8k	183.1k

Figure 4.2: Estimates of DE(α), IE$(0,\alpha)$ and TE$(0,\alpha)$ for various α based on $\mu^{IPW}(z,\alpha)$ (solid lines), $\mu^{IV}(z,\alpha)$ (dotted lines) and the bounds based on $\mu^{LB}(z,\alpha)$ and $\mu^{UB}(z,\alpha)$ (shaded area) for the AGE outcome (first row) and the RGE outcome (second row).



APPENDIX A: TECHNICAL DETAILS FOR CHAPTER 3

A.1 Proof of Proposition 3.1

We seek the asymptotic distribution of

$$\sqrt{n}\left\{\widehat{\delta}^j(t) - \delta^j(t)\right\} = \sqrt{n}dp^{-1}\left\{\widehat{dF}^j(t) - \delta^j(t)\widehat{dp}\right\} \text{ as } n \to \infty$$

which is normally distributed with mean 0 because $\widehat{dF}^j(t)$ and $\widehat{dp}$ are each asymptotically normally distributed and the IV estimator is asymptotically consistent for the difference in the subdistribution curves for the compliers principal strata. Using Slutsky's theorem, the variance of $\widehat{\delta}(t)$ is then given by

$$\begin{aligned}
&\text{var}\left\{\widehat{\delta}^j(t)\right\} = dp^{-2}\text{var}\left\{\widehat{dF}^j(t) - \widehat{dp}\delta^j(t)\right\} \\
&= dp^{-2}\left[\text{var}\{\widehat{dF}^j(t)\} - 2\delta^j(t)\text{cov}\{\widehat{dF}^j(t), \widehat{dp}\} + \delta^j(t)^2\text{var}(\widehat{dp})\right].
\end{aligned}$$

The influence functions, $L^i_{\widehat{\beta}}$ for $\widehat{\beta} = \widehat{dF}^j(t)$ and $\widehat{dp}$ are given by

$$\begin{aligned}
L^i_{\widehat{dF}^j(t)} &= \sum_r (-1)^r \left[\int_0^t \frac{S_r(u)}{y_r(u)} dN_r^{ji}(u) - \int_0^t \frac{Y_r^i(u)S_r(u)}{y_r(u)}\lambda_r^j(u)du \right. \\
&\qquad \left. - \int_0^t S_r(u)\lambda_r^j(u)\left\{\int_0^u \frac{1}{y_r(s)} dN_r^i(s) - \int_0^u \frac{Y_r^i(s)}{y_r(s)}\lambda_r(s)ds\right\}du\right] \\
L^i_{\widehat{dp}} &= \sum_r (-1)^{1-r}\left(I[Z_i = 1|R_i = r] - p_r\right)
\end{aligned}$$

(Pepe, 1991). Using Le Cam's third lemma we have the following:

$$\begin{aligned}
\text{var}\{\widehat{dF}^j(t)\} &= E\left[\{L^i_{\widehat{dF}^j(t)}\}^2\right] \\
&= \sum_r \int_0^t S_r(t)^2 \frac{\lambda_r^j(u)}{y_r(u)}du - 2\int_0^t S_r(u)\frac{\lambda_r^j(u)}{y_r(u)}\left\{\int_u^t S_r(s)\lambda_r^j(s)ds\right\}du \\
&\quad + \int_0^t \frac{\lambda_r(u)}{y_r(u)}\left\{\int_u^t S_r(s)\lambda_r^j(s)ds\right\}^2 du
\end{aligned}$$

$$\begin{aligned}
\text{cov}\{\widehat{dF}^j(t), \widehat{dp}\} &= E[L^i_{\widehat{dF}^j(t)} L^i_{\widehat{dp}}] \\
&= \sum_r \left[\int_0^t \frac{y_{r1}(u) S_r(u)}{y_r(u)} \left\{ \lambda^j_{r1}(u) - \lambda^j_r(u) \right\} du \right. \\
&\qquad \left. - \int_0^t S_r(u) \lambda^j_r(u) \int_0^u \frac{y_{r1}(s)}{y_r(s)} \left\{ \lambda_{r1}(s) - \lambda_r(s) \right\} ds\, du \right] \\
\text{var}(\widehat{dp}) &= E\{(L^i_{\widehat{dp}})^2\} = \sum_r p_r(1 - p_r)
\end{aligned}$$

(Pepe, 1991). Results for the asymptotic distribution of $\sqrt{n}\left\{\widehat{\delta}(t) - \delta(t)\right\}$ are obtained using similar arguments.

A.2 Proof of Proposition 3.2

We seek the asymptotic distribution of

$$\sqrt{n}\left\{\widehat{\delta}_w(t_0) - \delta_w(t_0)\right\} = \sqrt{n} dp^{-1} \left\{ \int_0^{t_0} W(u) \widehat{dS}(u) du - \delta_w(t_0) \widehat{dp} \right\} \text{ as } n \to \infty$$

which is normally distributed with mean 0. Using the continuous mapping theorem $\int_0^{t_0} W(u)\widehat{dS}(u)du$ is normally distributed. Because $\widehat{dp}$ is also normally distributed, Slutsky's theorem yields that the variance of $\widehat{\delta}_w(t)$ is given by

$$\begin{aligned}
\text{var}\left\{\widehat{\delta}_w(t)\right\} &= dp^{-2} \text{var}\left\{ \int_0^{t_0} W(u)\widehat{dS}(u)du - \widehat{dp}\delta(t) \right\} \\
&= dp^{-2} \left[\text{var}\left\{ \int_0^{t_0} W(u)\widehat{dS}(u)du \right\} - 2\delta_w(t_0)\text{cov}\left\{ \int_0^{t_0} W(u)\widehat{dS}(u)du, \widehat{dp} \right\} \right. \\
&\qquad \left. + \delta_w(t_0)^2 \text{var}(\widehat{dp}) \right].
\end{aligned}$$

where under the null hypothesis that $\delta_w(t_0) = 0$ the above variance reduces to $\text{var}\left\{\int_0^{t_0} W(u)\widehat{dS}(u)du\right\}$ which is given by

$$dp^{-2}\left[\int_0^{t_0}\left\{\int_t^{t_0} w(u)S(u)du\right\}^2\left\{y_0(t)^{-1}+y_1(t)^{-1}\right\}\lambda(t)dt\right]$$

(Pepe, 1991). Results for the asymptotic distribution of $\sqrt{n}\left\{\widehat{\delta}_w^j(t_0) - \delta_w^j(t_0)\right\}$ are obtained using similar arguments.

APPENDIX B: TECHNICAL DETAILS FOR CHAPTER 4

B.1 Proof that 4.3 identifies $E[\overline{Y}_{ij}(z;\alpha_p)]$

To see that $\mu^{IPW}(z,\alpha_p)$ is equal to $E[\overline{Y}_{ij}(z,\alpha_p)]$ under Assumptions 4.7–4.8 note

$$
\begin{aligned}
& E\left\{E\left[\frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{Z}_{i,-j}|Z_{ij};\alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i}(\overline{Z}_{i,-j}|V_i,R_{ij})dF_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i)}I(Z_{ij}=z)Y_{ij}(Z_{ij},\overline{Z}_{ij},R_{ij})\right]\right\} \\
= \; & E\left[\sum_{z_{ij}=0}^{1}\int_0^1\left\{\frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{Z}_{i,-j}|Z_{ij};\alpha_p)Y_{ij}(z_{ij},\overline{z}_{i,-j},R_{ij})I[z_{ij}=z]}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i,R_{ij})dF_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i,R_{ij})}\times\right.\right. \\
& \qquad \left.\left. dF_{\overline{Z}_{i,-j}|\mathcal{Y}_{ij}(\mathcal{Z}_{ij})}(\overline{z}_{i,-j}|\mathcal{Y}_{ij}(\mathcal{Z}_{ij}))dF_{Z_{ij}|\mathcal{Y}_{ij}(\mathcal{Z}_{ij})}(z_{ij}|\mathcal{Y}_{ij}(\mathcal{Z}_{ij}))\right\}\right] \\
= \; & E\left[\int_0^1\left\{\frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{Z}_{i,-j}|Z_{ij};\alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i)dF_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i,R_{ij})}Y_{ij}(z_{ij},\overline{z}_{i,-j},R_{ij})\times\right.\right. \\
& \qquad \left.\left. dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{z}_{i,-j}|V_i)dF_{Z_{ij}|V_i,R_{ij}}(z_{ij}|V_i,R_{ij})\right\}\right] \\
= \; & \int_0^1\int_0^1 Y_{ij}(z,\overline{z}_{i,-j},R_{ij})dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{z}_{i,-j}|z;\alpha_p)dF_{\overline{Y}_{ij}(z,\alpha_p)}(y) \\
= \; & \int_0^1 \overline{Y}_{ij}(z,\alpha_p)dF_{\overline{Y}_{ij}(z,\alpha_p)}(y) \;=\; E[\overline{Y}_{ij}(z;\alpha_p)].
\end{aligned}
$$

The first equality comes from Assumption 4.1, the second from Assumptions 4.7–4.8. The third, fourth and fifth come from properties of expectations and algebraic manipulation.

B.2 Estimation of 4.3

To show that $\widehat{\mu}^{IPW}(z,\alpha_p)$ consistently estimates $\mu^{IPW}(z,\alpha_p)$, let

$$
g(Y_{ij},Z_{ij},\overline{Z}_{i,-j},z,\alpha_p) = dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{z}_{i,-j}|z;\alpha_p)I[Z_{ij}=z]Y_{ij} \text{ and}
$$

$$
\psi_{z,\alpha_p}(Y_{ij},Z_{ij},\overline{Z}_{i,-j},\mu(z,\alpha_p)) = \frac{g(Y_{ij},Z_{ij},\overline{Z}_{i,-j},z,\alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i,R_{ij})dF_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i,R_{ij})}
$$

then for $m = \sum_{i=1}^{N} n_i$

$$\widehat{\mu}(z, \alpha_p) = m^{-1} \sum_{i=1}^{N} \sum_{j=1}^{n_i} \frac{g(Y_{ij}, Z_{ij}, \overline{Z}_{i,-j}, z, \alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i, R_{ij}) dF_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i, R_{ij})}$$

and $\widehat{\mu}(z, \alpha_p)$ is a solution (for $\mu(z, \alpha_p)$) to the estimating equation

$$\sum_{i=1}^{N} \sum_{j=1}^{n_i} \psi_{z,\alpha_p}(Y_{ij}, Z_{ij}, \overline{Z}_{i,-j}, \mu(z, \alpha_p)) = 0.$$

Thus, by M-estimation theory $\widehat{\mu}(z, \alpha_p) \xrightarrow{p} \mu(z, \alpha_p)$ provided

$$\widehat{dF}_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i, R_{ij}) \widehat{dF}_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i, R_{ij}) \xrightarrow{p}$$

$$dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i, R_{ij}) dF_{Z_{ij}|V_i,R_{ij}}(Z_{ij}|V_i, R_{ij})$$

.

B.3 Proof that 4.4 and 4.5 are sharp bounds

To see that the bounds in (4.4) and (4.5) are sharp, let $L'_{ij} = \{\overline{Y}_{ij}(z, \alpha_p, r),\ Z_{ij}(r) : r \in \mathcal{R}, V_i, R_{ij}\}$, $\mathcal{L}$ the set of all possible realizations of L_{ij} and

$$p_{zr}(\alpha_p) = \int_{l \in \mathcal{P}_{zr}} \overline{Y}_{ij}(z, \alpha_p, r) dF_{L_{ij}}(l)$$

where $\mathcal{P}_{zr}$ is the set of all realizations of L_{ij} such that $Z_{ij}(r) = z$. Note that

$$\begin{aligned}
& E\left\{E\left[\frac{dF_{\overline{Z}_{i,-j}}(\overline{Z}_{i,-j}|Z_{ij};\alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{Z}_{i,-j}|V_i,R_{ij})} I(Z_{ij}(R_{ij})=z) Y_{ij}(Z_{ij}(R_{ij}),\overline{Z}_{i,-j},R_{ij})|R_{ij}=r\right]\right\} \\
= \; & E\left[\int_0^1 \left\{\frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{z}_{i,-j}|z_{ij};\alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,r}(\overline{z}_{i,-j}|V_i,r)} Y_{ij}(Z_{ij}(r),\overline{z}_{i,-j}) I[Z_{ij}(r)=z]\times \right.\right. \\
& \qquad \left.\left. dF_{\overline{Z}_{i,-j}|\mathcal{Y}_{ij}(\mathcal{Z}_{ij})}(\overline{z}_{i,-j}|\mathcal{Y}_{ij}(\mathcal{Z}_{ij}))\right\}\right] \\
= \; & E\left[\int_0^1 \left\{\frac{dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{z}_{i,-j}|z_{ij};\alpha_p)}{dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{z}_{i,-j}|V_i,r)} Y_{ij}(z,\overline{z}_{i,-j},r) dF_{\overline{Z}_{i,-j}|V_i,R_{ij}}(\overline{z}_{i,-j}|V_i,r)\right\}\right] \\
= \; & \int\limits_{l\in\mathcal{P}_{zrv}} \int_0^1 Y_{ij}(z,\overline{z}_{i,-j},r) dF_{\overline{Z}_{i,-j}|Z_{ij}}(\overline{z}_{i,-j}|z_{ij};\alpha_p) dF_{L_{ij}}(l) \\
= \; & \int\limits_{l\in\mathcal{P}_{zr}} \overline{Y}_{ij}(z,\alpha_p,r) dF_{L_{ij}}(l) \;=\; p_{zr}(\alpha_p)
\end{aligned} \tag{B.3.1}$$

for all $r \in \mathcal{R}$ and $V_i \in \mathcal{V}$. The first line again comes from Assumption 4.1. The first equality (2nd line) comes from Assumption 4.5. The second equality (3rd line) comes from Assumption 4.7. Note also that

$$\begin{aligned}
E[\overline{Y}_{ij}(z,\alpha)] \;&=\; \int\limits_{l\in\mathcal{L}} \overline{Y}_{ij}(z,\alpha_p,r) dF_{L_{ij}}(l) \\
&=\; \int\limits_{l\in\mathcal{P}_{zr}} \overline{Y}_{ij}(z,\alpha_p,r) dF_{L_{ij}}(l) + \int\limits_{l\in\{\mathcal{L}-\mathcal{P}_{zr}\}} \overline{Y}_{ij}(z,\alpha_p,r) dF_{L_{ij}}(l) \\
&=\; p_{zr}(\alpha_p;V_i) + \int\limits_{l\in\{\mathcal{L}-\mathcal{P}_{zt}\}} \overline{Y}_{ij}(z,\alpha_p,r) dF_{L_{ij}}(l)
\end{aligned}$$

for all $r \in \mathcal{R}$ under Assumption 4.4. Because $p_{zr}(\alpha_p)$ is identified from the observable data, bounds for $E[\overline{Y}_{ij}(z,\alpha)]$ may be found by maximizing and minimizing

$$q_{zr}(\alpha_p) = \int\limits_{l\in\{\mathcal{L}-\mathcal{P}_{zr}\}} \overline{Y}_{ij}(z,\alpha_p,r) dF_{L_{ij}}(l).$$

A lower bound in q_{zr} is reached when $Y(z,\alpha_p,r) = 0$ for all $L_{ij} \in \mathcal{L} - \mathcal{P}_{zr}$ and an upper bound when $Y(z,\alpha_p,r) = 1$. Thus a lower bound is given by $\max_{r\in\mathcal{R}} p_{zr}$ (by Assumption 4.4) and an upper bound is given by $\min_{r\in\mathcal{R}}\{p_{zr} + \Pr[l \in \{\mathcal{L} - \mathcal{P}_{zr}\}]$. Under Assumption 4.6,

$\Pr[L_{ij} \in \{\mathcal{L} - \mathcal{P}_{zr}\}] = \Pr[Z_{ij}(r) = 1 - z] = \Pr[Z_{ij} = 1 - z]$. These results will hold within strata defined by $V_i = v$ and thus a tight lower and upper bound maybe computed by finding bounds within these strata and then integrating across V_i as in the bounds in (4.4)–(4.5).

B.4 Proof that 4.6 identifies 4.2

To see that $\text{DE}(\alpha_0)$ is given by the expression in (4.6) under Assumptions 4.3–4.7 and 4.9 note that the numerator of the estimand for $\text{DE}(\alpha_0)$

$$
\begin{aligned}
&E[Y_{ij}^{w}(\alpha_0; V_i, R_{ij})|R_{ij} = r_l] - E[\alpha_0; V_i R_{ij}|R_{ij} = r_u] \\
=&\sum_{z=0}^{1} E[Y_{ij}^{w}(\alpha_0; V_i, R_{ij}) I[Z(r_l) = z]|R_{ij} = r_l] - E[Y_{ij}^{w}(\alpha_0; V_i, R_{ij}) I[Z_{ij}(r_u) = z]|R_{ij} = r_u] \\
=&E[(\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij}))(Z_{ij}(r_u) - Z_{ij}(r_l))] \\
=&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij}(r_u) > Z_{ij}(r_l)] \Pr[Z_{ij}(r_u) > Z_{ij}(r_l)] \\
=&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij}(r_u) > Z_{ij}(r_l)](d_r E[Z_{ij}]). \qquad \text{(B.4.1)}
\end{aligned}
$$

The first equality comes from properties of expectations, the second from rearranging terms and applying the results in (B.3.1). The third and fourth equalities come from Assumption 4.6. Continuing,

$$
\begin{aligned}
&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij}(r_u) > Z_{ij}(r_l)] \\
=&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij}(r_u) = 1, Z_{ij}(r_l) = 0] \\
=&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij} = 1, R_{ij} = r_u] \text{ or} \\
&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij} = 0, R_{ij} = r_l]. \\
&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij} = 1, R_{ij} = r_u] \\
=&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij} = 1, R_{ij} = r] \text{ and} \\
&E[\overline{Y}_{ij}(0, \alpha_0) - \overline{Y}_{ij}(1, \alpha_0; V_i, R_{ij})|Z_{ij} = 0, R_{ij} = r_l]
\end{aligned}
$$

$$= E[\overline{Y}_{ij}(0,\alpha_0) - \overline{Y}_{ij}(1,\alpha_0;V_i,R_{ij})|Z_{ij}=0,R_{ij}=r]$$

for all $r \in \mathcal{R}$.

The first equality holds because Z_{ij} is dichotomous and the second and third equalities come from Assumption 4.1. The fourth and fifth equalities hold due to Assumption 4.9. The above equalities yield that $E[\overline{Y}_{ij}(0,\alpha_0) - \overline{Y}_{ij}(1,\alpha_0)] = E[\overline{Y}_{ij}(0,\alpha_0) - \overline{Y}_{ij}(1,\alpha_0;V_i,R_{ij})|Z_{ij}(r_u) > Z_{ij}(r_l)]$ and thus that $\mathrm{DE}(\alpha_0)$ is identified and given by the expression in (4.6).

To see that $\mathrm{IE}(\alpha_0,\alpha_1)$ is identified and equal to the estimand in (4.6) note that

$$\begin{aligned}
&E[Y_{ij}^w(\alpha_0;V_i,R_{ij}) - Y_{ij}^w(\alpha_1;V_i,R_{ij})|R_{ij}=r_l] \\
=&\sum_{z=0}^{1} E[(Y_{ij}^w(\alpha_0;V_i,R_{ij}) - Y_{ij}^w(\alpha_1;V_i,R_{ij}))I[Z(r_l)=z]|R_{ij}=r_l] \\
=&\{E[\overline{Y}_{ij}(1,\alpha_0) - \overline{Y}_{ij}(0,\alpha_0)|Z=1,R=r_l] \\
&- E[\overline{Y}_{ij}(1,\alpha_1) - \overline{Y}_{ij}(0,\alpha_1)|Z=1,R=r_l]\}E[Z_{ij}|R_{ij}=r_l] + \mathrm{IE}(\alpha_0,\alpha_1). \qquad \text{(B.4.2)}
\end{aligned}$$

And similarly,

$$\begin{aligned}
&E[Y_{ij}^w(\alpha_0;V_i,R_{ij}) - Y_{ij}^w(\alpha_1;V_i,R_{ij})|R_{ij}=r_u] \\
=&\sum_{z=0}^{1} E[(Y_{ij}^w(\alpha_0;V_i,R_{ij}) - Y_{ij}^w(\alpha_1;V_i,R_{ij}))I[Z(r_l)=z]|R_{ij}=r_u] \\
=&\{E[\overline{Y}_{ij}(1,\alpha_0) - \overline{Y}_{ij}(0,\alpha_0)|Z=1,R=r_u] \\
&- E[\overline{Y}_{ij}(1,\alpha_1) - \overline{Y}_{ij}(0,\alpha_1)|Z=1,R=r_u]\}E[Z_{ij}|R_{ij}=r_u] + \mathrm{IE}(\alpha_0,\alpha_1) \\
=&\{E[\overline{Y}_{ij}(1,\alpha_0) - \overline{Y}_{ij}(0,\alpha_0)|Z=1,R=r_l] \\
&- E[\overline{Y}_{ij}(1,\alpha_1) - \overline{Y}_{ij}(0,\alpha_1)|Z=1,R=r_l]\}E[Z_{ij}|R_{ij}=r_u] + \mathrm{IE}(\alpha_0,\alpha_1). \qquad \text{(B.4.3)}
\end{aligned}$$

Solving (B.4.2) for $\{E[\overline{Y}_{ij}(1,\alpha_0) - \overline{Y}_{ij}(0,\alpha_0)|Z=1,R=r_l] - E[\overline{Y}_{ij}(1,\alpha_1) - \overline{Y}_{ij}(0,\alpha_1)|Z=1,R=r_l]\}E[Z_{ij}|R_{ij}=r_u]$ and plugging this into (B.4.3) and solving for $\mathrm{IE}(\alpha_0,\alpha_1)$ yields the estimand in (4.6) which is composed of identifiable quantities.

BIBLIOGRAPHY

Aalen, O. O. and Johansen, S. (1978), "An empirical transition matrix for non-homogeneous Markov chains based on censored observations," *Scandinavian Journal of Statistics*, 5, 141–150.

Abbring, J. and van den Berg, G. (2005), "Social experiments and instrumental variables with duration outcomes," IFS Working Papers W05/19, Institute for Fiscal Studies.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1995), *Statistical Models Based on Counting Processes (Springer Series in Statistics)*, Springer, 2nd ed.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996), "Identification of causal effects using instrumental variables," *J. Am. Stat. Assoc.*, 91, 444–455.

Armstrong, C. S., Guay, W. R., and Weber, J. P. (2010), "The role of information and financial reporting in corporate governance and debt contracting," *Journal of Accounting and Economics*, 50, 179–234.

Aronow, P. M. and Samii, C. (2011), "Estimating average causal effects under general interference," For presentation at the NYU Devt Economics workshop.

Baker, S. G. (1998), "Analysis of Survival Data from a Randomized Trial with All-or-None Compliance: Estimating the Cost-Effectiveness of a Cancer Screening Program," *Journal of the American Statistical Association*, 93, 929–934.

Baker, S. G. and Lindeman, K. S. (1994), "The paired availability design: A proposal for evaluating epidural analgesia during labor," *Statist. Med.*, 13, 2269–2278.

Balke, A. and Pearl, J. (1993), "Nonparametric bounds on causal effects from partial compliance data," Tech. rep., University of California, Los Angeles.

— (1997), "Bounds on treatment effects from studies with imperfect compliance," *J. Am. Stat. Assoc.*, 92, 1171–1177.

Bowers, J., Fredrickson, M., and Panagopoulos, C. (2012), "Reasoning about interference between units," Tech. rep., University of Illinois at Urbana-Champaign.

Braun, T. M. and Yuan, Z. (2007), "Comparing the small sample performance of several variance estimators under competing risks," *Statistics in medicine*, 26, 1170–1180.

Brookhart, M. A. and Schneeweiss, S. (2007), "Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results," *The International Journal of Biostatistics*, 3, 1–23, article 14.

Brumback, B. A., Hernán, M. A., Haneuse, S. J. P. A., and Robins, J. M. (2004), "Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures," *Statistics in Medicine*, 23, 749–767.

Bugni, F. A. (2010), "Bootstrap inference in partially identified models defined by moment inequalities: coverage of the identified set," *Econometrica*, 78, 735–753.

Cai, Z., Kuroki, M., Pearl, J., and Tian, J. (2008), "Bounds on direct effects in the presence of confounded intermediate variables," *Biometrics*, 64, 695–701.

Cain, L. E., Cole, S. R., Greenland, S., Brown, T. T., Chmiel, J. S., Kingsley, L., and Detels, R. (2009), "Effect of highly active antiretroviral therapy on incident AIDS using calendar period as an instrumental variable." *Am J Epidemiol*, 169, 1124–1132.

Chasela, C. S., Hudgens, M. G., Jamieson, D. J., Kayira, D., Hosseinipour, M. C., Kourtis, A. P., Martinson, F., Tegha, G., Knight, R. J., Ahmed, Y. I., et al. (2010), "Maternal or infant antiretroviral drugs to reduce HIV-1 transmission," *New England Journal of Medicine*, 362, 2271–2281.

Chernozhukov, V., Hong, H., and Tamer, E. (2007), "Estimation and confidence regions for parameter sets in econometric models," *Econometrica*, 75, 1243–1284.

Chiburis, R. C. (2010), "Semiparametric bounds on treatment effects," *Journal of Econometrics*, 159, 267 – 275.

Chickering, D. M. and Pearl, J. (1996), "A clinician's tool for analyzing non-compliance," in *AAAI-96 Proceedings*, AAAI Press, pp. 1269–1276.

Cole, S. R. and Frangakis, C. E. (2009), "The consistency statement in causal inference: a definition or an assumption?" *Epidemiology*, 20, 3–5.

Cole, S. R., Hernán, M. A., Margolick, J. B., Cohen, M. H., and Robins, J. M. (2005), "Marginal structural models for estimating the effect of highly active antiretroviral therapy initiation on CD4 cell count." *Am. J. Epidemiol.*, 162, 471–478.

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, E. L. (1959), "Smoking and lung cancer: recent evidence and a discussion of some questions," *J. Natl. Cancer Inst.*, 22, 173–203.

Cox, D. R. (1958), *Planning of experiments*, Wiley series in probability and mathematical statistics: Applied probability and statistics, Wiley.

Cuzick, J., Sasieni, P., Myles, J., and Tyrer, J. (2007), "Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination," *J. R. Stat. Soc.*, 69, 565–588.

Dawid, A. P. (2003), "Causal inference using influence diagrams: The problem of partial compliance (with Discussion)," in *Highly Structured Stochastic Systems*, eds. Green, P., Hjort, N., and Richardson, S., Oxford University Press, no. 2, pp. 45–81.

Frangakis, C. E. and Rubin, D. B. (2002), "Principal stratification in causal inference," *Biometrics*, 58, 21–29.

Gaynor, J. J., Feuer, E. J., Tan, C. C., Wu, D. H., Little, C. R., Straus, D. J., Clarkson, B. D., and Brennan, M. F. (1993), "On the use of cause-specific failure and conditional failure probabilities: examples from clinical oncology data," *Journal of the American Statistical Association*, 88, 400–409.

Gilbert, P. B., Bosch, R. J., and Hudgens, M. G. (2003), "Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials," *Biometrics*, 59, 531–541.

Grilli, L. and Mealli, F. (2008), "Nonparametric bounds on the causal effect university studies on job opportunities using principal stratification," *J. Educ. Behav. Stat.*, 33, 111–130.

Gustafson, P. (2010), "Bayesian inference for partially identified models," *Int. J. Biostatistics*, 6, 1–18.

Gustafson, P., McCandless, L. C., Levy, A. R., and Richardson, S. (2010), "Simplified Bayesian sensitivity analysis for mismeasured and unobserved confounders," *Biometrics*, 66, 1129–1137.

Hafeman, D. M. (2011), "Confounding of indirect effects: a sensitivity analysis exploring the range of bias due to a cause common to both the mediator and the outcome," *American Journal of Epidemiology*, 174, 710–717.

Hahn, J., Hausman, J., and Kuersteiner, G. (2004), "Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations," *The Econometrics Journal*, 7, 272–306.

Halloran, M. E. and Struchiner, C. J. (1995), "Causal inference in infectious diseases," *Epidemiology Cambridge Mass*, 6, 142–151.

Hausman, J. A., Newey, W. K., Woutersen, T., Chao, J. C., and Swanson, N. R. (2012), "Instrumental variable estimation with heteroskedasticity and many instruments," *Quantitative Economics*, 3, 211–255.

Heckman, J. J. (2001), "Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture," *Journal of Political Economy*, 109, 673–748.

Heller, R., Rosenbaum, P. R., and Small, D. S. (2009), "Split samples and design sensitivity in observational studies," *J. Am. Stat. Assoc.*, 104, 1090–1101.

Hernán, M. and Robins, J. (2006), "Instruments for causal inference: An epidemiologist's dream?" *Epidemiology*, 17, 360–372.

Hernán, M. A. and Robins, J. M. (1999), "Assessing the sensitivity of regression results to unmeasured confounders in observational studies [letter]," *Biometrics*, 55, 1316–1317.

Holland, P. W. (1986), "Statistics and causal inference," *Journal of the American Statistical Association*, 81, 945–960.

Hong, G. and Raudenbush, S. W. (2006), "Evaluating kindergarten retention policy," *Journal of the American Statistical Association*, 101, 901–910.

Horowitz, J. L. and Manski, C. F. (2000), "Nonparametric analysis of randomized experiments with missing covariate and outcome data," *Journal of the American Statistical Association*, 95, 77–84.

— (2006), "Identification and estimation of statistical functionals using incomplete data," *Journal of Econometrics*, 132, 445–459.

Hu, J. C., Williams, S. B., OMalley, A. J., Smith, M. R., Nguyen, P. L., and Keating, N. L. (2012), "Androgen-deprivation therapy for nonmetastatic prostate cancer is associated with an increased risk of peripheral arterial disease and venous thromboembolism," *European Urology*, 61, 1119–1128.

Hudgens, M. G. and Halloran, M. E. (2006), "Causal vaccine effects on binary postinfection outcomes," *J. Am. Stat. Assoc.*, 101, 51–64.

— (2008), "Toward causal inference with interference," *Journal of the American Statistical Association*, 103, 832–842.

Hudgens, M. G., Hoering, A., and Self, S. G. (2003), "On the analysis of viral load endpoints in HIV vaccine trials," *Stat. Med.*, 22, 2281–2298.

Imai, K., Keele, L., and Yamamoto, T. (2010), "Identification, inference and sensitivity analysis for causal mediation effects," *Stat. Sci.*, 25, 51–71.

Imbens, G. and Angrist, J. (1994), "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Imbens, G. W. and Manski, C. F. (2004), "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Joffe, M. (2011), "Principal stratification and attribution prohibition: good ideas taken too far," *The International Journal of Biostatistics*, 7, 1–22.

Kaufman, S., Kaufman, J. S., and MacLehose, R. F. (2009), "Analytic bounds on causal risk differences in directed acyclic graphs involving three observed binary variables," *J. Stat. Plan. Inference*, 139, 3473–3487.

Lee, D. S. (2009), "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 76, 1071–1102.

Lee, M. J. (2005), *Micro-Econometrics for Policy, Program, and Treatment Effects*, Oxford University Press, USA.

Lin, D. Y., Psaty, B. M., and Kronmal, R. A. (1998), "Assessing the sensitivity of regression results to unmeasured confounders in observational studies," *Biometrics*, 54, 948–963.

Loeys, T. and Goetghebeur, E. (2003), "A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance." *Biometrics*, 59, 100–105.

Long, D. M. and Hudgens, M. G. (2013), "Sharpening bounds on principal effects with covariates," *Biometrics*, 69, 812–819.

Manski, C. F. (1990), "Nonparametric bounds on treatment effects," *Am. Econ. Rev.*, 80, 319–323.

— (1997), "Monotone treatment response," *Econometrica*, 65, 1311–1334.

— (2013), "Identification of treatment response with social interactions," *The Econometrics Journal*, 16, S1–S23.

Manski, C. F. and Pepper, J. (2000), "Monotone instrumental variables: with an application to the returns to schooling," *Econometrica*, 68, 997–1010.

Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V., and Klungel, O. H. (2006), "Instrumental variables: application and limitations," *Epidemiology*, 17, 260–267.

McCandless, L. C., Gustafson, P., and Levy, A. (2007), "Bayesian sensitivity analysis for unmeasured confounding in observational studies," *Stat. Med.*, 26, 2331–2347.

Moon, H. and Schorfheide, F. (2012), "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80, 755–782.

Morgan, S. and Winship, C. (2007), *Counterfactuals and Causal Inference*, New York: Cambridge University Press.

Neyman, J. (1923), "Sur les applications de la thar des probabilities aux experiences Agaricales: Essay des principle. Excerpts reprinted (1990) in English," *Statistical Science*, 5, 463–472.

Nie, H., Cheng, J., and Small, D. (2011), "Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring," *Biometrics*, 67, 1397–1405.

Panozzo, C. A., Becker-Dreps, S., Pate, V., Weber, D. J., Funk, M. J., Stürmer, T., and Brookhart, M. A. (2014), "Direct, indirect, total, and overall effectiveness of the rotavirus vaccines for the prevention of gastroenteritis hospitalizations in privately insured US children, 2007–2010," *American Journal of Epidemiology*, 179, 895–909.

Pearl, J. (2001), "Direct and indirect effects," in *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., UAI '01, pp. 411–420.

— (2009), *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University Press, 2nd ed.

— (2010), "On the consistency rule in causal inference: axiom, definition, assumption, or theorem?" *Epidemiology*, 21, 872–875.

— (2011), "Principal stratification: A goal or a tool?" *Int. J. Biostatistics*, 7, 1–14.

Pepe, M. (1991), "Inference for events with dependent risks in multiple endpoint studies," *Journal of the American Statistical Association*, 86, 770–778.

Pepe, M. S. and Fleming, T. R. (1989), "Weighted Kaplan-Meier statistics: a class of distance tests for censored survival data," *Biometrics*, 497–507.

Perez-Heydrich, C., Hudgens, M. G., Halloran, M. E., Clemens, J. D., Ali, M., and Emch, M. E. (2014), "Assessing effects of cholera vaccination in the presence of interference," *Biometrics*.

Preziosi, M. and Halloran, M. E. (2003), "Effects of pertussis vaccination on severity: vaccine efficacy for clinical severity," *Clin. Infect. Dis.*, 37, 772–779.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., Benenson, M., Gurunathan, S., Tartaglia, J., McNeil, J. G., Francis, D. P., Stablein, D., Birx, D. L., Chunsuttiwat, S., Khamboonruang, C., Thongcharoen, P., Robb, M. L., Michael, N. L., Kunasol, P., and Kim, J. H. (2009), "Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand," *N. Engl. J. Med.*, 361, 2209–2220.

Richardson, T. S., Evans, R., and Robins, J. (2011), "Transparent parametrizations of models for potential outcomes," In Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West, eds.) 569–610.

Robins, J. M. (1989), "The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies," *Health Service Research Methodology: A Focus on AIDS*, , 113–159.

— (1994), "Correcting for non-compliance in randomized trials using structural nested mean models," *Commun. Stat.*, 23, 2379–2412.

— (1997), "Non-response models for the analysis of non-monotone non-ignorable missing data," *Stat. Med.*, 16, 21–37.

— (1999), "Association, causation, and marginal structural models," *Synthese*, 121, 151–179.

— (2002), "Comment on "Covariance adjustment in randomized experiments and observational studies"," *Statistical Science*, 17, 309–321.

— (2003), "Semantics of causal DAG models and the identification of direct and indirect effects," *Oxford Statistical Science Series*, , 70–82.

Robins, J. M. and Greenland, S. (1992), "Identifiability and exchangeability for direct and indirect effects," *Epidemiology*, 3, 143–155.

— (1996), "Comment on "Identification of causal effects using instrumental variables" by Angrist, Imbens, and Rubin," *J. Am. Stat. Assoc.*, 91, 456–458.

Robins, J. M. and Richardson, T. S. (2010), "Alternative graphical causal models and the identification of direct effects," in *P. Shrout (Ed.): Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures*, Oxford Univ. Press.

Robins, J. M., Rotnitzky, A., and Scharfstein, D. (1999), "Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference," *Statistical Models in Epidemiology: The Environment and Clinical Trials*, 116, 1–92.

Robins, J. M. and Tsiatis, A. A. (1991), "Correcting for non-compliance in randomized trials using rank preserving structural failure time models," *Communications in Statistics - Theory and Methods*, 20, 2609–2631.

Romano, J. and Shaikh, A. (2008), "Inference for identifiable parameters in partially identified econometric models," *J. Stat. Plan. Inference*, 138, 2786–2807.

Rosenbaum, P. (2007), "Interference between units in randomized experiments," *Journal of the American Statistical Association*, 102, 191–200.

— (2010a), "Design sensitivity and efficiency in observational studies," *J. Am. Stat. Assoc.*, 105, 692–702.

— (2010b), "Evidence factors in observational studies," *Biometrika*, 97, 333–345.

— (2011), "Some approximate evidence factors in observational studies," *J. Am. Stat. Assoc.*, 106, 285–295.

Rosenbaum, P. and Rubin, D. (1983), "Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome," *J. R. Stat. Soc. Series B Stat. Methodol.*, 45, 212–218.

Rosenbaum, P. R. (1999), "Choice as an alternative to control in observational studies," *Statistical Science*, 14, 259–278.

— (2002), *Observational Studies*, New York: Springer, 2nd ed.

Rothman, K. J. (1976), "Causes," *American Journal of Epidemiology*, 104, 587–592.

Rotnitzky, A. and Jemiai, Y. (2003), "Sharp bounds and sensitivity analysis for treatment effects in the presence of censoring by death," in *Harvard Schering-Plough Workshop on Development and Approval of Oncology Drug Products: Impact of Statistics.*

Rubin, D. (1978), "Bayesian inference for causal effects: the role of randomization," *The Annals of Statistics*, 6, 34–58.

— (1980), "Discussion of "Randomization analysis of experimental data in the Fisher randomization test," by D. Basu," *J. Am. Stat. Assoc.*, 75, 591–593.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of educational Psychology*, 66, 688.

— (2000), "Comment on "Causal inference without counterfactuals"," *J. Am. Stat. Assoc.*, 95, 435–437.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. (1999), "Adjusting for nonignorable drop-out using semiparametric nonresponse models," *J. Am. Stat. Assoc.*, 94, 1096–1146.

Schlesselman, J. (1978), "Assessing effects of confounding variables." *Am. J. Epidemiol.*, 108, 3–8.

Shepherd, B. E., Gilbert, P. B., and Mehrotra, D. V. (2007), "Eliciting a counterfactual sensitivity parameter," *Am. Stat.*, 61, 56–63.

Sianesi, B. (2004), "An evaluation of the Swedish system of active labor market programs in the 1990s," *The Review of Economics and Statistics*, 86, 133–155.

Sjölander, A. (2009), "Bounds on natural direct effects in the presence of confounded intermediate variables." *Stat. Med.*, 28, 558–571.

Sobel, M. (2006), "What do randomized studies of housing mobility demonstrate?" *Journal of the American Statistical Association*, 101, 1398–1407.

Stoye, J. (2009), "More on confidence intervals for partially identified parameters," *Econometrica*, 77, 1299–1315.

Tchetgen Tchetgen, E. J., Glymour, M. M., Weuve, J., and Robins, J. (2012a), "A cautionary note on specification of the correlation structure in inverse-probability-weighted estimation for repeated measures." Tech. Rep. 140., Harvard University Biostatistics Working Paper Series.

— (2012b), "Specifying the correlation structure in inverse probability weighting estimation for repeated measures." *Epidemiology*, 23, 644–6.

Tchetgen Tchetgen, E. J. and Vanderweele, T. J. (2012), "On causal inference in the presence of interference," *Statistical Methods in Medical Research*, 21, 55–75.

Todem, D., Fine, J., and Peng, L. (2010), "A global sensitivity test for evaluating statistical hypotheses with non-identifiable models," *Biometrics*, 66, 558–566.

van der Laan, M. J. and Robins, J. M. (2003), *Unified methods for censored longitudinal data and causality*, Springer.

VanderWeele, T. J. (2008), "Sensitivity analysis: distributional assumptions and confounding assumptions." *Biometrics*, 64, 645–649.

— (2010), "Bias formulas for sensitivity analysis for direct and indirect effects," *Epidemiology*, 21, 540–551.

Vanderweele, T. J. (2011), "Controlled direct and mediated effects: definition, identification and bounds," *Scand. J. Stat.*, 38, 551–563.

VanderWeele, T. J. (2011), "Principal stratification - uses and limitations," *Int. J. Biostatistics*, 7, 1–14.

VanderWeele, T. J. and Arah, O. A. (2011), "Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders." *Epidemiology*, 22, 42–52.

VanderWeele, T. J. and Halloran, M. E. (2014), "Interference and sensitivity analysis," *Statistical Science*, In press.

VanderWeele, T. J. and Hernández-Diaz, S. (2011), "Is there a direct effect of pre-eclampsia on cerebral palsy not through preterm birth?" *Paediatric and Perinatal Epidemiology*, 25, 111–115.

VanderWeele, T. J., Mukherjee, B., and Chen, J. (2011), "Sensitivity analysis for interactions under unmeasured confounding." *Stat. Med.*, 31, 2552–2564.

Vansteelandt, G., Goetghebuer, E., Kenward, M. G., and Molenberghs, G. (2006), "Ignorance and uncertainty regions as inferential tools in a sensitivity analysis," *Stat. Sinica*, 16, 953–979.

Vansteelandt, S. and Goetghebeur, E. (2001), "Analyzing the sensitivity of generalized linear models to incomplete outcomes via the IDE algorithm," *Journal of computational and graphical statistics*, 10, 656–672.

Ver Steeg, G. and Galstyan, A. (2010), "Ruling out latent homophily in social networks." in *NIPS workshop on Social Computing*.

Zhang, J. L. and Rubin, D. B. (2003), "Estimation of causal effects via principal stratification when some outcomes are truncated by death," *Journal of Educational and Behavioral Statistics*, 28, 353–368.