CAUSAL INFERENCE AND PRINCIPAL STRATIFICATION: COMPETING RISKS, BOUNDS, AND SURROGATES

Dustin M. Long

A dissertation submitted to the faculty of 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
2012

Approved by:

Dr. Michael G. Hudgens
Dr. Jianwen Cai
Dr. Stephen R. Cole
Dr. Jason P. Fine
Dr. Amy H. Herring

Abstract

DUSTIN M. LONG: Causal Inference and Principal Stratification: Competing Risks, Bounds, and Surrogates
(Under the direction of Dr. Michael G. Hudgens)

Establishing statistical methods for quantifying the effects of interventions to prevent infectious diseases is the overall objective of this research. The principal stratification framework is frequently implemented to make causal comparisons where naive methods fail. For HIV vaccine trials, estimates of the causal effect of vaccine on viral load or post-infection survival is challenging using standard methods because all individuals do not become infected during the trial. In this scenario, the "principal" effect, which is the causal effect within a principal stratum, is the effect of vaccine on viral load for subjects who would be infected during the trial regardless of treatment assignment. Without strong assumptions, the principal effect is not identifiable and usually requires bounding, or sensitivity analysis, of the principal effect often resulting in bounds that are often large and uninformative. Methods for estimating, i.e., bounding, the principal effect of treatment on competing risks outcomes have not been developed. Furthermore, situations where bounds on the principal effect can be improved by using baseline covariates have not been investigated. The principal stratification framework can also be used to determine surrogates of vaccine protection, i.e., biomarkers measured during a trial that are correlated with the desired outcome (infection). Repeated low-dose challenge studies are often used to evaluate potential vaccines. While these studies more accurately mimic exposure, the assessment of the potential surrogates greatly depends on the study design. Evaluation and comparison of different study designs have not

been performed. Therefore, we propose to 1) develop methods to analyze the principal effect of treatment on competing risks outcomes, 2) investigate the improvement of the bounds on the principal effect based on baseline covariates, and 3) evaluate designs of repeated low-dose challenge experiments to assess surrogates of vaccine protection.

To my parents, this work and everything that I have and do is thanks to you.

To Leann, there is not another person I would have been able to do this entire journey with and have them not kill me. You are the reason that I was able to get to this point and the reason I will continue to move forward.

To Charlie, I hope that everything that I have accomplished teaches you that with help, and your mother, you can do anything.

Acknowledgments

This work was supported in part by NIH grants R01-AI085073 and U48-DP001944. Many thanks go to my committee, Drs. Jianwen Cai, Stephen Cole, Jason Fine, and Amy Herring for there helpful comments. My deepest appreciation goes to the Breastfeeding, Antiretoviral, and Nutrition (BAN) study participants and investigators for all their work and access to the data from their study.

This dissertation and my degree would not have been possible with the help of many people, especially those in the Department of Biostatistics. In particular, I would like to thank Melissa Hobgood, who was the first person Leann and I meet at UNC and helped convince us to come here, and provided a friendly ear when the struggles of earning the degree were becoming overwhelming. Nothing written in this dissertation could have been produced without great computing and support which was provided by Scott Zentz and David Hill. I would also like to thank Dr. Jane Monaco, who is the finest teaching mentor anyone could have ask for.

I would not have been able to accomplish this degree without first finishing classes. Those classes were made more tolerable by Jennifer Clark, Annie Green Howard, Beth Horton, Andrea Byrnes, and Matt Wheeler. And of course, Leann Long.

Lastly, my deepest thanks go to Dr. Michael Hudgens. His guidance, knowledge, motivation, and encouragement have been key in all of my work. I often wonder if he had known, contrary to fact, what he was getting into when he hired me for BAN, if he still would have hired me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Literature Review

1.1 Motivating Examples

Mother-to-Child Transmission

Recent research indicates that at least 40% of HIV infections in infants where infection time is known can be attributed to mother-to-child transmission (MTCT) through breastfeeding (Bulterys, Ellington, and Kourtis 2010). In resource limited areas such as sub-Saharan Africa, options in place of breastfeeding, such as replacement feeding, which have made HIV MTCT through breast milk uncommon in developed countries are not viable.

The risk of other diseases such as diarrhea and lower respiratory infections associated with poor water supplies in these settings make formula use unattractive (Bulterys *et al.* 2004). Also, the added cost of animal milk prevents mothers from utilizing that approach to reducing MTCT in resource-limited settings (Mofenson 2009). As such, an important area of research in the prevention of MTCT is identifying effective prophylactic treatment(s) for use on the mother (infant) prior to (during) during breastfeeding. Examples include the KiBS study which assessed the efficacy of a maternal

triple-antiretroviral regimen given to women daily beginning at 34-36 weeks gestation and until 6 months post-partum, the ZEB study which assessed the efficacy of early weaning, the SWEN study which assessed the efficacy of prophylactic therapy given to infants during breastfeeding and the BAN study which assessed the efficacy of prophylactic therapy given to women or infants during breastfeeding (KiBS Study 2011; Kuhn, *et al.* 2008; SWEN Study Team 2008; Chasela *et al.* 2010).

A significant problem with MTCT trial analysis involves early HIV infections. For each of the trials above, randomization occurred at birth or shortly before. Since MTCT of HIV can occur anytime before, at, or after birth (De Cock *et al.* 2000), an infant who tests positive for HIV early in the trial does not have a clear method of transmission. Most MTCT trials are interested in only HIV transmission through breast milk. Thus, if randomization occurs at time 0, a time point $\tau_0 > 0$ is often chosen prior to the beginning of the trial and only randomized infants alive and uninfected at τ_0 are considered for analysis. The removal of these infants from the analysis, while eliminating the potential bias due to other modes of HIV infection, creates that potential for selection bias, since a mechanism other than randomization determined who was included.

A second issue in the analysis of the risk of HIV infection in MTCT trials is the presence of competing risks (Alioum *et al.* 2001), in particular infant death and cessation of breastfeeding prior to HIV infection. Often infants experiencing HIV-free death are treated as right censored in the primary analysis, e.g., when computing the Kaplan-Meier estimator of the cumulative probability of HIV infection (for instance, see Figure 2a of Kumwenda *et al.* 2008). Unfortunately, interpretation of such estimates is not straight-forward (Pepe and Mori 1993; Lawless 2003) and a preferred approach is to

estimate the cumulative incidence of HIV treating death as a competing risk. Similarly, cessation of breastfeeding is also a competing risk of HIV infection because once an infant is weaned, the likelihood of HIV infection from the mother is essentially zero.

Vaccine Development in Macaques

The use of animal models in pre-clinical vaccine trials can reduce the risk, time and cost of later clinical trials involving human subjects by providing precursory evidence of potential risks and efficacy of a candidate vaccine (Girard *et al.* 2011; Shedlock, Silvestri and Weiner 2009; Koff 2005). While chimpanzees are the only non-human primate that can be infected with HIV-1, research on chimps is unfeasible due to ethical and financial constraints due to expense and endangerment of the species (Shedlock *et al.* 2009; Smith 2002; Nath, Schumann, and Boyer 2000). Thus, the majority of pre-clinical studies of HIV-1 vaccines have used macaques and viral surrogates of HIV, such as simian immunodeficiency viruses (SIVs), as disease progression of SIVs in macaques parallels that of HIV in humans (Girard *et al.* 2011; Shedlock *et al.* 2009). Virus challenges in these pre-clinical trials have previously been administered via a single high-dose intravenous or mucosal inoculation which frequently resulted in near certain infection of all animals (Hudgens *et al.* 2009).

The appeal of single high-dose challenge studies is that high infection rates create greater chance of observing vaccine efficacy, however this type of trial does not mimic real exposure. Individuals are more likely to have repeated exposure to low doses of the disease in non-trial settings implying the high-dose challenge studies are overestimating infections probabilities. For example, the high infection in high-dose challenge studies do not parallel the low probability of HIV transmission per heterosexual sex

act, estimated to < 0.01 in various studies of different populations (Gray *et al.* 2001; Boily *et al.* 2009). It is doubtful that vaccines are equally efficacious against high-dose and low-dose challenges, i.e., the vaccine may create enough protection to prevent lower doses from causing infections while a high dose could overwhelm the host immune response. This implies that potential vaccines that would be efficacious against low-dose challenges may be rejected due to poor effectiveness in the high-dose challenge studies (Kim *et al.* 2006).

1.2 Causal Inference

1.2.1 Introduction

Determining the cause(s) of an outcome is the aim of most public health studies, both observational and experimental. We wish to say with certainty that event A causes event B, or that the absence of event A causes the absence of event B, e.g., vaccination prevents disease. This idea of causation implies a certain "causal pathway" where we can identify which event(s) cause the other (Rothman 1976). Most researchers desire to perform and analyze these studies with simple or conventional methods. In many cases, causation cannot be assessed from either type of research using conventional methods due to mitigating circumstances. In observational studies, a lack of temporality or unmeasurable variables often hinder the ability to determine causation without sophisticated methods or strong assumptions (Rubin 1974). While most randomized studies are designed to answer, and are able to answer, causal questions using simple approaches, situations can arise that would prevent these naive methods from determining causation, i.e., the BAN study. When these conventional methods fail, methods and frameworks are developed to determine causation. The general term for this type

of research is causal inference.

One quantity of interest in causal inference is the causal effect (CE) or the effect of an intervention on some outcome of interest. For example, the causal effect of angiotensin-converting enzyme (ACE) inhibitor use on the prevention of a second heart attack is the ability of the ACE inhibitor to prolong the time to a second heart attack. This can be quantified as a difference of survival times (e.g. average time to second heart attack while using an ACE inhibitor minus the average time to second heart attack when not using an ACE inhibitor). In infection prevention studies, the causal effect of vaccine is measured by vaccine efficacy (VE), defined as the reduction of disease incidence within vaccinated individuals versus disease incidence in unvaccinated individuals (Halloran, Longini, Struchiner 1999). Inference about these effects can be made using a number of modeling techniques including four major types for health-sciences research: graphical models (causal diagrams), potential outcome models, sufficient component cause models, and structural equations models (Greenland and Brumback 2002).

Using the potential outcomes or counterfactual approach, suppose n individuals are observed or selected and each individual is randomly assigned one of two treatments. Without loss of generality, assume that there are two possible treatments and let $Z_i = 0$ if subject i is assigned treatment 0 (control) and $Z_i = 1$ if assigned treatment 1 (active treatment). A necessary condition for estimating causal effects is that the possibility for each individual to receive any of the treatments exists. This condition allows individuals to have specific outcomes had they been assigned a specific treatment, i.e., individuals have potential outcomes respective to each treatment. If we let $Y_i(Z_i)$ be some outcome for individual i when assigned treatment Z_i, the potential outcomes under our scenario are $Y_i(0)$ and $Y_i(1)$. These potential outcomes are assumed to be

fixed and observable if assigned the respective treatment. This allows us to manipulate which potential outcome is observed by changing the treatment assigned since there is "no causation without manipulation"(Holland 1986). The major drawback to this approach is that we can only observe one of the potential outcomes, since an individual can only be assigned one treatment. The other potential outcome is termed as a counterfactual, since it is the outcome that is contrary to fact. Thus, we define the observed outcome as $Y_i^{obs} = Z_i Y_i(1) + (1 - Z_i)Y_i(0)$ (Cole and Frangakis 2009). A standard assumption states that the treatment assignment of individual i does not affect the potential outcomes of other individuals (i.e., there is no interference) and there are not multiple forms of treatment, i.e., the stable unit treatment value assumption (SUTVA) holds (Rubin 1980).

1.2.2 Principal Stratification

The principal stratification framework was developed by Frangakis and Rubin (2002) to provide causal interpretations where standard procedures would not allow. Developed using the potential outcomes approach (Rubin 1974, Neyman *et al.* 1990), principal strata are defined by the potential outcomes of a variable measured post-randomization or post-treatment. Two goals of principal stratification are to reduce post-treatment selection bias created by making causal comparisons of outcomes that do not have a common set of units or individuals and identify principal surrogates.

In most trials of infectious disease prevention the post-randomization variable of interest is infection status at some predetermined time τ_0 with the outcome measured in only those infected (viral loads) or uninfected (time to infection) at τ_0. In these cases, the principal strata are defined by the potential infection status at τ_0 under all

possible treatment assignments. Let $S_i(Z_i)$ be individual i's potential infection status at τ_0 when given treatment Z_i. In the simple case of only one treatment group and one control group, there are four principal strata; always infected (AI) ($S_i(0) = S_i(1) = 1$), protected ($S_i(0) = 1, S_i(1) = 0$), never infected (NI) ($S_i(0) = S_i(1) = 0$), and harmed ($S_i(0) = 0, S_i(1) = 0$). For example, a person who is assigned $Z_i = 0$ and becomes infected at τ_0 ($S_i = 1$), they are either in the AI or the protected strata.

The principal effect is the causal comparison of treatment within the principal strata of interest. When the principal strata are non-empty, principal effects are not identifiable from the data without strong assumptions. The monotonicity assumption states that the treatment does no harm to patients or more formally

$$S_i(0) \geq S_i(1), \ \forall i. \tag{1.1}$$

Under monotonicity and SUTVA, one half of the causal effect of interest can be identified from the data, since there is not a harmed principal strata. While the role principal stratification in research has caused recent debate (Pearl 2011; Gilbert, Hudgens, and Wolfson 2011), principal stratification can be used to identify causal effects under scenarios where they are otherwise impossible.

For example, in HIV vaccine trials where viral load post-infection is of importance, the AI principal strata is of interest. Here the causal comparison would be the difference in average viral load when taking control versus treatment among individuals who would have been infected regardless of treatment assignment. A fundamental problem with this approach involves principal strata membership. Since only one of the two potential outcomes $S_i(0)$ and $S_i(1)$ is observed, an individuals principal strata membership is unidentifiable from the data without possibly strong assumptions.

1.2.3 Sensitivity Analysis and Bounds Within Principal Strata

Assuming monotonicity in an HIV vaccine trial, all vaccinated individuals are in the AI strata. On the other hand, individuals assigned control who become infected are a mixture of individuals from the AI and protected principal strata. Thus the distribution of potential outcomes when assigned control in the AI strata is not identifiable from the observed data. However, the proportion of infected controls who are in the AI strata is identifiable. Let $VE = \Pr[S_i(1) = 0 | S_i(0) = 1]$ implying that $1 - VE = \Pr[S_i(1) = 1 | S_i(0) = 1]$. Then, $1 - VE$ is the proportion we seek. We can think of $1 - VE$ as a measure of potential selection bias (Hudgens, Hoering, and Self 2003 (HHS)). HHS developed a framework where stochastic lower and upper bounds of this causal effect can be estimated corresponding to extreme selection bias models. Specifically, let $F_C(y)$ be the cumulative distribution function (CDF) of viral load among infected control participants such that the causal effect of interest is $CE = F_T^{AI}(y) - F_C^{AI}(y)$. Then,

$$F_C(y) = (1 - VE)F_C^{AI}(y) + (VE)F_C^{protected}(y), \tag{1.2}$$

where $F_C^{AI}(y)$ ($F_C^{protected}(y)$) is the CDF of viral load for infected vaccinees in the AI(protected) principal strata. Under monotonicity, $F_V^{AI}(y) = F_V(y)$ and is thus identifiable from the observed data. For what HHS calls the extreme lower selection bias model, the control component of the causal effect is

$$F_{C,LB}^{AI}(y) = \max\left\{\frac{F_C(y) - VE}{1 - VE}, 0\right\}.$$

The extreme upper bound is

$$F_{C,UB}^{AI}(y) = \min\left\{\frac{F_C(y)}{1 - VE}, 1\right\}.$$

They then present simulation results demonstrating the performance of hypothesis tests using these bounds. Since $1-VE$ is only a measure of potential selection bias, the true bias is likely not as extreme as the bounds presented in HHS.

Building upon HHS, Gilbert, Bosch, and Hudgens (2003)(GBH) developed sensitivity analysis for a continuous outcome based on the mixing equation (1.2). With Z_i, $S_i(Z_i)$, $Y_i(Z_i)$, and the principal strata defined as before, they assumed a sensitivity parameter, β, that allowed for selection bias models between the extreme cases presented in HHS. Recalculating (1.2), using the probability density function, yields

$$f_C^{AI}(y) = W^{-1} w(y) f_C(y),$$

where $w(y) = \Pr[S_i(1) = 1 | Y_i(0) = y, S_i(0) = 1]$ and $W = \int_{-\infty}^{\infty} w(x) f_C(x) dx = 1 - VE$. Assuming a logistic relationship between β and $1 - VE$ gives $w(y) = w(y|\alpha, \beta) = \exp(\alpha + \beta y)/(1 + \exp(\alpha + \beta y))$. This results in

$$F_C^{AI}(y|\beta) = (1 - VE)^{-1} \int_{-\infty}^{y} \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} dF_C(x), \tag{1.3}$$

where for a fixed β, α is the solution to the equation $F_C(y|\beta) = 1$. Similar to HHS, GBH created test statistics and hypothesis tests based on the selection models with critical values computed using the "controls only" approach. Simulations were conducted under different selection bias models with varying values of β. They showed that when presuming the correct value of β, their tests had correct Type I error and decent power, independent of the value of VE assumed. However, an incorrectly specified β yielded poor performance. Their suggestion was to perform sensitivity analysis across a continuously indexed range of βs. Gains in power or precision may be achieved by restricting the range of β based on prior information elicited from subject matter

experts (Scharfstein *et al.* 2006; Shepherd, Gilbert, Mehrotra 2007). Hudgens and Halloran (2006) extended this work to a binary outcome.

A method for sensitivity analysis within principal strata with a time-to-event outcome was developed by Shepherd, Gilbert, and Lumley (2007)(SGL). Consider the same setup as GBH but instead of a continuous potential outcome $Y_i(Z_i)$ we have a time-to-event outcome $T_i(Z_i)$. Allowing for potential censoring times, the observed outcome for an individual who become infected is the pair $Y_i^{obs} = \min(T_i, C_i)$ and $\Delta_i = I(Y_i = T_i)$, where $I()$ is the usual indicator function. Let $F_z(t)$ be the CDF of failure times for subjects randomized to the group $z = 0, 1$. The causal effect of interest is $SCE(t) = F_0^{AI}(t) - F_1^{AI}(t)$, the survival causal effect in the AI stratum at time t. Assuming the same mixing equation (1.2), SGL proposed nonparametric extreme selection models equivalent to HHS and GBH with similar upper and lower bounds. SGL's extreme bounds are

$$F_{0,LB}^{AI}(t) = \max\left\{\frac{F_0(t) - VE}{1 - VE}, 0\right\}$$

and

$$F_{0,UB}^{AI}(t) = \min\left\{\frac{F_0(t)}{1 - VE}, 1\right\},$$

which mimic those in HHS. Following GBH, they assume a semiparametric selection bias model with logistic weighting that allow for bias models between the nonparametric extreme models. In fact, SGL's selection bias model is found by replacing $F_C(y)$, the CDF of the continuous outcome Y_i for the control group, in (1.3) with $F_0(t)$. It requires an additional set of assumptions to estimate SGL's model in practice, since $1 - VE = \int_0^\infty w(x) dF_0(x) dx$ requires knowledge of $F_0(t)$ after time τ_1, the end of the trial follow-up period, and a solution for α for a fixed β. For simplicity, they assume

that the logistic weights are constant after τ_1, which makes α identifiable from the data. For both types of models, SGL present analytic variance estimates for each estimator of the causal effect. SGL also present the setup for a parametric model but focus on performance of the non- and semiparametric models. Their simulations show both types of model have minimal bias when the correct β is specified. For most cases, confidence intervals based on their estimators have proper coverage independent of the choice of β.

1.2.4 Covariates and Principal Stratification

Frangakis and Rubin (2002) formed the principal stratification framework assuming that the analysis was performed within cells defined by baseline covariates. Shepherd *et al.* (2006) developed sensitivity analysis of the causal effect conditional on baseline covariates in a fully parametric setting. Jemiai *et al.* (2007) expanded the methods of Shepherd *et al.* (2006) allowing for semiparametric estimation. There are other examples of principal stratification within levels of baseline covariates, e.g., Sjölander *et al.* (2009).

Grilli and Mealli (2008) presented nonparametric unadjusted bounds on the causal effect, CE, within principal strata under a number of different assumptions. They suggest that these unadjusted bounds for CE can be improved by creating bounds within cells defined by a baseline covariate and then recovering then adjusting the bounds through a weighted average. Let X be a baseline covariate and CE_x be the causal effect conditional on $X = x$ such that $CE = \sum CE_x \Pr[X = x]$. Let $[CE_x^l, CE_x^u]$ be the bounds on CE_x. The adjusted bounds for CE would be $CE_X^l = \sum CE_x^l \Pr[X = x]$ and $CE_X^u = \sum CE_x^u \Pr[X = x]$. Grilli and Mealli (2008) performed this method on

data from an employment study with mixed results. For some choices of X the adjusted bounds of CE had improvement on only one side of the unadjusted bounds, i.e., $CE^l_X > CE^l$ and $CE^u_X > CE^u$ so improvement was seen on only one side of the bounds. Other choices of X actually worsened the bounds. The reason for only partial improvement was not addressed.

Using baseline covariates to model the probability of being in a certain principal strata or "principal score" was first established in Hill, Waldfogel, and Brooks-Gunn (2002). The process outlined within was more fully developed by Jo and Stuart (2009). Their idea was to model the probability that an individual was a 'complier' using baseline covariates. In the compliance literature the basic principal strata are defined by a patients treatment assignment (Z_i) and actual use of treatment ($S_i(Z_i)$). Compliers are subjects who would take the treatment when assigned treatment and would not take the treatment when assigned control, i.e., $S_i(0) = 0,\ S_i(1) = 1$. Jo and Stuart assume that no patient assigned control has access to treatment which removes the possibility of two principal strata, defiers ($S_i(0) = 1,\ S_i(1) = 0$) and always-takers ($S_i(0) = S_i(1) = 1$). That implies that all patients are either compliers or never-takers ($S_i(0) = S_i(1) = 0$). Thus, the principal score is the probability that an individual is a complier. An analogous 'compliance score' was created previously by Follmann (2000), which while mathematically equivalent, was created before, therefore outside, the principal stratification framework. In both Jo and Stuart (2009) and Follmann (2000), a two-step model is performed by first estimating the principal scores using logistic regression then using a parametric model for the outcome estimated by matching or weighting by the principal scores.

A similar notion exists within the 'truncation by death' literature. Zhang, Rubin,

and Mealli (2009) developed likelihood based methods that estimated the principal scores and the outcome model simultaneously. Using HIV vaccine trial notation, let G_i represent the principal strata membership for subject i, i.e., $G_i = (S_i(0), S_i(1))$. In general, $\pi_{S_i(0)S_i(1)} = \Pr[G_i = g]$, e.g., π_{11} is the probability that a subject is in the AI strata. Their model allows for these probabilities to vary by subject.

1.2.5 Principal Surrogates

In HIV vaccine trials, participants cannot be followed forever, implying the primary clinical outcome of HIV infection cannot be measured in all participants, preventing an accurate measurement of VE. A surrogate endpoint that is predictive of the primary outcome that can be measured for all subjects within the trial would be useful.

While not the first to use the term, Prentice (1989) laid the foundation for the current surrogate literature. Let Z_i be treatment assignment $Z_i = 0, 1$. Let $S_i(Z_i)$ be a binary post-randomization variable and assume monotonicity, i.e., $S_i(0) \geq S_i(1)$ for all i. Additionally, let $Y_i(Z_i)$ be the outcome of interest. For Prentice, for S_i to be a surrogate endpoint for the true outcome Y_i under treatment Z_i, Y_i would be independent of Z_i conditional on S_i. Frangakis and Rubin (2002) term this a 'statistical' surrogate and demonstrate how a statistical surrogate can exist while not having a causal relationship to the outcome. They state that a variable is a 'principal' surrogate if all comparisons of the outcome within all strata where $S_i(0) = S_i(1)$ results in equality, i.e., there is no effect of treatment on the outcome given $S_i(0) = S_i(1) = s$ for all s. Let $S_i = 1$ indicate if a person has a high potential surrogate value ,$S_i = 0$ otherwise, and let $f(x)$ be some measurement of interest, i.e. risk of infection. S_i would be a principal surrogate if $f(Y_i(0)|S_i(0) = S_i(1) = s) - f(Y_i(1)|S_i(0) = S_i(1) = s) = 0$ for $s = 0, 1$. However, since principal strata membership is unknown, only $f(Y_i(z)|S_i(z) = s)$, $z = 0, 1$ can be identified from the data, which allows a variable to

meet Prentice's criteria (a statistical surrogate) but would not be a principal surrogate. Likewise, a variable can be a principal surrogate but not a statistical surrogate. Using Example 1 in Table 1.1, S_i is a principal surrogate but is not a statistical surrogate since the treatment effect is 0 in the strata where $S_i(0) = S_i(1)$ but the treatment effect is nonzero when $S_i = 0$ in the observed data. Example 2 illustrates the reverse situation.

Since Frangakis and Rubin (2002) defined principal surrogacy, researchers have developed further criteria, especially in the vaccine literature. A vaccine-induced immune response (surrogate endpoint) that is predictive of infection, a correlate of protection, is the 'holy grail' of vaccine trials (Halloran, Longini, and Struchiner 2009). Qin *et al.* (2007) defined three different levels of confidence in a biomarker to be a correlate of protection. The first level indicates a measurement that predicts the primary endpoint in a particular population and is called a correlate of risk. The higher levels are called surrogates of protection (SoP) and have two levels of generalizability. A SoP is a correlate of risk that predicts a vaccines efficacy based on comparisons between the vaccinated and unvaccinated subjects immunological measurements. The level 1 SoP, or specific SoP (Sadoff and Wittes 2007), can predict VE in the same setting in which it was identified (same vaccine, population, etc). The level 2 SoP can predict VE in a variety of settings and is termed a general SoP or a bridging SoP (Pearl and Barenboim 2011). While a general SoP is desirable, a specific SoP is still of scientific use (Gilbert, Hudgens, and Wolfson 2011).

Joffe and Greene (2009) summarized four competing frameworks that have been used for evaluating potential surrogate endpoints. The first two frameworks considered are under the causal-effects paradigm, where the effect of treatment on the surrogate and the effect of the surrogate on the clinical outcome are used to predict the effect of

treatment on the clinical outcome. The first, developed by Prentice (1989), defines a surrogate S as a valid surrogate when a hypothesis test of no relationship between S and treatment is also a valid test of no relationship between treatment and the clinical outcome. He proposed three criteria where when all criteria are met the previous holds and S is a 'true surrogate'. The second framework involves modeling the direct and indirect effects of the surrogate on the clinical outcome. In general, both causal-effects paradigm frameworks require direct manipulation of S to determine the causal effects, which makes determination of proxy surrogates, a measurable related variable to an unmeasurable S, possible. The second set of two frameworks are under a causal-association paradigm, where the effect of treatment on the surrogate is associated with the effect of treatment on the outcome. The first design in the causal-association paradigm uses meta-analysis to examine the effect of a randomized treatment on the clinical outcome across a number of studies. The second is principal surrogacy as described previously.

Gilbert and Hudgens (2008) developed a principal stratification estimand they call the "causal effect predictiveness (CEP) surface" to measure a biomarkers accuracy as a specific surrogate, or surrogate value. Built upon Frangakis and Rubin (2002), and Follmann (2006), CEP is conditional on not yet having the primary outcome under either treatment assignment at time τ, the time that the biomarker is collected. Formally, CEP is defined as

$$\begin{aligned} CEP(s_1, s_0) &= P(Y(0) = 1 | S(1) = s_1, S(0) = s_0) - P(Y(1) = 1 | S(1) = s_1, S(0) = s_0) \\ &= E(Y(0) - Y(1) | S(1) = s_1, S(0) = s_0) \end{aligned}$$

(or some different contrast), where $Y(Z)$ $(S(Z))$ is primary binary outcome (potential biomarker) if assigned treatment Z, $Z = 0$ is control and $Z = 1$ is treatment (Gilbert, Hudgens, and Wolfson 2011). They argue that one can use $CEP(s_1, s_0)$ to determine the association between $S(1) - S(0)$ and $Y(1) - Y(0)$, the goal of predicting VE, since previous estimands only measured the association between $S(1)$ and $Y(1)$ (Gilbert and Hudgens 2008; Gilbert, Hudgens, and Wolfson 2011). Gilbert and Hudgens (2008) also consider summary functions of the associative and dissociative effects of a biomarker, namely the expected associative effect (EAE) and the expected dissociative effect (EDE). They also define the proportion associative (PA) effect by

$$PAE^w = |EAE^w|/|EDE| + |EAE^w|, \tag{1.4}$$

with the convention that $|0|/(|0| + |0|) = 0.5$. PAE^w is the relative proportion of primary outcome effects for those with and without surrogate effects. Hudgens and Gilbert (2009) assessed the vaccine effects in repeated-low-dose experiments. They suggest that PAE, PAE^w from (1.4) with $w(.,.) = 1$, can be used as a summary measure of the surrogate value of S. Using the framework for Follmann (2006), they define the transition probability from uninfected to infected as a probit model which allows PAE to be estimated easily. Huang and Gilbert (2011) developed a method to evaluate the joint surrogacy of multiple biomarkers.

Table 1.1: Examples demonstrate possible scenarios where a biomarker is a statistical surrogate but not a principal surrogate and vice versa.

Average potential post-treatment variable		Average potential outcome		Observed	
$S_i(0)$	$S_i(1)$	$Y_i(0)$	$Y_i(1)$	S_i, Y_i $Z_i = 0$	S_i, Y_i $Z_i = 1$
Example 1: S_i a principal surrogate but not a statistical surrogate					
0	0	10	10	0, 15	0, 10
0	1	20	40		1, 40
1	1	40	40	1, 40	
Example 2: S_i a not principal surrogate but is a statistical surrogate					
0	0	15	20	0, 20	0, 20
0	1	25	25		1, 40
1	1	40	55	1, 40	

Chapter 2

Competing Risks Outcomes Within Principal Strata

2.1 Introduction

Every year approximately 200,000 infants become infected with HIV through breastfeeding; in the absence of treatment, half of these infants will die within two years of birth (WHO 2007; UN AIDS 2007). In clinical trials to prevent MTCT of HIV through breast milk, investigators are often interested in comparing interventions conditional on the infant being alive and uninfected up to a certain time point during the trial (van der Horst *et al.* 2009; Chasela *et al.* 2010; Kilewo *et al.* 2009; Kumwenda *et al.* 2008). Specifically, when randomization occurs at birth (time 0), a time point $\tau_0 > 0$ is often chosen prior to the beginning of the trial and only randomized infants alive and uninfected at τ_0 are considered for analysis. For example, in the Breastfeeding, Antiretroviral, and Nutrition (BAN) study (van der Horst *et al.* 2009; Chasela *et al.* 2010) infants were randomized at birth but the primary analysis included only infants HIV uninfected and alive at $\tau_0 = 2$ weeks. Infants infected prior to 2 weeks were excluded because these transmissions likely occurred in utero or during labor and delivery, whereas the primary objective of the trial was to assess the effects of interventions to

prevent infection due to breast milk. Similar exclusions were made in the primary analysis of the SWEN and PEPI trials (Bedri *et al.* 2008; Kumwenda *et al.* 2008).

There are two aspects of the analysis described above that are the focus of this paper. First, an analysis comparing risk of HIV infection between trial arms among infants who are alive and uninfected at time τ_0 after randomization is subject to selection bias. One method to protect against selection bias in this scenario entails principal stratification (Frangakis and Rubin 2002). Principal stratification uses the potential outcomes of a variable collected post-randomization to define strata of individuals. In the MTCT trial setting, the principal stratum of interest is infants who would be alive and uninfected by time τ_0 under either treatment assignment. Because principal stratum membership is not affected by treatment assignment, comparisons between trial arms within a particular principal stratum are not subject to selection bias. For a recent discussion of the strengths and weaknesses of principal stratification, see Pearl (2011) and subsequent responses such as VanderWeele (2011).

The second aspect in the analysis of the effect of treatment on the risk of HIV infection in MTCT trials is the presence of competing risks (Alioum 2001). In particular, death or weaning prior to HIV transmission are competing risks for HIV infection since these events (death, weaning) can preclude HIV infection from occurring. Likewise, HIV infection precludes the possibility of an HIV-free death or weaning prior to HIV infection. One analytical approach that avoids the complication of competing risks is to use a composite endpoint, such as time until HIV infection or death. Using a composite endpoint simplifies analysis and has the advantage of providing a single measure of the overall effect of treatment. However, such an analysis does not provide inference about whether the treatment is having an effect on the risk of HIV infection, death, or both endpoints. Another common approach in the analysis of MTCT trials is to treat infants experiencing HIV-free death as right censored, e.g., when computing the Kaplan-Meier

estimator of the cumulative probability of HIV infection (for instance, see Figure 2a of Kumwenda *et al.* 2008). It is well known that computing the Kaplan-Meier estimator by right censoring competing events does not in general yield a consistent estimator of the cumulative risk of the event of interest (Tsiatis 1998; Andersen, Abildstrom, and Rosthøj 2002); in the MTCT setting such Kaplan-Meier estimators will tend to overestimate the risk of HIV infection when there is a non-zero probability of death prior to HIV infection. A third approach, adopted in this paper, is to estimate the cumulative incidence functions of each competing event, namely HIV, death, and weaning. The resulting estimates have a straightforward interpretation as the cumulative risk of each event in settings such as the trial where the other events may occur. Contrasts is the estimated risks between trial arms can then be used to assess treatment effects on each of the competing events. While Bekaert, Vansteelandt, and Martens (2010) investigate the effect of time-varying covariates in the presence of competing risks, they assume that there is a potential event time for each competing risk which we will avoid.

Previous work on estimating treatment effects within principal strata has considered binary outcomes (e.g., Hudgens and Halloran 2006), continuous outcomes (e.g., Gilbert *et al.* 2003) and survival outcomes (e.g., Hayden *et al.* (2005) and Shepherd *et al.* (2007)). In this paper we develop methods for estimating treatment effects within principal strata for a survival outcome in the presence of competing risks. In the absence of competing risks the developed methods essentially reduce to those of Shepherd *et al.* (2007). The outline of the remainder of this chapter is as follows. In Section 2 notation and assumptions are discussed. In Section 3 inferential methods for the causal effect of interest are presented. The finite sample performance of the methods are assessed in a simulation study in Section 4. These simulations also illustrate how misleading inferences can arise if selection bias are ignored. In Section 5 the methods are applied to investigate the effect of infant antiretroviral therapy (ART) on the cumulative risk

of HIV infection in the BAN trial. A brief discussion is given in Section 6.

2.2 Notation and Assumptions

Suppose n individuals are randomly assigned one of two treatments, 0 or 1, at baseline (birth or time 0). For $i = 1, \ldots, n$, let $Z_i = 0$ if subject i is assigned treatment 0 and $Z_i = 1$ otherwise. Let $n_0 = \sum(1-Z_i)$ and $n_1 = \sum Z_i$, where here and throughout $\sum = \sum_{i=1}^{n}$. Without loss of generality, assume $Z_i = 0$ corresponds to placebo or control, and $Z_i = 1$ corresponds to active treatment. In the BAN study analysis, $Z_i = 1$ will refer to the infant ART arm and $Z_i = 0$ will refer to the control arm. Suppose the primary objective is to assess the effect of treatment on the time T_i (from baseline) until some particular event occurs. Assume there are k possible causes or types of events and let J_i denote the event type for individual i with $J_i \in \{1, \ldots, k\}$. In the BAN study there are $k = 3$ competing risks: HIV infection ($J_i = 1$), death prior to HIV infection or weaning ($J_i = 2$), or cessation of breastfeeding prior to HIV infection ($J_i = 3$).

Suppose in the analysis of the effect of treatment Z_i on (T_i, J_i) we would like to condition on some binary post-randomization variable S_i (taking on values 0 or 1) measured at some pre-specified post-randomization time $\tau_0 > 0$. For instance, in the analysis of BAN it is desired to assess the effect of treatment in infants alive and uninfected at time τ_0; in this case we let $S_i = 1$ if an infant becomes infected or dies by τ_0 and $S_i = 0$ otherwise. Note for the BAN example that $S_i = I(T_i \leq \tau_0, J_i \leq 2)$ where $I(\cdot)$ is the usual indicator function, however in the methods developed below S_i need not be defined in terms of T_i or J_i.

Define C_i to be a potential right censoring time and assume $\tau_0 \leq C_i$, i.e., no individuals drop out of the study prior to τ_0 such that S_i is always observed. Let τ_1 denote the maximum length of follow-up for the study such that any individual who has not had an event or dropped out of the study by time τ_1 is administratively censored at

that time, i.e., $C_i \leq \tau_1$. Let $Y_i = \min\{T_i, C_i\}$ and $\Delta_i = I(Y_i = T_i)$. Due to censoring, instead of (T_i, C_i, J_i) we only observe $(Y_i, J_i\Delta_i)$; i.e., T_i and J_i are observed if and only if individual i is not right censored.

Let $T_i(z)$ be the potential survival time when assigned treatment z for $z = 0, 1$ such that $T_i = (1 - Z_i)T_i(0) + Z_iT_i(1)$. Define $C_i(z)$, $S_i(z)$, and $J_i(z)$ similarly. Assume the treatment assignment of individual i does not affect the potential outcomes of other individuals (i.e., there is no interference) and there are not multiple forms of treatment, i.e., the stable unit treatment value assumption (SUTVA) holds (Rubin 1980). Let $W_i = (S_i(0), S_i(1), T_i(0), T_i(1), J_i(0),\ J_i(1), C_i(0), C_i(1))$ denote the vector of potential outcomes and $O_i = (Z_i, S_i, Y_i, J_i\Delta_i)$ denote the vector of observable random variables. Assume individuals in the study are a random sample from a larger population such that $W_1, \ldots, W_n$ and $O_1, \ldots, O_n$ are iid copies of W and O respectively.

Principal strata can be defined by sets of individuals with the same potential outcome pair $(S_i(0) = s_0, S_i(1) = s_1)$. Define the never infected (NI) principal stratum to be individuals with $S_i(0) = S_i(1) = 0$, i.e., individuals who would be alive and uninfected at τ_0 regardless of treatment assignment. Similarly define the harmed stratum as those individuals with $S_i(0) = 0, S_i(1) = 1$; the protected stratum as those individuals with $S_i(0) = 1, S_i(1) = 0$; and the doomed stratum as those individuals with $S_i(0) = S_i(1) = 1$. Motivated by MTCT studies of HIV, we focus on drawing inference about causal effects in the NI principal stratum. For example, in the BAN study we are interested in the principal stratum of infants who would be alive and not infected with HIV by $\tau_0 = 2$ weeks under either randomization assignment.

In the presence of competing risks, a quantity of interest is the cumulative incidence function (CIF) or subdistribution function of (T, J). Let $F(t, j) = P(T \leq t, J = j)$ denote the CIF, i.e., the probability of having event j at or before time t. Define the causal estimand of interest to be $CE(t, j) = F_1^{NI}(t, j) - F_0^{NI}(t, j)$ for $t \in [\tau_0, \tau_1]$ where

$F_z^{NI}(t,j) = \Pr[T_i(z) \leq t, J_i(z) = j | S_i(0) = S_i(1) = 0]$ for $z = 0, 1$. In words, $CE(t,j)$ is the difference in the probability of having an event of type j by time t for treatment 0 compared to treatment 1 within the NI principal stratum. For example, in the BAN study (where $j = 1$ corresponds to HIV infection), $CE(28,1)$ is the difference in the probability of HIV infection by 28 weeks between the two study arms among infants who would be alive and HIV negative by τ_0 weeks regardless of treatment assignment. In the analysis of BAN, $CE(28,1)$ was of particular interest because per protocol a primary endpoint of the trial was HIV infection by 28 weeks (van der Horst *et al.* 2009).

To draw inference about $CE(t,j)$ we make the following assumptions:

Assumption 2.1 Independent treatment assignment: $Z_i \perp W_i$

Assumption 2.2 Monotonicity: $S_i(1) \leq S_i(0)$ for all i

Assumption 2.3 Independent censoring: $C_i(z) \perp \{T_i(z), J_i(z), S_i(z)\}$ for $z = 0, 1$

Assumption 2.1 is plausible in randomized clinical trials. Assumption 2.2 is a strong assumption that must be considered carefully and is discussed further in Section 2.5 in the context of the BAN study. Methods not requiring the monotonicity assumption are discussed in Section 6. Assumption 2.3 is a common assumption when analyzing competing risks data. In the infant ART and control arms of BAN, 15% of participants were administratively censored at $\tau_1 = 28$ weeks and 12% were censored at earlier time points due to drop-out from the study prior to week 28.

Under Assumptions 2.1 and 2.2, $Z_i = 0$ and $S_i = 0$ imply $S_i(0) = S_i(1) = 0$; i.e., individuals who are alive and uninfected by τ_0 when assigned control must be members of the NI principal stratum. Letting $F_0(t,j) = \Pr[T_i(0) \leq t, J_i(0) = j | S_i(0) = 0]$, it follows under Assumptions 2.1 – 2.2 that $F_0^{NI}(t,j) = F_0(t,j)$, which is identifiable from the observable data under Assumption 2.3. However $F_1^{NI}(t,j)$ remains unidentifiable under Assumptions 2.1 – 2.3 because individuals who are alive and uninfected by τ_0

when assigned treatment ($Z_i = 1$) are a mixture of individuals from the NI and protected principal strata. In particular, following Gilbert *et al.* (2003)nocitegilbert2003b, one can show

$$F_1(t,j) = \gamma F_1^{NI}(t,j) + (1-\gamma) F_1^{prot}(t,j), \tag{2.1}$$

where $\gamma = \Pr[S_i(0) = 0 | S_i(1) = 0]$ is the probability an individual is uninfected under control given they would be uninfected under treatment, $F_1(t,j) = \Pr[T_i(1) \leq t, J_i(1) = j | S_i(1) = 0]$ and $F_1^{prot}(t,j) = \Pr[T_i(1) \leq t, J_i(1) = j | S_i(0) = 1, S_i(1) = 0]$.

To proceed, one can introduce an additional assumption about the selective effect of conditioning on S_i which renders $F_1^{NI}(t,j)$ identifiable. For example, following Hudgens and Halloran (2006), large-sample upper and lower bounds can be obtained by considering extreme selection bias models. The upper bound selection model is given by assuming either $F_1^{prot}(t,j) = 0$ or $F_1^{NI}(t,j) = 1$, while the lower bound selection model is given by assuming either $F_1^{prot}(t,j) = 1$ or $F_1^{NI}(t,j) = 0$. By (2.1), these models are equivalent to assuming either

$$F_1^{NI}(t,j) = \min\left\{\gamma^{-1} F_1(t,j), 1\right\}, \tag{2.2}$$

or

$$F_1^{NI}(t,j) = \max\left\{\frac{F_1(t,j) - (1-\gamma)}{\gamma}, 0\right\}. \tag{2.3}$$

Estimating $CE(t,j)$ under (2.2) or (2.3) is useful in bounding the estimate of the causal effect above and beyond any possible selective effects induced by conditioning on $S_i = 0$.

The true degree of selection bias may be considerably less than that assumed by (2.2) or (2.3). Therefore, we consider a class of selection models that includes the extreme models above as special cases. Through sensitivity analysis over the entire class (as in Robins *et al.* 2000 and Gilbert *et al.* 2003), the relationship between the assumed degree of selection bias and inference about $CE(t,j)$ can be explored. These selection

models are semiparametric in the sense that no additional restrictions are placed on the distribution of the observable random variables $O_1, \ldots, O_n$ but an unidentifiable parameter (β_j in the model below) is used to quantify the selection bias. One possible selection model is:

Assumption 2.4:

$$\exp(\beta_j) = \frac{F_1^{NI}(t,j)/\{1 - F_1^{NI}(t,j)\}}{F_1^{prot}(t,j)/\{1 - F_1^{prot}(t,j)\}}. \tag{2.4}$$

The parameter β_j equals the log odds ratio of having an event of type j by time t under treatment assignment $z = 1$ in the NI principal stratum versus the protected principal stratum. Note Assumption 2.4 allows for the log odds to differ across event types as indicated by the subscript on β. Also note (2.4) is unverifiable since β_j is not identifiable from the observable data. For fixed β_j, under Assumptions 2.1 – 2.4 $F_1^{NI}(t, j; \beta_j)$ is identifiable from the observable data and $CE(t, j)$ can be estimated as described in Section 2.3 below. The extreme models (2.2) and (2.3) can be viewed as special cases of Assumption 2.4 as $\beta_j \to \infty$ and $\beta_j \to -\infty$. We refer to $\beta_j = 0$ as the no selection bias model because in this case the odds of having an event of type j by time t are the same in the NI and protected principal strata. Sensitivity analysis of inference about $CE(t, j)$ can be conducted by letting β_j range from $-\infty$ to ∞. Gains in power or precision may be achieved by restricting the range of β_j based on prior information about β_j elicited from subject matter experts (Scharfstein *et al.* 2006; Shepherd *et al.* 2007).

2.3 Inference

In this section we first consider nonparametric estimation of $CE(t, j)$ under the extreme selection models (2.2) and (2.3). Then inference for $CE(t, j)$ under the semiparametric selection model (2.4) given some value of β_j is discussed in Section 2.3.2.

The construction of uncertainty intervals about $CE(t,j)$ is considered in Section 2.3.3.

2.3.1 Nonparametric Estimation: Bounds

Under Assumptions 2.1 - 2.3 consistent estimators of $F_1^{NI}(t,j)$ assuming (2.2) or (2.3) are given, respectively, by

$$\hat{F}_1^{NI,up}(t,j) = \min\left\{\hat{\gamma}^{-1}\hat{F}_1(t,j), 1\right\} \text{ and } \hat{F}_1^{NI,low}(t,j) = \max\left\{\frac{\hat{F}_1(t,j) - (1-\hat{\gamma})}{\hat{\gamma}}, 0\right\}, \tag{2.5}$$

where

$$\hat{\gamma} = \min\left\{\frac{\sum(1-S_i)(1-Z_i)/n_0}{\sum(1-S_i)Z_i/n_1}, 1\right\},$$

and $\hat{F}_1(t,j)$ is the Aalen-Johansen estimator (Aalen and Johansen 1978) of $F_1(t,j)$ calculated using $(Y_i, J_i\Delta_i)$ for individuals with $Z_i = 1$ and $S_i = 0$. It can be shown that $\hat{\gamma}$ and $\hat{F}_1(t,j)$ are nonparametric maximum likelihood estimators (NPMLEs) of γ and $F_1(t,j)$. Thus the estimators in (2.5) can be viewed as NPMLEs of $F_1^{NI}(t,j)$. Because Assumptions 2.1 and 2.2 imply $F_0^{NI}(t,j) = F_0(t,j)$, consistent estimators of $CE(t,j)$ assuming either (2.2) or (2.3) are $\widehat{CE}^{up}(t,j) = \hat{F}_1^{NI,up}(t,j) - \hat{F}_0(t,j)$ or $\widehat{CE}^{low}(t,j) = \hat{F}_1^{NI,low}(t,j) - \hat{F}_0(t,j)$, where $\hat{F}_0(t,j)$ is the Aalen-Johansen estimator of $F_0(t,j)$ calculated using $(Y_i, J_i\Delta_i)$ for individuals with $Z_i = S_i = 0$. In the nomenclature of Vansteelandt *et al.* (2006), the interval $[\widehat{CE}^{low}(t,j), \widehat{CE}^{up}(t,j)]$ is an estimated ignorance region of $CE(t,j)$.

If $0 < \gamma < 1$, then $\hat{\gamma}$ is asymptotically normal. The Aalen-Johansen estimators $\hat{F}_z(t,j)$, for $z = 0, 1$ are asymptotically normal assuming $0 < F_z(t,j) < 1$ and certain regularity conditions (Aalen, Borgan, and Gjessing 2008). Therefore, $\hat{F}_1^{NI,up}(t,j)$ is asymptotically normal if, in addition to these conditions,

$$F_1(t,j) < \gamma. \tag{2.6}$$

If (2.6) does not hold, then $\hat{F}_1^{NI,up}(t,j) \xrightarrow{p} 1$ and hence is not asymptotically normal. Under conditions where $\hat{F}_1^{NI,up}(t,j)$ is asymptotically normal, a consistent estimator of the variance of $\hat{F}_1^{NI,up}(t,j)$ is

$$\widehat{\text{var}}\{\hat{F}_1^{NI,up}(t,j)\} = \frac{\widehat{\text{var}}\{\hat{F}_1(t,j)\}}{\hat{\gamma}^2} + \left\{\frac{\hat{F}_1(t,j)}{\hat{\gamma}}\right\}^2 \left(\frac{1}{N_0} - \frac{1}{n_0} + \frac{1}{N_1} - \frac{1}{n_1}\right), \qquad (2.7)$$

where $\widehat{\text{var}}\{\hat{F}_1(t,j)\}$ is a consistent estimator of the variance of $\hat{F}_1(t,j)$ (e.g., see Aalen *et al.* 2008, Section 3.4.5) and $N_z = \sum I(S_i = 0, Z_i = z)$. Similarly $\hat{F}_1^{NI,low}(t,j)$ is asymptotically normal if, in addition to the conditions above,

$$1 - \gamma < F_1(t,j). \qquad (2.8)$$

If (2.8) does not hold, $\hat{F}_1^{NI,low}(t,j) \xrightarrow{p} 0$ and hence is not asymptotically normal. If $\hat{F}_1^{NI,low}(t,j)$ is asymptotically normal, the variance can be consistently estimated by

$$\widehat{\text{var}}\{\hat{F}_1^{NI,low}(t,j)\} = \frac{\widehat{\text{var}}\{\hat{F}_1(t,j)\}}{\hat{\gamma}^2} + \left\{\frac{1-\hat{F}_1(t,j)}{\hat{\gamma}}\right\}^2 \left(\frac{1}{N_0} - \frac{1}{n_0} + \frac{1}{N_1} - \frac{1}{n_1}\right). \qquad (2.9)$$

Derivations of (2.7) and (2.9) are given in the appendix. When (2.6) and (2.8) hold, pointwise Wald-type confidence intervals for $CE(t,j)$ can be constructed in the usual manner. Alternatively, the bootstrap percentile method can be used for computing confidence intervals of $CE(t,j)$. If (2.6) and (2.8) do not hold, then $\hat{F}_1^{NI,up}(t,j) \xrightarrow{p} 1$ and $\hat{F}_1^{NI,low}(t,j) \xrightarrow{p} 0$, i.e., the bounds are non-informative. Note that conditions (2.6) and (2.8) can be assessed based on observed data by comparing $\hat{\gamma}$ and $\hat{F}_1(t,j)$.

2.3.2 Semiparametric Estimation

Under Assumptions 2.1 – 2.4, for fixed β_j a semiparametric estimator of $F_1^{NI}(t,j)$ can be constructed by plugging $\hat{F}_1(t,j)$ and $\hat{\gamma}$ into equation (2.1) and then simultaneously solving (2.1) and (2.4) for $F_1^{NI}(t,j)$. This can be accomplished by expressing $F_1^{prot}(t,j)$ as a function of β_j and $F_1^{NI}(t,j)$ using (2.4), replacing $F_1^{prot}(t,j)$ by this expression in (2.1), and finding the solution to (2.1) using a one-dimensional line search. Define the solution as $\hat{F}_1^{NI}(t,j;\beta_j)$ and let the corresponding estimator of the causal effect be $\widehat{CE}(t,j;\beta_j) = \hat{F}_1^{NI}(t,j;\beta_j) - \hat{F}_0(t,j)$. Without a closed form for $\hat{F}_1^{NI}(t,j;\beta_j)$, confidence intervals of $F_1^{NI}(t,j)$ and $CE(t,j)$ for an assumed value of β_j can be constructed using the bootstrap percentile method; alternatively, Wald-type confidence intervals can be constructed based on bootstrap estimates of $\text{var}\{\hat{F}_1^{NI}(t,j;\beta_j)\}$ and $\text{var}\{\widehat{CE}(t,j;\beta_j)\}$.

Note $\lim_{\beta_j \to \infty} \widehat{CE}(t,j;\beta_j) = \widehat{CE}^{up}(t,j)$ and $\lim_{\beta_j \to -\infty} \widehat{CE}(t,j;\beta_j) = \widehat{CE}^{low}(t,j)$, i.e., the estimators that arise from the extreme selection models (2.2) and (2.3) are special cases of the estimators from the semiparametric bias model (2.4). Under the no selection model $\beta_j = 0$, $\widehat{CE}(t,j;\beta_j) = \hat{F}_1(t,j) - \hat{F}_0(t,j)$, i.e., the causal effect is estimated by the difference in Aalen-Johansen estimators from the two treatment groups as in a standard competing risks analysis. In other words, assuming the no selection model gives rise to a naive or "net" estimator (Frangakis and Rubin 2002) which simply compares subsets of the two randomization groups conditional on being observed HIV free and alive at τ_0.

2.3.3 Uncertainty Regions

The pointwise confidence intervals described in Sections 2.3.1 and 2.3.2 will contain $CE(t,j)$ with the stated coverage probability provided the correct value of β_j is assumed. However, the true value of β_j is not identifiable from the observed data. Therefore, following Vansteelandt *et al.* (2006), it is useful to also construct a $(1-\alpha)100\%$ uncertainty interval which contains $CE(t,j)$ with probability $1-\alpha$ without conditioning on any assumption about the value of β_j. Under the assumptions given in Section 2.3.1 where $\widehat{CE}^{up}(t,j)$ and $\widehat{CE}^{low}(t,j)$ are consistent and asymptotically normal, a large sample $(1-\alpha)100\%$ pointwise uncertainty interval for $CE(t,j)$ is given by

$$[\widehat{CE}^{low}(t,j) - c^*_{\alpha/2}\widehat{\text{var}}\{\widehat{CE}^{low}(t,j)\}^{1/2}, \quad \widehat{CE}^{up}(t,j) + c^*_{\alpha/2}\widehat{\text{var}}\{\widehat{CE}^{up}(t,j)\}^{1/2}]$$

where $c^*_{\alpha/2}$ can be computed using equation (4.3) of Vansteelandt *et al.* (2006), $\widehat{\text{var}}\{\widehat{CE}^{low}(t,j)\} = \widehat{\text{var}}\{\hat{F}_1^{NI,low}(t,j)\} + \widehat{\text{var}}\{\hat{F}_0(t,j)\}$ and $\widehat{\text{var}}\{\widehat{CE}^{up}(t,j)\} = \widehat{\text{var}}\{\hat{F}_1^{NI,up}(t,j)\} + \widehat{\text{var}}\{\hat{F}_0(t,j)\}$.

2.4 Simulation Study

Simulations were conducted to evaluate the performance of the methods described in Section 2.3 for drawing inference about $CE(t,j)$. Data were simulated based on the BAN study under five models: $\beta_j = -\infty, -1, 0, 1, \infty$ for fixed j. These five choices of β_j correspond to the two extreme selection models ($\beta_j = -\infty, \infty$), two intermediate selection models ($\beta_j = -1, 1$), and the no selection bias model ($\beta_j = 0$). The Gompertz distribution was used to simulate competing risks data (Jeong and Fine 2006). Under the Gompertz distribution the CIF can be expressed as $F(t,j) = 1 - \exp\left[\lambda_j\{1-\exp(\alpha_j t)\}/\alpha_j\right]$ where $\{\alpha_1, \ldots, \alpha_k, \lambda_1, \ldots, \lambda_k\}$ are chosen such that

$\sum_{j=1}^{k} \Pr[J = j] = \sum_{j=1}^{k} F(\infty, j) = 1$. For the simulation study $k = 3$ and the parameters $\{\alpha_1, \alpha_2, \alpha_3, \lambda_1, \lambda_2, \lambda_3\}$ were selected such that $F_1(28, 1) = 0.02$, $F_1(28, 2) = 0.02$, $F_1(28, 3) = 0.70$, and $\sum_{j=1}^{3} F_1(\infty, j) = 1$. These probabilities correspond roughly to the estimated risk of HIV infection ($j = 1$), death ($j = 2$) prior to HIV infection or weaning, and cessation of breastfeeding prior to HIV infection ($j = 3$) at 28 weeks in the BAN study among infants randomized to the infant ART arm who were HIV negative and alive at 2 weeks.

Simulations were conducted under two scenarios (for each of the five models). For the first scenario we let $\gamma = 0.9884$, corresponding to the estimated value of γ from the BAN study. In this scenario we considered estimating $CE(28, 1)$, i.e., the effect of treatment on risk HIV infection at 28 weeks. Note (2.6) and (2.8) hold in this scenario for $t = 28$ and $j = 1$ such that the estimators of the bounds are asymptotically normal. Because $\gamma = 0.9884$ is near the boundary value of 1, for the second scenario we let $\gamma = 0.75$. In order for (2.6) and (2.8) to hold in the second scenario, we considered estimating $CE(28, 3)$, i.e., the effect of treatment on weaning at 28 weeks. For the first scenario simulations were conducted under the alternative hypothesis $CE(28, 1) = -0.05$, i.e., the risk of HIV infection is lowered by 5% due to treatment. For the second scenario simulations were conducted where $CE(28, 3) = 0.05$, i.e., women are more likely to breastfeed at 28 weeks when the infant receives ART. For each model and each scenario, data sets of $n = 1520$ iid copies of W were simulated according to the following steps. The description below is for the first scenario where $j = 1$ is the event of interest; simulations were conducted analogously for the second scenario where $j = 3$ is the event of interest.

Step 1. $S_i(1)$ was drawn from a Bernoulli(0.0458), where 0.0458 was the estimated risk of infection or death at two weeks in the infant ART arm of BAN.

Step 2. If $S_i(1) = 1$, then by monotonicity $S_i(0) = 1$. In this case we let $T_i(0) =$

$J_i(0) = T_i(1) = J_i(1) = *$ because the survival time and failure type for individuals with $S_i = 1$ are not used by any of the estimators of $CE(t, j)$.

Step 3. If $S_i(1) = 0$, then $(T_i(1), J_i(1))$ were generated according to the Gompertz models described above. In particular, first $J_i(1)$ was generated from a multinomial distribution with cell probabilities $1 - \exp(\lambda_j/\alpha_j)$ for $j = 1, 2, 3$. Then $T_i(1)$ was set equal to $\tau_0 + U_i$ where U_i was randomly generated from the conditional distribution $\Pr[T_i(1) \leq t | J_i(1) = j] = F(t, j)/\Pr[J_i(1) = j]$ using the inverse probability transformation. Generating $T_i(1)$ in this fashion guarantees that $T_i(1) > \tau_0 = 2$ whenever $S_i(1) = 0$.

Step 4. If $S_i(1) = 0$, $S_i(0)$ was generated as follows. For $\beta_1 = -\infty$, $S_i(0) = I(T_i(1) < q_1^{(1-\gamma)}, J_i(1) = 1)$ where $q_j^{(1-\gamma)}$ is defined in general such that $\Pr[T_i(1) \leq q_j^{(1-\gamma)}, J_i(1) = j | S_i(1) = 0] = 1 - \gamma$. Note for the first scenario (2.8) holds for $t = 28$ and $j = 1$, guaranteeing the existence of $q_1^{(1-\gamma)}$. For, $\beta_1 = -1, 0, 1$, the value of $F_1^{prot}(28, 1; \beta_1)$ was found by solving (2.1) and (2.4) simultaneously, and then $S_i(0) \sim$ Bernoulli(p_{β_1}) where $p_{\beta_1} = (1 - \gamma)I(T_i(1) < 28, J_i(1) = 1)F_1^{prot}(28, 1; \beta_1)/F_1(28, 1) + (1 - \gamma)\{1 - I(T_i(1) < 28, J_i(1) = 1)\}\{1 - F_1^{prot}(28, 1; \beta_1)\}/\{1 - F_1(28, 1)\}$. For $\beta_1 = \infty$, $S_i(0) \sim$ Bernoulli(p_∞) where $p_\infty = (1-\gamma)\{1-I(T_i(1) < 28, J_i(1) = 1)\}/\{1-F_1(28, 1)\}$. Note for the first scenario (2.6) holds for $t = 28$ and $j = 1$, implying $1 - \gamma < 1 - F_1(28, 1)$ thus ensuring $p_\infty < 1$.

Step 5. If $S_i(0) = 0$, then we let $J_i(0) = J_i(1)$. If $S_i(0) = 0$ and $J_i(0) = 1$, then $T_i(0) = T_i(1)/\epsilon$, where ϵ was chosen such that $CE(28, 1) = -0.05$. If $S_i(0) = 0$ and $J_i(0) \neq 1$, then $T_i(0) = T_i(1)$. If $S_i(0) = 1$, then we set $T_i(0) = J_i(0) = *$.

Step 6. $C_i(0)$ and $C_i(1)$ were generated from exponential distributions with means 29 weeks and 18 weeks respectively.

Step 7. Z_i was randomly assigned such that $n_1 = 852$ and $n_0 = 668$.

Step 8. Given Z_i, we set $Y_i = \min\{T_i(Z_i), C_i(Z_i)\}$, $\Delta_i = I(Y_i = T_i(Z_i))$, $J_i = J_i(Z_i)$, and $S_i = S_i(Z_i)$.

These steps resulted in simulated data sets satisfying Assumptions 2.1 – 2.4 with $CE(28, 1) = -0.05$ for the first scenario. For each data set simulated, $\widehat{CE}(28, 1; \beta_j)$ was computed for $\beta_j = -\infty, -1, 0, 1, \infty$. Bootstrap percentile and Wald 95% confidence intervals as well as the uncertainty intervals described in Section 2.3.3 were also computed for each simulated data set, assumed value of β_1, and estimator of $CE(28, 1)$.

Table 2.1 reports the mean relative bias of $\widehat{CE}(28, j; \beta_j)$ based on 10,000 simulated data sets for both scenarios ($\gamma = 0.9884, j = 1$, and $\gamma = 0.75, j = 3$) and each model ($\beta_j = -\infty, -1, 0, 1, \infty$). The proposed estimator $\widehat{CE}(28, j; \beta_j)$ is approximately unbiased when β_j is correctly specified; for incorrectly specified β_j the relative bias can be quite large. For example, if β_1 is (incorrectly) assumed to be zero, corresponding to the naive analysis that simply compares infants HIV free and alive at two weeks from each study arm, when in fact $\beta_1 = -\infty$, then the relative bias of $\widehat{CE}(28, 1; \beta_1)$ is 23%. This demonstrates how a naive analysis that ignores the potential for selection bias can yield incorrect inference. This is demonstrated further in the scenario where $\gamma = 0.75$, in which case misspecifying β_3 leads to even greater relative bias.

Table 2.2 shows the empirical coverage probabilities of 95% pointwise bootstrap confidence intervals based on 500 bootstrap replications per simulated data set. When the correct β_j is specified, the confidence intervals associated with $\widehat{CE}(28, j; \beta_j)$ have approximately 95% coverage. Similar results were found using Wald confidence intervals (results not shown). Because β_j is not identifiable from the observable data, coverage of the uncertainty regions is perhaps of more practical interest. For the 50,000 simulated data sets from the first scenario (i.e., combining across the 10,000 data sets for each of the five values of β_j), the empirical coverage of the 95% pointwise uncertainty regions

was 97%. Similarly for the second scenario, the empirical coverage of the uncertainty intervals was 97%.

2.5 Application to BAN Study

The BAN study was a randomized clinical trial to assess interventions for the prevention of breast milk transmission of HIV in 2369 HIV infected mothers and their infants in Lilongwe, Malawi (van der Horst *et al.* 2009; Chasela *et al.* 2010). There were three arms in the BAN study: daily ART for the infant, daily ART for the mother, or control. While the primary analysis of the study considered comparisons of both ART arms to control, we will focus on comparing the infant ART and control arms only. In March 2008 the data and safety monitoring board stopped the control arm due to efficacy but recommended continued enrollment of mother/infant pairs into the two active treatment arms. This led to an imbalance in the final number of infants randomized to the three arms, with 852 infants in the infant ART arm and 668 infants in the control arm. In the infant ART arm there were 37 HIV infections and 2 deaths before $\tau_0 = 2$ weeks, while the control arm had 36 HIV infections and 2 infant deaths prior to 2 weeks. Thus $\hat{\gamma} = (630/668)/(813/852) = 0.9884$, as in the first scenario of the simulations in Section 2.4. Among infants HIV free and alive at 2 weeks, in the infant ART (control) arm 12 (32) became HIV infected, 588 (384) weaned prior to HIV infection, and 5 (6) died prior to HIV infection or weaning by 28 weeks. Figure 2.1 shows the Aalen-Johansen estimates of the cumulative risk of HIV, death prior to HIV infection or weaning, and cessation of breastfeeding prior to HIV infection for infants who were alive and uninfected at 2 weeks as in a standard analysis, i.e., assuming the no selection model $\beta_j = 0$ holds for all j. Figure 1(a) suggests a difference in the risk of HIV infection between the infant ART arm and the control arm, however direct comparison between the arms is subject to selection bias.

Figure 2.2 shows the semiparametric sensitivity analysis described in Section 2.3.2. The plot depicts $\widehat{CE}(28, 1; \beta_1)$ and pointwise 95% Wald confidence intervals for each value of β_1 (using bootstrap variance estimates). Note for the infant ART arm $\hat{F}_1(28, 1) = 0.0141$, suggesting (2.6) and (2.8) hold for $t = 28$ and $j = 1$. The estimated ignorance region for $CE(28, 1)$ equals [-0.056, -0.044] and the estimated 95% uncertainty interval equals [-0.078, -0.025]. This estimated uncertainty interval was computed using bootstrap variance estimates; using the analytical variance estimates (2.7) and (2.9) yielded a slightly wider uncertainty interval of [-0.084, -0.025]. In either case, because the uncertainty interval excludes 0, we conclude there is evidence of a causal effect of infant ART on the cumulative incidence of HIV at 28 weeks in the NI stratum. Moreover, without any assumptions about the selection bias mechanism, we are 95% confident daily infant ART lowers the risk of HIV infection at 28 weeks between 3% and 8%.

The veracity of these results relies on several key assumptions. While interference between infants was not likely, SUTVA could have been violated by changes in the infant ART regimen. Per protocol, if an infant on ART had an adverse event due to the study drug (nevirapine), the ART was changed (to lamivudine) and the infant remained in the study. Thus not all infants were on the same treatment for the duration of the study. Therefore, the effect of ART being estimated can be viewed as an average causal effect over all administered ARTs (Vanderweele 2011). While this interpretation answers the hypothesis proposed for the BAN study, it does not indicate which particular ART causes the greatest reduction in risk of HIV infection. Assumption 2.1 seems reasonable because treatment was randomized. While mothers were not blinded, they were counseled to breastfeed their infants regardless of randomization assignment and self-reported frequency of exclusive breastfeeding was comparable between study arms (Chasela *et al.* 2010). The BAN study principal investigator, Dr. Charles van der

Horst, indicated that monotonicity (Assumption 2.2) is reasonable (personal communication). Dr. van der Horst conjectured that an infant could have an adverse reaction to ART leading to increased susceptibility to HIV infection but he felt this was "highly unlikely." Monotonicity is also supported by the estimated risk of HIV infection or death at two weeks being lower in the infant ART arm than in the control arm.

Finally, note that two of the three endpoints in BAN were interval censored. In particular, the HIV infection times of the infants were interval censored, known only to be between the last negative and first positive HIV tests. Similarly, the actual timing of weaning is known only to be visits where the mother reported still breastfeeding and weaning. On the other hand, the time of death was known exactly for all infants. Other analyses of the BAN data have found that formally accounting for interval censoring almost always gives nearly the same result as using the midpoint or right endpoint of the interval. This is not surprising given the visits in the BAN study were fairly close together, typically two to four weeks apart. In settings where the intervals are wider, midpoint or right endpoint imputation may yield misleading results. Instead, a non-parametric estimator of $F_1(t, j)$ that allows for interval censored event times (Hudgens, Satten, and Longini 2001) can be employed in place of the Aalen-Johansen estimator. Inference that formally accounts for interval censoring is challenging however, owing to slow rates of convergence and non-standard limiting distributions of non-parametric estimators (for continuous time models; Groeneboom, Maathuis, and Wellner 2008a,b).

2.6 Discussion

The objective of many MTCT trials is to determine differences in the cumulative risk of breastfeeding transmission of HIV between study arms conditional on infants being HIV free and alive by some time point $\tau_0 > 0$. Here we have presented methods for evaluating the effect of treatment on the cumulative risk of HIV within a principal

stratum when death and weaning are competing risks. Large sample non-parametric bounds and a semi-parametric sensitivity analysis model were developed, and the methods were applied to the BAN study, a large, recent MTCT trial. A simulation study was presented demonstrating that the proposed methods perform well in finite samples similar to the BAN study. The simulations also illustrated how analyses that ignore the potential for selection bias by simply conditioning on being HIV free and alive at τ_0 can give misleading results in settings similar to the BAN study.

The analysis of the BAN study indicates infant ART reduces the risk of HIV infection by 28 weeks in infants who would be HIV free and alive at two weeks regardless of treatment assignment. The proposed methods could be applied in other settings as well. For example, BAN investigators (personal communication) were interested in comparing the risk of HIV infection or death by 48 weeks conditional on infants being HIV free and alive at 28 weeks; here $\tau_0 = 28$ weeks is further from time 0 and the potential for selection bias is even greater than the analysis presented in Section 2.5. Another example is given by the Zambia Exclusive Breastfeeding (ZEB) study, a randomized MTCT study conducted to evaluate whether abrupt weaning at four months compared with continued breastfeeding increases survival of children of HIV-infected mothers (Kuhn *et al.* 2008). Randomization occurred at one month postpartum in the ZEB study, however Kuhn *et al.* (2008) presented a comparison of the randomized groups conditional on infants being HIV free and breastfeeding at four months.

A key assumption of the methods described in this paper is monotonicity, which implies that the treatment is no worse than control for any individual in terms of the intermediate variable S. This assumption seems reasonable in the analysis of the BAN study presented in Section 5, but in other settings it may be unrealistic. For example, monotonicity might be considered dubious in an analysis comparing the two active arms of the BAN trial, i.e., maternal ART versus infant ART. In such settings methods that

relax or do not require this assumption would be needed. Following Zhang and Rubin (2003), nonparametric bounds analogous to those in Section 2 can be derived without assuming monotonicity. Specifically, note that $F_0(t,j) = \phi F_0^{NI}(t,j) + (1-\phi)F_0^{harm}(t,j)$, where $\phi = \Pr[S_i(1) = 0 | S_i(0) = 0]$ and $F_0^{harm}(t,j) = \Pr[T_i(0) \leq t, J_i(0) = j | S_i(0) = 0, S_i(1) = 1]$. If γ and ϕ were identifiable, then bounds for $F_0^{NI}(t,j)$ can be constructed analogous to (2.2) and (2.3) and combined with bounds for $F_1^{NI}(t,j)$ to obtain the following bounds on $CE(t,j)$:

$$CE^{low}(t,j) = \max\left\{\frac{F_1(t,j) - (1-\gamma)}{\gamma}, 0\right\} - \min\left\{\frac{F_0(t,j)}{\phi}, 1\right\} \tag{2.10}$$

and

$$CE^{up}(t,j) = \min\left\{\frac{F_1(t,j)}{\gamma}, 1\right\} - \max\left\{\frac{F_0(t,j) - (1-\phi)}{\phi}, 0\right\}. \tag{2.11}$$

However, without the monotonicity assumption γ and ϕ are not identifiable. Let $\pi = \Pr[S_i(0) = 0, S_i(1) = 1]$ and note that

$$\gamma = \Pr[S_i(0) = 0 | S_i(1) = 0] = \frac{\Pr[S_i(0) = 0, S_i(1) = 0]}{\Pr[S_i(1) = 0]} = \frac{\Pr[S_i(0) = 0] - \pi}{\Pr[S_i(1) = 0]}$$

and

$$\phi = \Pr[S_i(1) = 0 | S_i(0) = 0] = \frac{\Pr[S_i(0) = 0, S_i(1) = 0]}{\Pr[S_i(0) = 0]} = \frac{\Pr[S_i(0) = 0] - \pi}{\Pr[S_i(0) = 0]}$$

are identifiable from the observed data for a fixed value of π. Thus, the lower bound of $CE(t,j)$ is found by minimizing (2.10) over π where $\max\{0, \Pr[S_i(0) = 0] - \Pr[S_i(1) = 0]\} \leq \pi \leq \min\{\Pr[S_i(0) = 0], \Pr[S_i(1) = 1]\}$. Likewise, the upper bound of $CE(t,j)$ is found by maximizing (2.11) over the same range of π. Sensitivity analysis could be performed by adapting the methods of Shepherd, Gilbert, and Dupont (2011). For instance, similar to Assumption 2.4, a selection model for $F_0^{NI}(t,j)$ could be assumed,

such as:

Assumption 2.5:

$$\exp(\eta_j) = \frac{F_0^{NI}(t,j)/\{1 - F_0^{NI}(t,j)\}}{F_0^{harm}(t,j)/\{1 - F_0^{harm}(t,j)\}}. \tag{2.12}$$

Sensitivity analysis under Assumptions 2.1, 2.3, 2.4, and 2.5 would be performed by varying π over $\max\{0, \Pr[S_i(0) = 0] - \Pr[S_i(1) = 0]\} \leq \pi \leq \min\{\Pr[S_i(0) = 0], \Pr[S_i(1) = 1]\}$ and η_j, β_j each over $(-\infty, \infty)$. The resulting inference will be more precise if the ranges of π, η_j, and β_j can be further restricted based on prior information elicited from subject matter experts.

For the MTCT research motivating this work, interest focused on the principal stratum of infants HIV free and alive at τ_0 under either treatment assignment. The methods developed could also be applied to infants HIV infected and alive at τ_0 under either treatment where T might denote the time until death from various causes. Beyond MTCT trials, the methods developed could be applied in other settings where inference about treatment effects within principal strata is of interest (e.g., truncation-by-death or non-compliance) and the endpoint is a time-to-event outcome subject to competing risks. Further research might entail allowing the cumulative incidence functions to depend on baseline covariates (e.g., as in Jeong and Fine (2007)).

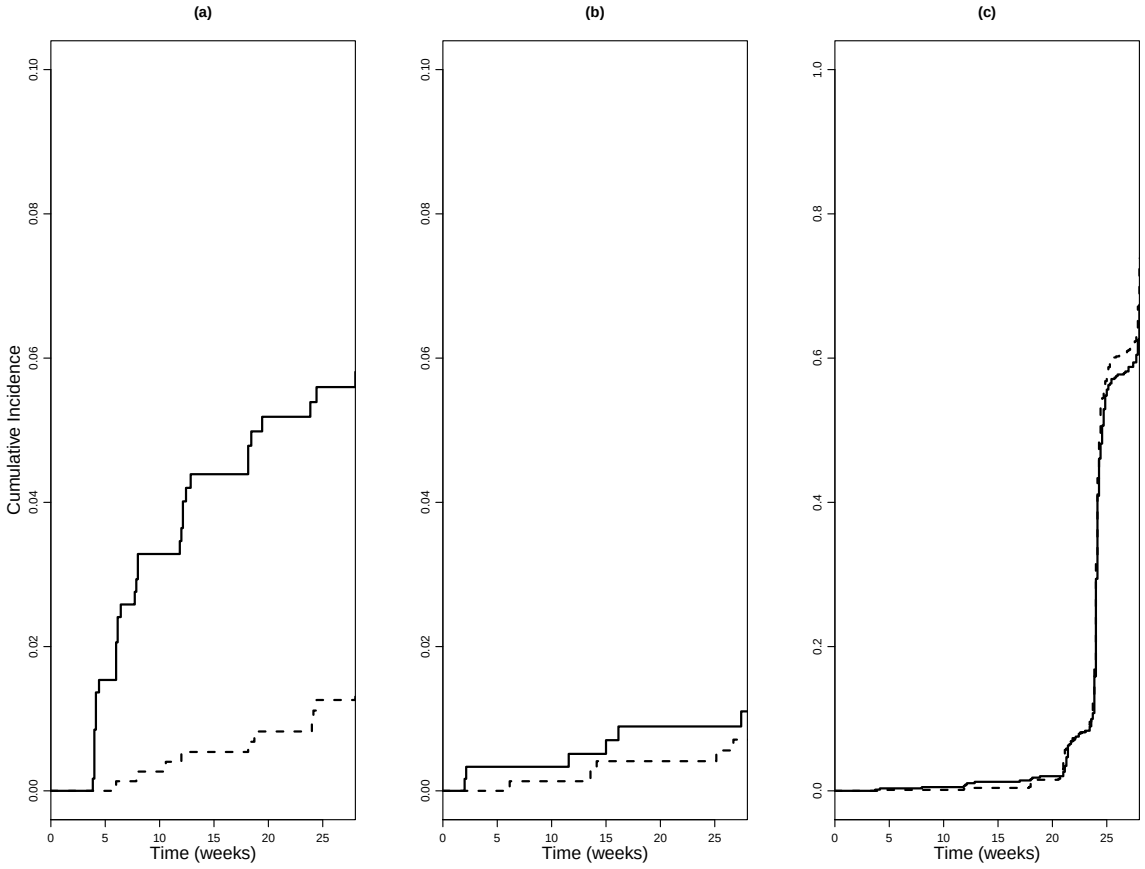

Figure 2.1: Estimated cumulative incidence functions, $\hat{F}_z(23, j)$, for the three events from the BAN study: (a) HIV infection, (b) HIV-free death prior to weaning, and (c) cessation of breastfeeding prior to HIV infection. For each panel, $Z_i = 0$ (control) is represented by the solid line (—) and $Z_i = 1$ (infant ART) is represented by the dashed line (− − −).

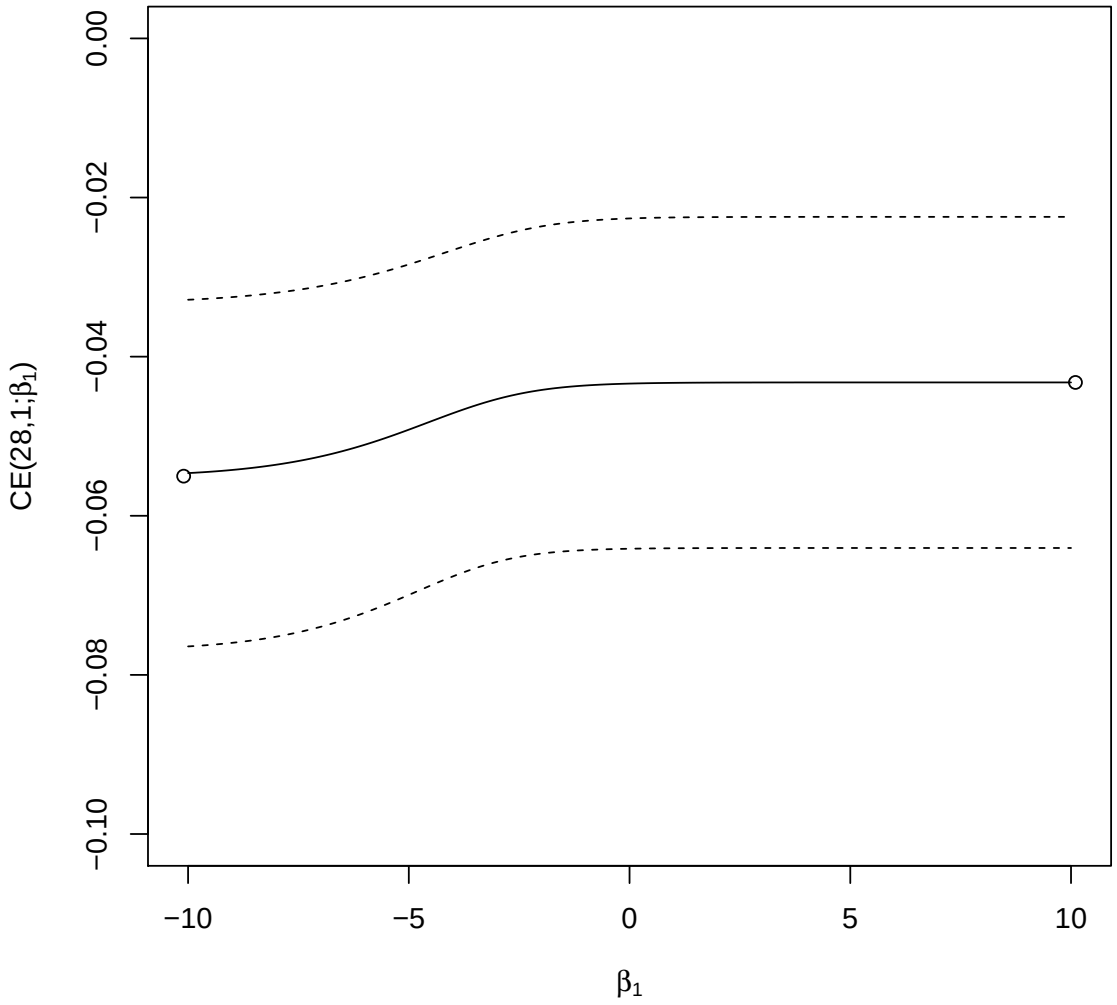

Figure 2.2: Sensitivity analysis of the effect of infant ART on the cumulative incidence of HIV at 28 weeks for the BAN study. The solid line — denotes $\widehat{CE}(28, 1; \beta_1)$ and the dotted lines $\cdots$ denote pointwise 95% confidence intervals. The estimated non-parametric bounds corresponding to $\beta_1 = -\infty$ and $\beta_1 = \infty$ are given by $\circ$.

Table 2.1: Empirical relative bias of estimates of $CE(28, j)$ from simulation study described in Section 2.4 for both scenarios. Bold entries correspond to estimates where the assumed β_j was correct. Relative bias of $\widehat{CE}(28, j; \beta_j)$ defined as $\{\widehat{CE}(28, j; \beta_j) - CE(28, j)\}/CE(28, j)$.

True parameters			Assumed β_j				
γ	CE	β_j	$-\infty$	-1	0	1	∞
0.9884	-0.05	$-\infty$	**0.02**	0.22	0.23	0.24	0.24
		-1	-0.19	**0.01**	0.01	0.02	0.02
		0	-0.21	-0.01	**0.00**	0.00	0.01
		1	-0.20	-0.01	0.00	**0.00**	0.00
		∞	-0.21	-0.01	-0.00	0.00	**0.00**
0.75	0.05	$-\infty$	**-0.01**	-1.10	-2.01	-3.13	-6.69
		-1	1.02	**-0.01**	-0.88	-1.96	-5.32
		0	2.00	0.91	**-0.01**	-1.12	-4.67
		1	2.66	1.77	0.98	**0.00**	-2.93
		∞	6.50	5.44	4.55	3.45	**0.00**

Table 2.2: Empirical coverage of pointwise 95% bootstrap percentile confidence intervals of $CE(28, j)$ from simulation study described in Section 2.4 for both scenarios. Bold entries correspond to estimates where the assumed β_j was correct.

True parameters			Assumed β_j				
γ	CE	β_j	$-\infty$	-1	0	1	∞
0.9884	-0.05	$-\infty$	**0.95**	0.81	0.80	0.80	0.79
		-1	0.90	**0.95**	0.94	0.94	0.94
		0	0.88	0.94	**0.94**	0.94	0.94
		1	0.89	0.95	0.95	**0.95**	0.95
		∞	0.89	0.95	0.95	0.95	**0.95**
0.75	0.05	$-\infty$	**0.95**	0.55	0.06	0.00	0.00
		-1	0.66	**0.94**	0.65	0.06	0.00
		0	0.14	0.66	**0.94**	0.47	0.00
		1	0.01	0.12	0.55	**0.94**	0.01
		∞	0.00	0.00	0.00	0.00	**0.94**

Chapter 3

Sharpening Bounds on Principal Effects with Covariates

3.1 Introduction

Often in randomized trials to evaluate the effect of a treatment, inference is hampered by possible selection bias induced by conditioning on or adjusting for a variable measured post-randomization. One approach that avoids potential selection bias is to focus inference on the causal effect within a principal strata of interest, i.e., the principal effect (Frangakis and Rubin 2002). Principal strata are defined by the pair of potential outcomes under either treatment assignment of the post-randomization variable. In vaccine trials, a principal stratum of interest may be individuals who would be infected at a certain time regardless of vaccine status (Shepherd *et al.* 2011). In studies of interventions to prevent mother-to-child transmission (MTCT) of HIV through breastfeeding, a principal stratum of interest is infants who would be uninfected at a certain time regardless of treatment (Nolen and Hudgens 2011). In either case, principal strata membership is unidentifiable from the observable data without strong assumptions because only one of the two post-randomization variable potential outcomes is ever observed for an individual. In turn, the principal effect of interest is

not identifiable. One approach to cope with lack of identifiability is to conduct sensitivity analysis wherein some model is assumed indexed by an unidentifiable parameter conditional on which the principal effect is identifiable. Inference about the principal effect is conducted conditional on some value of the unidentifiable parameter and then sensitivity of the inference is examined by considering different values of the parameter. An alternative approach entails drawing inference about bounds on the principal effects, e.g., Zhang and Rubin (2003). Informally, these extreme bounds provide the smallest and largest possible values of the principal effect consistent with the observed data. This approach is appealing in that typically bounds can be obtained under minimal assumptions. However, in many cases the bounds may be quite wide and therefore not particularly informative about the principal effect.

Grilli and Mealli (2008) derived nonparametric bounds on the principal effect under a number of different assumptions. They suggested these bounds can be improved (or narrowed) by creating bounds within strata defined by a baseline covariate and combining these stratum specific bounds by taking a weighted average to obtain new *adjusted* bounds on the principal effect. Grilli and Mealli (2008) performed this method on data from an employment study with mixed results. The adjusted bounds were an improvement on only one side of the unadjusted bounds, i.e., the adjusted lower bound was larger than the unadjusted lower bound but the adjusted upper bound was also larger than the unadjusted upper bound. The reason for only partial improvement was not addressed and characterization of which circumstances will lead to improved bounds was not investigated.

In this work we consider sharpening (or narrowing) the large sample bounds of a principal effect using information from a baseline categorical covariate, as proposed by Grilli and Mealli (2008). In Section 3.2, notation and assumptions are introduced. Section 3.3 addresses non-identifiability of the principal effect and in Section 3.4 the

unadjusted bounds are reviewed. In Section 3.5, adjusted bounds are presented by taking a weighted average of bounds within levels of the baseline covariate. Section 3.6 contains the main result, giving necessary and sufficient conditions under which the covariate adjusted bounds improve upon (i.e., are narrower than) the unadjusted bounds. Cases in which adjusting for the covariate will identify the principal effect are illustrated in Section 3.8. In Section 3.7, the adjusted and unadjusted bounds are considered using data from a recent, large MTCT study. A brief discussion is presented in Section 3.9. Proofs of the propositions in Section 3.6 are given in the Appendix.

3.2 Notation and Assumptions

To motivate, throughout we consider the MTCT example where infants of HIV positive mothers are randomized at birth to treatment or control. Suppose n infants are enrolled in a MTCT study and randomly assigned one of two treatments, 0 or 1, at baseline (birth or time 0). For $i = 1, \ldots, n$, let $Z_i = 0$ if infant i is assigned treatment 0 and $Z_i = 1$ otherwise. Without loss of generality, assume $Z_i = 0$ corresponds to control, and $Z_i = 1$ corresponds to active treatment. Let X_i be some binary variable measured at baseline (prior to randomization) taking on values 0 or 1. Let S_i denote whether infant i is infected at a pre-specified post-randomization time point $\tau_0 > 0$, i.e., $S_i = 1$ if the infant is infected at τ_0, $S_i = 0$ otherwise. Let Y_i be a binary outcome measured only in infants with $S_i = 0$. Let $S_i(z)$ denoted the potential value of S_i when assigned treatment z for $z = 0, 1$ such that $S_i = (1 - Z_i)S_i(0) + Z_i S_i(1)$. Define $Y_i(z)$ similarly. We assume the i^{th} infant's treatment assignment does not affect the potential outcomes of other infants (i.e., no interference) and that there are not multiple forms of treatment, i.e., the stable unit treatment value assumption (SUTVA) holds (Rubin 1980).

Principal strata are defined by sets of infants with the same potential outcome

pair $(S_i(0) = s_0, S_i(1) = s_1)$. Define the always infected (AI) principal stratum to be infants with $s_0 = s_1 = 1$, i.e., infants who would be infected at τ_0 regardless of treatment assignment. Similarly define the harmed stratum as those infants with $s_0 = 0, s_1 = 1$; the protected stratum as those infants with $s_0 = 1, s_1 = 0$; and the never infected (NI) stratum as those infants with $s_0 = s_1 = 0$.

Here and throughout assume *Assumption 1:* $Z_i \perp (X_i, S_i(0), S_i(1), Y_i(0), Y_i(1))$ *Assumption 2 (Monotonicity):* $S_i(1) \leq S_i(0)$ for all i Assumption 1 will hold in randomized trials. Monotonicity assumes that treatment does no harm, i.e., there are no infants who would be infected only if treated. Under Assumption 2, there are only three possible principal strata: AI, NI, and protected. In MTCT studies to prevent breastmilk transmission of HIV, investigators are interested in the NI stratum because infections prior to τ_0 could be due to the birthing process and not breastfeeding. The causal estimand of interest, the principal effect, is the effect of treatment on Y_i in infants who would be uninfected and alive at τ_0 under either treatment assignment, namely

$$CE = \Pr[Y_i(1) = 1 | S_i(0) = S_i(1) = 0] - \Pr[Y_i(0) = 1 | S_i(0) = S_i(1) = 0].$$

Below we consider large sample bounds for CE that do and do not adjust for baseline covariates.

3.3 Partial Identifiability

In this section we consider the identifiability of CE. Let $\theta_{zst} = \Pr[Y_i(z) = 1 | S_i(1) = s, S_i(0) = t]$, $\pi_z = \Pr[Y_i(z) = 1 | S_i(z) = 0]$, and $\gamma = \Pr[S_i(0) = 0 | S_i(1) = 0]$, such that $CE = \theta_{100} - \theta_{000}$. Assume $\gamma > 0$ as otherwise the NI stratum is empty with probability 1. Under Assumptions 1 and 2, $\theta_{000} = \Pr[Y_i = 1 | S_i = 0, Z_i = 0]$, which is identifiable from the observed data. However, θ_{100} is not identifiable. Following Hudgens and

Halloran (2006) note

$$\begin{aligned}\Pr[Y_i(1)=1|S_i(1)=0] &= \Pr[Y_i(1)=1|S_i(1)=S_i(0)=0]\Pr[S_i(0)=0|S_i(1)=0] \\ &+ \Pr[Y_i(1)=1|S_i(1)=0,S_i(0)=1]\Pr[S_i(0)=1|S_i(1)=0],\end{aligned}$$

i.e.,

$$\pi_1 = \gamma\theta_{100} + (1-\gamma)\theta_{101}. \tag{3.1}$$

Under Assumption 1 π_1 is identifiable. Under Assumptions 1 and 2 γ is identifiable because

$$\gamma = \frac{\Pr[S_i(0)=0]}{\Pr[S_i(1)=0]} = \frac{\Pr[S_i=0|Z=0]}{\Pr[S_i=0|Z=1]}.$$

On the other hand, θ_{100} and θ_{101} are not identifiable because infants who were treated and uninfected at τ_0 are a mixture of infants from the protected and NI strata. Solving (3.1) for θ_{100} yields,

$$\theta_{100} = \frac{\pi_1}{\gamma} - \frac{1-\gamma}{\gamma}\theta_{101}. \tag{3.2}$$

Equation (3.2) describes a line with intercept π_1/γ and slope $-(1-\gamma)/\gamma$. Any point on this line will give rise to the same observed data distribution. Two populations will have the same observable data if all else being equal $\pi_1 * \gamma + \Pr[Y_i(1)=1|S_i(0)=1, S_i(1)=0] * (1-\gamma)$ is the same in the two populations.

Note that if $\gamma = 1$, $\pi_1 = 1$, or $\pi_0 = 0$ then CE is identifiable. If $\gamma = 1$, then (3.2) is a horizontal line with intercept π_1 and thus $\theta_{100} = \pi_1$. If $\pi_1 = 1$, then (3.1) implies $\theta_{100} = 1$. Likewise, if $\pi_1 = 0$, then (3.1) implies $\theta_{100} = 0$. Otherwise, if $\gamma < 1$ and $0 < \pi_1 < 1$, under Assumptions 1 and 2, CE is not identifiable from the observable random variables.

3.4 Unadjusted Bounds

In this section, we present large sample bounds that ignore the baseline covariate X. Large sample bounds for CE are found by first bounding θ_{100}. The upper bound for θ_{100} is obtained by assuming $\theta_{101} = 0$ or $\theta_{100} = 1$. Likewise, the lower bound for θ_{100} is obtained by assuming $\theta_{101} = 1$ or $\theta_{100} = 0$. These bounds can be envisaged as corresponding to the point where the line (3.2) intersects the unit square (Hudgens and Halloran 2006).

In particular, the upper and lower bounds are

$$\theta^u_{100} = \min\left\{\frac{\pi_1}{\gamma}, 1\right\} \quad \text{and} \quad \theta^l_{100} = \max\left\{\frac{\pi_1 - (1-\gamma)}{\gamma}, 0\right\}. \tag{3.3}$$

Bounds for CE are found by replacing θ_{100} by θ^u_{100} and θ^l_{100}, i.e., $CE^u = \theta^u_{100} - \theta_{000}$ and $CE^l = \theta^l_{100} - \theta_{000}$. These bounds will be referred to as "unadjusted" bounds since no information from the covariate is used.

To illustrate, let the probabilities corresponding to a fictitious trial of MTCT of HIV be $\gamma = 0.95$, $\pi_1 = 0.02$, and $\pi_0 = 0.05$. Using (3.3) for this trial, $\theta^u_{100} = \min\{0.02/0.95,1\} = 0.021$ and $\theta^l_{100} = \max\{(0.02 - (1 - 0.95))/0.95, 0\} = 0$. This gives the unadjusted bounds as $[CE^l, CE^u] = [0 - 0.05, 0.021 - 0.05] = [-0.05, -0.029]$ since $\theta_{000} = \pi_0$. Let the probabilities for a second fictitious trial be $\gamma = 0.80$, $\pi_1 = 0.85$, and $\pi_0 = 0.95$. Thus, $\theta^u_{100} = 1$ and $\theta^l_{100} = 0.813$, implying the unadjusted bounds are $[CE^l, CE^u] = [-0.137, 0.05]$. These two fictitious trials will be revisited in the next section.

3.5 Adjusted Bounds

Here we consider the method proposed by Grilli and Mealli (2008) for adjusting the large sample bounds using the binary baseline covariate X, i.e., bounds will be

obtained within strata defined define by X and then weighted averages of the stratum specific bounds will be computed. Let $\theta_{zstx} = \Pr[Y_i(z) = 1|S_i(1) = s, S_i(0) = t, X_i = x]$, $\gamma_x = \Pr[S_i(0) = 0|S_i(1) = 0, X_i = x]$, $\pi_{zx} = \Pr[Y_i(z) = 1|S_i(1) = 0, X_i = x]$, $\phi_x = \Pr[X_i = x|S_i(1) = S_i(0) = 0]$, and $\lambda_x = \Pr[X_i = x|S_i(1) = 0]$ for $x = 0, 1$. Note

$$\theta_{100} = \sum_x \theta_{100x}\phi_x, \tag{3.4}$$

where here and in the sequel $\sum_x = \sum_{x=0}^{1}$. As in the unadjusted case, θ_{100x} is not identifiable but using arguments analogous to (3.2) for $X_i = x$ we have

$$\theta_{100x} = \frac{\pi_{1x}}{\gamma_x} - \frac{1-\gamma_x}{\gamma_x}\theta_{101x}, \tag{3.5}$$

and identifiable upper and lower bounds for θ_{100} are

$$\theta^u_{100x} = \min\left\{\frac{\pi_{1x}}{\gamma_x}, 1\right\} \quad \text{and} \quad \theta^l_{100x} = \max\left\{\frac{\pi_{1x} - (1-\gamma_x)}{\gamma_x}, 0\right\}. \tag{3.6}$$

Under Assumptions 1 and 2, ϕ_x is identifiable because $\Pr[X_i = x|Z_i = 0, S_i = 0] = \Pr[X_i = x|S_i(0) = 0] = \Pr[X_i = x|S_i(1) = S_i(0) = 0]$. Therefore, identifiable bounds for θ_{100} can be obtained by combining (3.4) and (3.6), namely

$$\theta^u_{100X} = \sum_x \theta^u_{100x}\phi_x \quad \text{and} \quad \theta^l_{100X} = \sum_x \theta^l_{100x}\phi_x. \tag{3.7}$$

This leads to adjusted bounds $CE^u_X = \theta^u_{100X} - \theta_{000}$ and $CE^l_X = \theta^l_{100X} - \theta_{000}$.

Table 3.1 contains the values of two baseline covariates, X_1 and X_2, for each of the fictional trials discussed in Section 3.4. For the first trial and X_1, by (3.6), we have $\theta^u_{1000} = \min\{0.035/0.995, 1\} = 0.035$, $\theta^l_{1000} = \max\{(0.035-(1-0.995))/0.995, 0\} = 0.03$, $\theta^u_{1001} = 0.011$, and $\theta^l_{1001} = 0$. Thus, $\theta^u_{100X} = 0.035 * 0.419 + 0.011 * 0.581 = 0.021$ and

$\theta^l_{100X} = 0.013$. Therefore, there is improvement to the lower bound on the causal effect when adjusting for X_1, from $[CE^l, CE^u] = [-0.05, -0.029]$ to $[CE^l_{X_1}, CE^u_{X_1}] = [-0.037, -0.029]$. However, using X_2 in the first trial there is no improvement since $\theta^u_{100X} = 0.021 = \theta^u_{100}$ and $\theta^l_{100X} = 0 = \theta^l_{100}$.

For the second trial and X_1, $\theta^u_{1000} = 0.854$, $\theta^l_{1000} = 0.73$, $\theta^u_{1001} = 1$, and $\theta^l_{1001} = 0.878$. Thus, $\theta^u_{100X} = 0.935$ and $\theta^l_{100X} = 0.813$. Here adjusting for X_1 yields a smaller upper bound resulting in narrower bounds, i.e., $[CE^l, CE^u] = [-0.137, 0.05]$ to $[CE^l_{X_1}, CE^u_{X_1}] = [-0.137, -0.015]$. In fact, the adjusted upper bound is less than the null value of 0, indicating treatment has an effect in the NI principal stratum. On the other hand, using X_2 there was again no improvement since $\theta^u_{100X} = 1 = \theta^u_{100}$ and $\theta^l_{100X} = 0.813 = \theta^l_{100}$.

A graphical depiction of the unadjusted and adjusted bounds is given in Figure 3.1. The unadjusted bounds are found where the solid lines intersect the unit square. Bounds within strata defined by X correspond to where the dashed and dotted lines intersect the unit square. The adjusted bounds, represented by $\circ$ and $+$, are weighted averages of these stratum specific bounds. For example, in the upper left panel corresponding to trial 1 and X_1, we see that θ^l_{100X} is larger than θ^l_{100} since the $+$ is above 0, the point where the solid line intersects the horizontal axis.

3.6 Improvement of the Bounds

The examples in the preceding section illustrate that adjusting for a baseline covariate may or may not improve the bounds on CE. In this section, we give necessary and sufficient conditions for when the adjusted bounds (3.7) will be narrower than the unadjusted bounds of (3.3). Proofs for all propositions are given in the Appendix.

Proposition 1. $[\theta^l_{100X}, \theta^u_{100X}] \subseteq [\theta^l_{100}, \theta^u_{100}]$ *for any baseline binary covariate* X.

According to Proposition 1, the adjusted bounds will be at least as narrow as the unadjusted bounds no matter the choice of X. To characterize the conditions under which the adjusted bounds are strictly narrower than the unadjusted bounds, assume X_i takes on values x and x' and consider the following two criteria:

$$\pi_{1x} < \gamma_x \text{ and } \pi_{1x'} > \gamma_{x'}, \tag{3.8}$$

and

$$\pi_{1x} > (1 - \gamma_x) \text{ and } \pi_{1x'} < (1 - \gamma_{x'}), \tag{3.9}$$

where the value of x in (3.8) and (3.9) is not necessarily the same. In words, (3.8) and (3.9) indicate that X_i is informative about relation of the distribution of $S_i(0)$ given $S_i(1) = 0$ and the distribution of $Y_i(1)$ given $S_i(1) = 0$. On the other hand, if X_i is uninformative about this relation then neither (3.8) nor (3.9) will hold. For example, if X_i is independent of $S_i(0)$ given $S_i(1) = 0$ and if X_i is independent of $Y_i(1)$ given $S_i(1) = 0$, then neither (3.8) nor (3.9) will hold. Using (3.8) and (3.9), the following propositions characterize exactly the situations when the adjusted bounds will be narrower.

Proposition 2. *$\theta^u_{100X} < \theta^u_{100}$ if and only if X satisfies (3.8).*

Proposition 2 states (3.8) is a necessary and sufficient condition for the adjusted upper bound for θ_{100} to be smaller than the unadjusted bound. This proposition is exemplified in the second fictional trial from Section 3.5 using X_1, where $\pi_{10} < \gamma_0$ and $\pi_{11} > \gamma_1$.

Proposition 3. *$\theta^l_{100X} > \theta^l_{100}$ if and only if X satisfies (3.9).*

Proposition 3 provides the necessary and sufficient condition for the adjusted lower bound to be θ_{100} larger than the unadjusted bound. This proposition is illustrated in

the first fictional trial from Section 3.5 using X_1, where $\pi_{10} > (1-\gamma_0)$ and $\pi_{11} < (1-\gamma_1)$.

It follows immediately from Propositions 2 and 3 that if (3.8) and (3.9) both hold then the adjusted bounds are strictly contained within the unadjusted bounds and if neither hold, the adjusted and unadjusted bounds are equal.

3.7 Illustration

Grilli and Mealli (2008) analyzed data on academic careers and job opportunities of university students and found that the estimated adjusted bounds were not strictly contained within the estimated unadjusted bounds, apparently contradicting Proposition 1. In this section we consider a MTCT trial where a similar relationship between the estimated adjusted and unadjusted bounds is found.

The Breastfeeding, Antiretoviral, and Nutrition (BAN) study was a randomized clinical trial to assess interventions for the prevention of breast milk transmission of HIV in 2369 HIV infected mothers and their infants in Lilongwe, Malawi (Chasela *et al.* 2010). There were three arms in the BAN study: daily antiretroviral therapy (ART) for the infant, daily ART for the mother, or control. While the primary analysis of the study considered comparisons of both ART arms to control, we will focus on comparing the infant ART and control arms only. Per protocol, infants who died or were infected in the first 2 weeks post-treatment were excluded from the primary analysis, creating the potential for selection bias. Let $S_i = 1$ if the infant became HIV positive or died by 2 weeks, $S_i = 0$ otherwise. Furthermore, let X_i be an indicator of low birth weight (< 2.5 kg), i.e., $X_i = 1$ if the infant had low birth weight, 0 otherwise. Define Y_i as HIV infection status at 28 weeks where $Y_i = 1$ if an infant is infected by 28 weeks. As there is right censoring in the BAN study, a more formal analysis would correct for censoring using a survival outcome but as this is for illustration we will assume Y_i is observed for all infants. The principal effect of interest is the difference in risk of HIV

infection at 28 weeks between the infant ART and control arms for infants in the NI stratum.

For estimation, the consistent moment-based estimators under Assumptions 1 and 2 will be used. Let

$$\hat{\gamma} = \min\left\{\frac{\sum(1-S_i)(1-Z_i)/\sum(1-Z_i)}{\sum(1-S_i)Z_i/\sum Z_i}, 1\right\}$$

and

$$\hat{\gamma}_x = \min\left\{\frac{\sum(1-S_i)(1-Z_i)I(X_i=x)/\sum(1-Z_i)I(X_i=x)}{\sum(1-S_i)Z_iI(X_i=x)/\sum Z_iI(X_i=x)}, 1\right\},$$

where $\sum = \sum_{i=1}^{n}$ and $I()$ is the usual indicator function. Likewise, let $\hat{\pi}_z = \sum Y_i(1-S_i)I(Z_i=z)/\sum(1-S_i)I(Z_i=z)$, $\hat{\pi}_{1x} = \sum Y_i(1-S_i)Z_iI(X_i=x)/\sum(1-S_i)Z_iI(X_i=x)$, $\hat{\lambda}_x = \sum(1-S_i)Z_iI(X_i=x)/\sum(1-S_i)Z_i$ and $\hat{\phi}_x = \sum(1-S_i)(1-Z_i)I(X_i=x)/\sum(1-S_i)(1-Z_i)$. For $\hat{\pi}_z$, the Aalen-J The estimates $\hat{\theta}^u_{100}$, $\hat{\theta}^l_{100}$, $\hat{\theta}^u_{100x}$, and $\hat{\theta}^l_{100x}$ are calculated using the above estimators in (3.3) and (3.6) and finally $\hat{\theta}^u_{100X} = \sum_x \hat{\theta}^u_{100x}\hat{\phi}_x$ and $\hat{\theta}^l_{100X} = \sum_x \hat{\theta}^l_{100x}\hat{\phi}_x$.

Table 3.2 presents data from this study by treatment and X_i. Using these data and the above estimators, we have $\hat{\gamma} = (630/668)/(813/852) = 0.9884$, $\hat{\pi}_1 = 0.0148$, and $\hat{\pi}_0 = 0.0507$. Thus the estimated unadjusted upper and lower bounds are $\hat{\theta}^u_{100} = 0.0149$, $\hat{\theta}^l_{100} = 0.0032$ and $[\widehat{CE}^l, \widehat{CE}^u] = [-0.0476, -0.0359]$.

Using the data stratified by X_i, $\hat{\gamma}_0 = 1$ and $\hat{\gamma}_1 = 0.8612$. Furthermore, $\hat{\pi}_{10} = 0.0107$, $\hat{\pi}_{11} = 0.0645$, $\hat{\theta}^u_{1000} = 0.0107$, $\hat{\theta}^l_{1000} = 0.0107$, $\hat{\theta}^u_{1001} = 0.0749$, $\hat{\theta}^l_{1001} = 0$, $\hat{\lambda}_0 = 0.9237$, $\hat{\lambda}_1 = 0.0763$, $\hat{\phi}_0 = 0.9270$, and $\hat{\phi}_1 = 0.0730$. Therefore, $\hat{\theta}^u_{100X} = 0.0153$ and $\hat{\theta}^l_{100X} = 0.0099$ and $[\widehat{CE}^l_X, \widehat{CE}^u_X] = [-0.0409, -0.0354]$. Since X_i satisfies (3.9) empirically, i.e., $\hat{\pi}_{10} > (1-\hat{\gamma}_0)$ and $\hat{\pi}_{11} < (1-\hat{\gamma}_1)$, the estimated adjusted lower bound is larger than the estimated unadjusted lower bound, consistent with Proposition 3. However, the estimated adjusted upper bound is larger than the estimated unadjusted bound which

seems to contradict Proposition 1.

In this example, $\hat{\gamma} = 0.9884 \neq 0.9894 = \sum_x \hat{\gamma}_x \hat{\lambda}_x$ even though $\gamma = \sum_x \gamma_x \lambda_x$. This suggests an alternative estimator for λ_x, namely $\tilde{\lambda}_x = \hat{\lambda}_x \hat{\alpha}_x$ where

$$\hat{\alpha}_x = \frac{\sum(1 - Z_i) I(X_i = x) / \sum(1 - Z_i)}{\sum Z_i I(X_i = x) / \sum Z_i}.$$

Note $\hat{\alpha}_x$ is a consistent estimator of $\Pr[X_i = x | Z_i = 0] / \Pr[X_i = x | Z_i = 1]$, which equals 1 by Assumption 1, implying $\hat{\alpha}_x \xrightarrow{p} 1$ and thus $\tilde{\lambda}_x \xrightarrow{p} \lambda_x$. Similarly, let $\tilde{\phi}_x = \hat{\phi}_x / \hat{\alpha}_x$. By design, the estimators $\tilde{\theta}^u_{100X} = \sum_x \hat{\theta}^u_{100x} \tilde{\phi}_x$ and $\tilde{\theta}^l_{100X} = \sum_x \hat{\theta}^l_{100x} \tilde{\phi}_x$ will satisfy the relationship given in Proposition 1, i.e., the adjusted bounds computed using $\tilde{\phi}_x$ and $\tilde{\phi}_x$ will always be at least as narrow as the unadjusted bounds.

Using the BAN data, $\hat{\alpha}_0 = 0.9912$, $\hat{\alpha}_1 = 1.0988$, $\tilde{\phi}_0 = 0.9346$, $\tilde{\phi}_1 = 0.0664$. Thus, $\tilde{\theta}^u_{100X} = 0.0149$, and $\tilde{\theta}^l_{100X} = 0.0100$, yielding adjusted bounds for CE of $[-0.0408, -0.0359]$. That is, the adjusted bounds are 58% narrower than the unadjusted bounds.

3.8 Identifiability

As noted at the end of Section 3.3, in the absence of covariate X, CE is identifiable if and only if one of the following three conditions occur: $\gamma = 1$, $\pi_1 = 1$, or $\pi_0 = 0$. When at least one of these conditions holds, CE is identifiable and $\theta^u_{100} = \theta^l_{100}$, i.e., the bounds collapse to a single point. In this section we consider conditions under which adjusting for the binary covariate X renders CE identifiable in the sense that the adjusted bounds collapse to a point, i.e., $CE^l_X = CE^u_X$. By the form of the adjusted bounds given in (3.7), it follows that if $0 < \phi_0 < 1$, then the adjusted bounds yield a

single point if and only if

$$\gamma_x = 1, \pi_{1x} = 1, \text{ or } \pi_{1x} = 0 \tag{3.10}$$

for $x = 0$ and $x = 1$. If $\phi_0 = 0$ (and thus $\phi_1 = 1$), then $CE_X^l = CE_X^u$ if and only if (3.10) holds for $x = 1$. Analogously, If $\phi_0 = 1$, then $CE_X^l = CE_X^u$ if and only if (3.10) holds for $x = 0$.

Ding et al. (2011) also considered identifiability of a principal effect when outcomes are truncated by death, which is mathematically identical to the problem considered here. In addition to Assumptions 1 and 2 above, Ding et al. provided two additional assumptions which are sufficient for identifiability: (i) $X_i \perp Y_i|\{S_i(0), S_i(1), Z_i\}$ and (ii) $\Pr[X_i = x|S_i(0) = S_i(1) = 0] \neq \Pr[X_i = x|S_i(0) = 1, S_i(1) = 0]$. Unfortunately, assumption (i), which under randomization can be equivalently stated as $X_i \perp Y_i(z)|\{S_i(0), S_i(1)\}$ for $z = 0, 1$, cannot be verified from the observable data. Ding et al. also gave sufficient identifiability conditions that do not require (i) but instead require that X takes on at least three levels or is continuous and that the mean of Y satisfies a particular linear model.

In contrast, condition (3.10) can be assessed from the observable data because γ_x, π_{1x} and π_{0x} are all identifiable under Assumptions 1 and 2 only. Moreover, (3.10) suggests a strategy for selecting X. In particular, if a covariate X can be found such that (3.10) holds for $x = 0, 1$, then CE will be identifiable. If no such covariate is available, then selecting X such that (3.10) approximately holds for $x = 0, 1$ should yield adjusted bounds with width close to zero. For instance, in the MTCT from the previous section, the low birth weight indicator covariate X yields $\hat{\gamma}_0 = 1$, i.e., (3.10) empirically holds for $x = 0$; while (3.10) does not hold empirically for $x = 1$, $\hat{\pi}_{11} = 0.065$ is not too far from zero and indeed the birth weight adjusted bounds for CE are substantially narrower than the unadjusted bounds.

Finally, we note two special examples where X_i identifies CE. First, suppose $X_i = 1$ if and only if $S_i(0) = S_i(1) = 0$, i.e., X_i is a perfect predictor of membership in the NI principal stratum. Because $\gamma_1 = 1$, the causal effect is identifiable within the stratum where $X_i = 1$, i.e., $\theta^l_{1001} = \theta^u_{1001}$. Furthermore, because $\phi_1 = 1$ and $\phi_0 = 0$, it follows that $\theta^u_{100X} = \theta^u_{1001}$ and $\theta^l_{100X} = \theta^l_{1001}$, implying $CE^l_X = CE^u_X$. This example is related to the "principal score," i.e., the probability an individual is within a principal stratum conditional on one or more covariates (Jo and Stuart, 2009). In practice, principal scores are not known but predicted based on fitted models using the observed data. In the MTCT setting, such a model can be fit using data from infants with $Z_i = 0$ and $S_i = 0$, because under Assumptions 1 and 2 such infants are necessarily in the NI stratum. This first example illustrates that if a set of one or more covariates (not necessarily binary or even discrete) can be found such that the principal scores for the NI stratum equal zero or one for all individuals and dichotomized at some threshold, then CE is identified. In practice this may not be possible; however, if covariates can be found such that the principal scores for the NI stratum are all close to zero or one, i.e., the principal scores are highly predictive of NI stratum membership, then the adjusted bounds constructed by stratifying on the dichotomized principal scores should have width near zero. For the second example, suppose $Y_i = 1$ if and only if $X_i = 1$, i.e., X_i is a perfect predictor of Y_i. Then $\pi_{10} = 0$ and $\pi_{11} = 1$ implying (3.10) holds for $x = 0$ and $x = 1$ and therefore $CE^l_X = CE^u_X$. In settings where Z_i has an effect on Y_i and Z_i is assigned randomly, no such perfect predictor X_i will exist (because X_i is measured pre-randomization), such that the second example seems to have little practical implication.

3.9 Conclusion

In summary, this paper considers when adjusted bounds of the principal effect will be sharper than unadjusted bounds. In particular, Proposition 1 shows that the adjusted bounds cannot be worse, i.e., wider, than the unadjusted bounds. Propositions 2 and 3 give necessary and sufficient conditions for the adjusted upper and lower bounds to be an improvement. Throughout it was assumed that X was a binary covariate, although the results from Section 3.6 can be extended to any categorical baseline covariate. Specifically, Proposition 1 will hold for any categorical baseline covariate X_i with k levels, k finite. Similarly, if there exists any two levels of X that satisfy either (3.8) or (3.9), then either Proposition 2 or Proposition 3 will hold respectively.

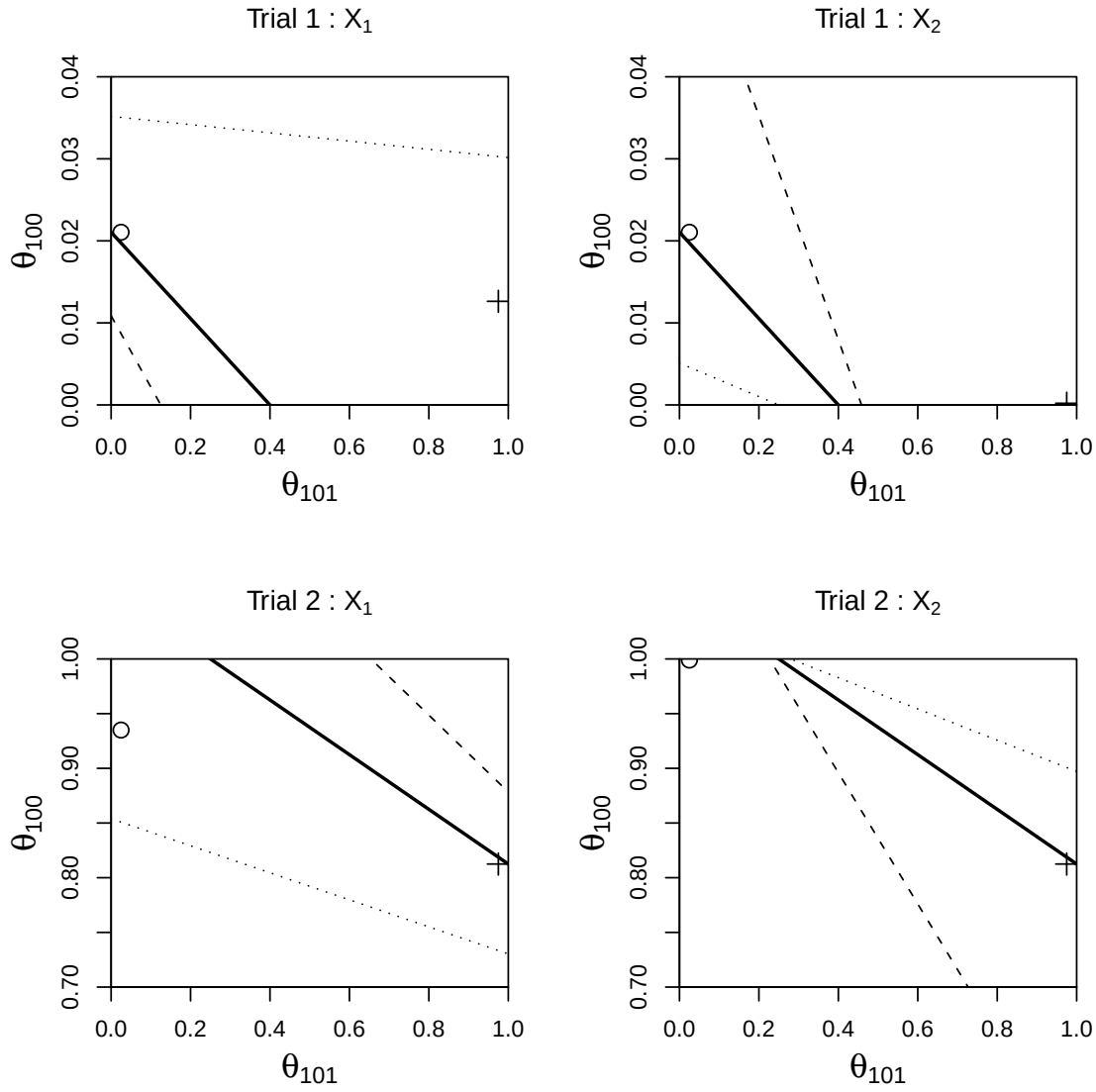

Figure 3.1: Graphical depiction of bounds discussed in Section 3.5 for the two fictional trials from Section 3.5 stratified by two covariates. The solid lines depict equation (3.2) with $\pi_1 = 0.02$ and $\gamma = 0.95$ in the upper panels and $\pi_1 = 0.85$ and $\gamma = 0.8$ in the lower panels. The $\cdots$ $(- - -)$ lines represent (3.5) for $X_i = 0$ ($X_i = 1$). The vertical value of $\circ$ $(+)$ corresponds to θ^u_{100X} (θ^l_{100X}).

Table 3.1: Probabilities from the fictional trials described in Section 3.4 stratified by X_1 and X_2

Trial 1		X_1		X_2	
		0	1	0	1
	γ_x	0.995	0.920	0.980	0.880
	π_{1x}	0.035	0.010	0.005	0.055
	λ_x	0.400	0.600	0.700	0.300
	ϕ_x	0.419	0.581	0.722	0.278
Trial 2		X_1		X_2	
		0	1	0	1
	γ_x	0.890	0.740	0.875	0.625
	π_{1x}	0.760	0.910	0.910	0.710
	λ_x	0.400	0.600	0.700	0.300
	ϕ_x	0.445	0.555	0.766	0.234

Table 3.2: Data from the BAN study stratified by X (low birth weight < 2.5 kg).

	Control			Treatment		
	X			X		
	0	1	Total	0	1	Total
$S = 1$	28	10	38	36	3	39
$S = 0$	584	46	630	751	62	813
$Y = 1$	31	1	32	8	4	12

Chapter 4

Principal Surrogates in Repeated Low-Dose Challenge Experiments

4.1 Introduction

Experiments in non-human primates (NHPs) play an essential role in the development, screening, and evaluation of preventive HIV vaccines (Morgan *et al.* 2008; Fauci *et al.* 2008). Historically, these studies have usually entailed challenges with very high doses of a hybrid simian-human immunodeficiency virus (SHIV), resulting in infection of all NHPs in the experiment. Such an approach uses doses that are much higher than those experienced in natural transmission, so that the failure of a vaccine candidate to protect from infection in this setting is not necessarily indicative of a vaccine without utility in humans (Feinberg and Moore 2002). Thus researchers have recently begun to conduct repeated low-dose challenge (RLC) studies in NHPs (Garcia-Lerma *et al.* 2008; Ellenberger *et al.* 2006; Subbarao *et al.* 2006; Otten *et al.* 2005) that may more closely mimic typical exposure in natural human transmission settings. Recent investigations (Regoes *et al.* 2005; Hudgens and Gilbert 2009; Hudgens *et al.* 2009) have shown that these experiments can be adequately powered to detect vaccine efficacy against infection with a clinically feasible numbers of NHPs. Since investigators

control exposure, challenge experiments can assess vaccine effects of interest that may be difficult to observe in typical human efficacy trials. For example, since infection can be assessed after each exposure via viral load assays, one can easily estimate the per-exposure or "per-contact" effect of vaccination. Likewise, challenge studies allow precise characterization of immunological and virological parameters very early after infection, which are practically impossible to evaluate in humans (except possibly in neonates).

In addition to assessing vaccine effects on infection and post-infection endpoints, evaluation of immunological surrogates of protection (SoPs) are a vital component of NHP challenge experiments, providing points of reference to judge new vaccine candidates and for retrospective analyses of candidates evaluated previously in high-dose challenge experiments (Regoes *et al.* 2005). In general, a SoP is defined to be an immunological variable S_i such that a vaccine effect on S_i is predictive of a vaccine effect on the risk of infection or disease. The utility of such a SoP includes guiding vaccine development, providing guidance for regulatory and immunization policy decisions, and bridging efficacy of a vaccine observed in a trial to a new setting. For RLC challenge studies, knowledge of an immunological surrogate can inform comparisons of vaccine candidates in NHPs and support predictions of vaccine efficacy in humans.

Despite the importance of finding immune SoP, the literature on methods for their quantitative assessment is still quite limited. Moreover, there exists considerably confusion about what constitutes an immune correlate or surrogate of protection and how it is appropriately evaluated. Recently Qin *et al.* (2007) and Gilbert *et al.* (2008) proposed a hierarchical three-tier framework for evaluating immune correlates: correlate of risk (CoR), specific SoP, and general SoP. A CoR is an immunological measurement that correlates with the rate or level of a clinical endpoint in a defined population. A specific SoP is a CoR that is predictive of vaccine efficacy in a particular setting. A

general SoP is a specific SoP that is also predictive of VE_S across different settings (e.g., across populations or across vaccine formulations). Meta-analysis of multiple vaccine studies is required for evaluating a general SoP whereas one study may be sufficient for evaluating a specific SoP. Here attention is restricted to the evaluation of a specific SoP from a single RLC challenge study of a candidate HIV vaccine.

Traditionally, identification of potential specific SoPs has relied on solely assessing whether a immune response was a CoR, i.e., associated with risk of infection or disease. For example, in the first phase III trial of an HIV vaccine, a significant negative association was found between risk of HIV infection and antibody (Ab) response to the vaccine (Gilbert *et al.* 2005). Unfortunately, this purely correlational analysis provides no information to distinguish between two possibilities: (i) a greater vaccine effect on the immune response predicted a greater vaccine effect on infection risk, or (ii) the immune response simply marked an innate ability to escape infection but did not predict vaccine efficacy. In other words, it was not possible based on the observed data to conclude whether Ab response to the vaccine was a SoP or just a CoR. A similar example is given by Ellenberger *et al.* (2006), who found an association between vaccine induced ELISpot Gag responses and risk of SHIV infection in a RLC challenge study evaluating an HIV DNA/MVA vaccine candidate.

Recently, novel experimental designs and corresponding statistical methodology have been proposed for evaluating potential specific SoPs in the context of human efficacy trials (Follmann 2006; Gilbert and Hudgens 2008). The central premise behind these designs is to attempt to elicit the immune response control NHPs would have had if they had been vaccinated. The first design, which is referred to as the baseline immunogenicity predictor or BIP design, entails measuring a baseline covariate(s) W that is correlated with the immune response that NHP would have to the HIV vaccine being evaluated. For example, W might be an immune response to a non-HIV vaccine.

The missing HIV vaccine immune response for NHPs in the control arm can then be predicted from their W and a prediction model based on observed data from the vaccine group. In turn, the association between the vaccine induced immune response and the vaccine effect to prevent infection can be assessed.

The second study design proposed for evaluation of SoPs in human efficacy trials is the close-out placebo vaccination design, where placebo recipients who are uninfected at the end of the trial are administered the HIV vaccine and their immune response is measured. In the RLC challenge study setting, this design may not be feasible since most, if not all, control NHPs are often infected after repeated challenges, e.g., see Ellenberger *et al.* (2006) and Subbarao *et al.* (2006). Therefore this paper presents a proposed modification of close-out placebo vaccination wherein after each challenge uninfected control NHPs are randomly assigned to receive vaccine according to a pre-specified crossover probability. This will be referred to as the crossover vaccination (CrV) design.

Simulation studies of large (e.g., Phase III) randomized human efficacy trials have demonstrated that the additional information provided by employing a BIP or close-out placebo vaccination design can enable assessing the extent to which a CoR is a SoP (Follmann 2006; Gilbert and Hudgens 2008). Similarly, Hudgens and Gilbert (2009) demonstrated the BIP design also holds promise in identifying SoPs in the RLC challenge study setting. However, the feasibility of the BIP design relies heavily on the existence of a baseline covariate W that is correlated with the HIV vaccine immune response. Whether such a covariate will be available may not be evident, especially when evaluating new vaccine candidates. The CrV design, on the other hand, does not require a BIP W. In order to further compare these two designs, methods are proposed for assessing potential immunological SoPs using either the BIP or CrV design. The operating characteristics of the different designs will then evaluated by simulating RLC

challenge studies.

4.2 Methods

To begin consider the usual two-arm RLC challenge study design where no NHPs crossover from control to vaccine. For NHP i, let Z_i be the treatment assignment subject i was randomized to, where $Z_i = 0$ is control and $Z_i = 1$ is vaccine. Let $T_i(z)$ be the potential survival time under assignment z. It is assumed throughout that survival time is measured by the number of exposures (i.e., challenges) until infection with SHIV; thus $T_i(z)$ is a positive integer. Let $S_i(Z_i)$ denote the HIV-specific immune response when assigned treatment Z_i. Assume $S_i(0) = 0$ for all i because vaccine antigens (absent in the control) are necessary to induce an HIV-specific immune response. Let $p(Z_i, S_i(1))$ denote the probability of infection from a single exposure (i.e., the transmission probability) under assignment to treatment Z_i.

The model and likelihoods developed below rely on several key assumptions. First, assume no interference between NHPs and no variations in the type of vaccine such that the stable unit treatment value assumption (SUTVA) holds. The lack of interference between NHPs should hold in this setting since investigators can prevent transmission of SHIV between NHPs. Second assume ignorable treatment assignment, i.e., $Z_i \perp (S_i(1))$, which is insured by the use of randomization in assigning NHPs to receive vaccine or serve as a control. Third, assume the per-challenge probability of infection is independent of the prior number of challenges.

Under these assumptions, following Follmann (2006), a model for the transmission probability is

$$p(Z_i, S_i(1), W_i; \beta) = \Phi\{\beta_1 + \beta_2 Z_i + \beta_3 S_i(1) + \beta_4 Z S_i(1)\}, \tag{4.1}$$

where Φ is the standard normal CDF and $\beta \equiv (\beta_1, \beta_2, \beta_3, \beta_4)$. The linear component of model (4.1) can easily be generalized to include baseline covariates. A key parameter of model (4.1) is β_4 since $\beta_4 \neq 0$ indicates that larger vaccine effects on the immune response S_i are associated with larger vaccine effects on the probability of infection. The parameter β_2 of model (4.1) is also important as it describes the vaccine effect on infection not associated with the vaccine effect on S_i. In a typical RLC challenge study design, $S_i(1)$ is only observed when $Z_i = 1$. Thus, intuitively, it is not expected for β to be fully identifiable without additional information. The Appendix contains the proof that β is indeed not identifiable when $S_i(1)$ follows a normal distribution. On the other hand, β is identifiable when either the BIP or CrV designs are employed (Gilbert and Hudgens 2008).

The average causal effect of the vaccine on survival is defined as $h(E\{T_i(0)|S_i(1)\}, E\{T_i(1)|S_i(1)\})$, where h is some contrast function such that $h(x, y) = 0$ iff $x = y$, e.g., $h(x, y) = x - y$, and $E\{T_i(Z_i)|S_i(1)\}$ is the expected time to infection under treatment Z_i given immune response $S_i(1)$. Thus, $T_i(Z_i)|S_i(Z_i)$ is a geometric random variable with $E(T_i(z)|S_i(z) = s) = 1/p(z, s)$. Following Hudgens and Gilbert (2009), choosing $h(x, y) = \Phi^{-1}(1/x) - \Phi^{-1}(1/y)$ yields convenient causal effects, i.e., $h(E\{T_i(0)|S_i(1) = s\}, E\{T_i(1)|S_i(1) = t\}) = -\beta_2 - \beta_3(t - s) - \beta_4 t$.

The surrogate value of S_i can be gleaned from the two curves $p(0, s; \beta)$ and $p(1, s; \beta)$. Generally, larger $|\beta_4|$ and smaller $|\beta_2|$ will reflect greater surrogate value. Hudgens and Gilbert (2009) suggest the proportion associative effect (PAE) statistic as a summary measure of the surrogate value of S_i. PAE measures the proportion of the total effect of treatment represented by the expected associative effect (EAE) of S_i versus the expected dissociative effect (EDE). EAE measures the association between the vaccine effect on S_i and the vaccine effect on infection with larger values of EAE implying stronger association. EDE measures the effect of vaccine in strata

where $S_i(1) = S_i(0)$ and is termed 'dissociative' since large values of EDE suggest the opposite of an associative effect. Small values of EDE paired with large values of EAE give strong indication that S_i is an SoP. Thus, $PAE = |EAE|/(|EAE| + |EDE|)$. In general, $EAE = E[h(E\{T_i(0)|S_i(1)\}, E\{T_i(1)|S_i(1)\})|S_i(1) > S_i(0)]$ and $EDE = E[h(E\{T_i(0)|S_i(1)\}, E\{T_i(1)|S_i(1)\})|S_i(1) = S_i(0)]$. Using the h described above $EAE = -\beta_2 - \kappa\beta_4$, where $\kappa = E[S_i(1)|S_i(1) > S_i(0) = 0]$, $EDE = -\beta_2$ and

$$PAE \equiv |\beta_2 + \kappa\beta_4|/\{|\beta_2| + |\beta_2 + \kappa\beta_4|\}, \tag{4.2}$$

where the convention $|0|/(|0| + |0|) = 0.5$ is used if $\beta_2 = \beta_4 = 0$. Note if $\beta_4 = 0$, then $PAE = 0.5$, corresponding to no association between the vaccine effect on S_i and the vaccine effect on infection. On the other hand, if $\beta_2 = 0$ and $\beta_4 \neq 0$, then $PAE = 1$, indicating there is a vaccine effect on infection if and only if there is a vaccine effect on S_i. Biomarkers with some surrogate value will have $PAE \in (0.5, 1]$, with the value increasing as PAE nears 1.

4.2.1 Baseline Immunogenicity Predictor Design

The likelihood for the BIP design can be constructed as follows. For NHP i, let $T_i \equiv \min\{T_i(0)(1-Z_i)+T_i(1)Z_i, c_{max}\}$ denote the observed number of exposures during the experiment where c_{max} denotes the right censoring time, i.e., the maximum allowable number of exposures. In general c_{max} may differ from NHP to NHP, but here for simplicity it is assumed c_{max} is the same for all NHPs. Let W_i denote a baseline covariate that might be correlated with $S_i(1)$. Let δ_i equal 1 (0) if subject i is infected (uninfected) by the end of the study and let $S_i \equiv S_i(0)(1-Z_i)+S_i(1)Z_i$ denote the observed immune response. Suppose n iid copies of $(Z_i, T_i, \delta_i, S_i, W_i)$ are observed. Letting G_w denote

the conditional distribution of $S_i(1)$ given W_i, the conditional likelihood is

$$L^{bip}(\beta, G_w) \equiv \prod_{i=1}^{n} \varphi(1, S_i, T_i, \delta_i; \beta)^{Z_i} \left\{ \int \varphi(0, s, T_i, \delta_i; \beta) dG_w(s|W_i) \right\}^{1-Z_i},$$

where $\varphi(Z, S, T, \delta; \beta) \equiv \{1 - p(Z, S; \beta)\}^{T-\delta} p(Z, S; \beta)^{\delta}$.

4.2.2 Crossover Vaccination Design

Next consider the likelihood for the CrV design. Recall in this design that after each challenge uninfected control NHPs are randomly assigned to receive vaccine according to a pre-specified crossover probability (c_p). Let T_i^0 denote the observed number of challenges when the NHP was unvaccinated and δ_i^0 be an indicator which equals one if challenge T_i^0 resulted in infection. Define T_i^1 and δ_i^1 similarly for when the NHP was vaccinated. For NHPs randomly assigned to receive vaccine at the start of the experiment let $T_i^0 = \delta_i^0 = 0$. For NHPs randomly assigned to control at the start of the experiment who become infected prior to crossover or reach the end of the experiment prior to crossover let $T_i^1 = \delta_i^1 = 0$. Let $\tilde{Z}_i$ indicate vaccination at some point during the study, i.e., $\tilde{Z}_i = Z_i + (1 - Z_i) I[T_i^1 > 0]$. Suppose n iid copies of $(\tilde{Z}_i, T_i^0, \delta_i^0, T_i^1, \delta_i^1, S_i(1) I[T_i^1 > 0])$ are observed. Letting G be the marginal distribution of $S_i(1)$, the conditional likelihood for the CrV design is

$$L^{crv}(\beta, G) \equiv \prod_{i=1}^{n} \left\{ (\varphi(0, S_i(1), T_i^0, \delta_i^0; \beta) \varphi(1, S_i(1), T_i^1, \delta_i^1; \beta) \right\}^{\tilde{Z}_i} \left\{ \int \varphi(0, s, T_i^0, \delta_i^0; \beta) dG(s) \right\}^{1-\tilde{Z}_i}. \quad (4.3)$$

If the CrV and BIP designs are combined such that W_i is also observed for $i = 1, \ldots, n$, then the conditional likelihood is the same as L^{crv} except that $dG(s)$ is replaced by $dG_w(s|W_i)$ in (4.3).

4.2.3 Inference

When $S_i(1)$ is normal β is unidentifiable even though G (G_w) can be identified. Therefore, maximum "estimated likelihood" (Pepe and Fleming 1991) or "pseudolikelihood" (Liang and Self 1996) can be used for inference. Maximum estimated likelihood estimation (MELE) entails two steps. First obtain an estimate $\hat{G}$ (or $\hat{G}_w$) using data from the vaccine arm of the study. Then, conditional on $\hat{G}$ (or $\hat{G}_w$), the MELE $\hat{\beta}$ is obtained by maximizing the likelihood L^{bip} (or L^{crv}) with respect to β. The estimator $\widehat{PAE}$ is computed by evaluating (4.2) at $\hat{\beta}$ and $\hat{\kappa} = \int_{s>0} s d\hat{G}(s) / \int_{s>0} d\hat{G}(s)$. A parametric bootstrap test (PBT) (Davidson and Hinkley 1997; Hudgens and Gilbert 2009) can be employed to assess $H_0 : PAE = 0.5$ versus $H_A : PAE > 0.5$, i.e., test the null S_i has no surrogate value versus the alternative S_i has some surrogate value. The null value of $PAE = 0.5$ is found by fixing $\beta_4 = 0$ or $\beta_4 = -2\kappa\beta_2$.

4.3 Simulation study

A simulation study was conducted to assess whether sample sizes typical of RLC challenge studies provide adequate power to detect immune responses with high surrogate value. Data were generated assuming 10, 15, 20, and 25 NHPs were randomized to each arm initially, a maximum number of exposures per NHP of $c_{max} = 10$, and model (4.1). Several RLC challenges studies with roughly 25 NHPs per arm are currently being conducted or planned by the NIH Vaccine Research Center (John Mascola, personal communication).

Simulation studies were conducted where values of $(S_i(1), W_i)$ were generated using the method of Emrich and Piedmonte (1991), which first generates a bivariate normal random variable and truncates it to a create a bivariate binary random variable with

$(S_i(1), W_i)$ each having success probability 1/2 and correlation ρ. Values of β were selected such that the average probability of infection per challenge for controls (vaccines) was 0.5 (0.1). Thus the vaccine efficacy (VE), i.e., percent reduction in the average probability of infection due to vaccination, was (1-0.1/0.5) x 100 = 80%. To reflect immune responses with varying surrogate values, simulated data sets were generated under five scenarios corresponding to $PAE \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$. For the first scenario $\beta = (-1, -1.85, 2, -1.12)$ such that $PAE = 0.7$. Values for β for the remaining scenarios were chosen such that the desired PAE was achieved with $p(1, S_i(1); \beta)$ the same as in the first scenario. Different values of the crossover probability c_p were also investigated to ascertain the optimal c_p to detect a SoP. For each scenario, 1000 simulated data sets were generated. For each data set maximum likelihood estimates $\hat{\beta}$ and $\widehat{PAE}$ were computed and a PBT using 100 boots was performed for $H_0 : PAE = 0.5$ versus $H_0 : PAE > 0.5$ at the $\alpha = 0.10$ significance level.

Simulation results for the case where $(S_i(1), W_i)$ are binary are given in Figure 4.1 and Tables 4.1 and 4.2. In general, there is power to reject $H_0 : PAE = 0.5$ if neither the BIP or CrV designs, which is surprising. However, as shown in the Appendix, β is identifiable, up to absolute value, when $S_i(1)$ is binary. There is an increase in power when using the CrV alone, but it appears that using a small c_p, < 0.2, may be preferred as there was an overall decrease in power for $c_p \geq 0.2$ versus $c_p \leq 0.2$.

One potential drawback of the CrV design is that there are fewer challenges of unvaccinated NHPs, potentially diminishing the power to detect VE. Therefore we also conducted simulation studies to assess how c_p affects the power to reject the null $H_0 : VE = 0$. Simulations were performed for $VE \in \{0, 0.25, 0.50, 0.75\}$ and $c_p \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. The null $H_0 : VE = 0$ is equivalent to testing $H_0 : p(1) - p(0) = 0$, where $p(z)$ is the probability of infection under treatment z, and was conducted using a likelihood ratio test. Results shown in Table 4.3 demonstrate the

CrV design results in modest diminution of power to detect a vaccine effect on infection.

4.4 Discussion

A drawback in determining if a biomarker is a SoP in a RLC study is estimation of $S_i(1)$ in the unvaccinated subjects. Previously, Follmann (2006) developed methods for a BIP to be used to more accurately determine an SoP and also outlined the close-out placebo vaccination (CPV). We have presented an innovative adaptation of his design, the CrV, that provides a number of improvements. If all control subjects are infected under the CPV design, there will be no subjects to receive vaccination at the end of the trial. This scenario is not as likely in a RLC study. Employing the CrV will increase the probability of measuring $S_i(1)$ for some control subjects, given the infection probability at the first challenge is not one. Also, Follmann found that without incorporating a BIP, the CPV design had little power to detect an SoP, whereas we have shown the CrV, while more powerful when a BIP is present, has decent power to detect an SoP. Additionally, the CrV design has little effect on power to detect VE.

Huang and Gilbert (2011) suggest alternative measures to the PAE for assessing SoPs when multiple biomarkers are present. Specifically, the suggest using the standardized total gain (STG), a graphical measure which estimates the amount of variability of the treatment effect that is characterized by the risk difference between treatment arms. Where PAE measures how well $S_i(0) - S_i(1)$ predicts treatment effects, STG measures the amount of variability in $\Pr[Y_i(0) = 1] - \Pr[Y_i(1) = 1]$ explained by $\Pr[Y_i(0) = 1|S_i(0), S_i(1)] - \Pr[Y_i(1) = 1|S_i(0), S_i(1)]$, where Y_i is the outcome of interest and S_i can be a vector of potential surrogates. Future research could entail investigating the applicability of STG in the RLC setting by adapting (4.3) to estimate STG under the CrV design, potentially gaining the flexibility to estimate the surrogate value of multiple biomarkers in a RLC study.

Figure 4.1: Empirical type I error and power to reject $H_0 : PAE = 0.5$ versus $H_A : PAE > 0.5$ from simulation study described in Section 4.3 with 25 NHPs in each arm and $(S_i(1), W_i)$ binary.

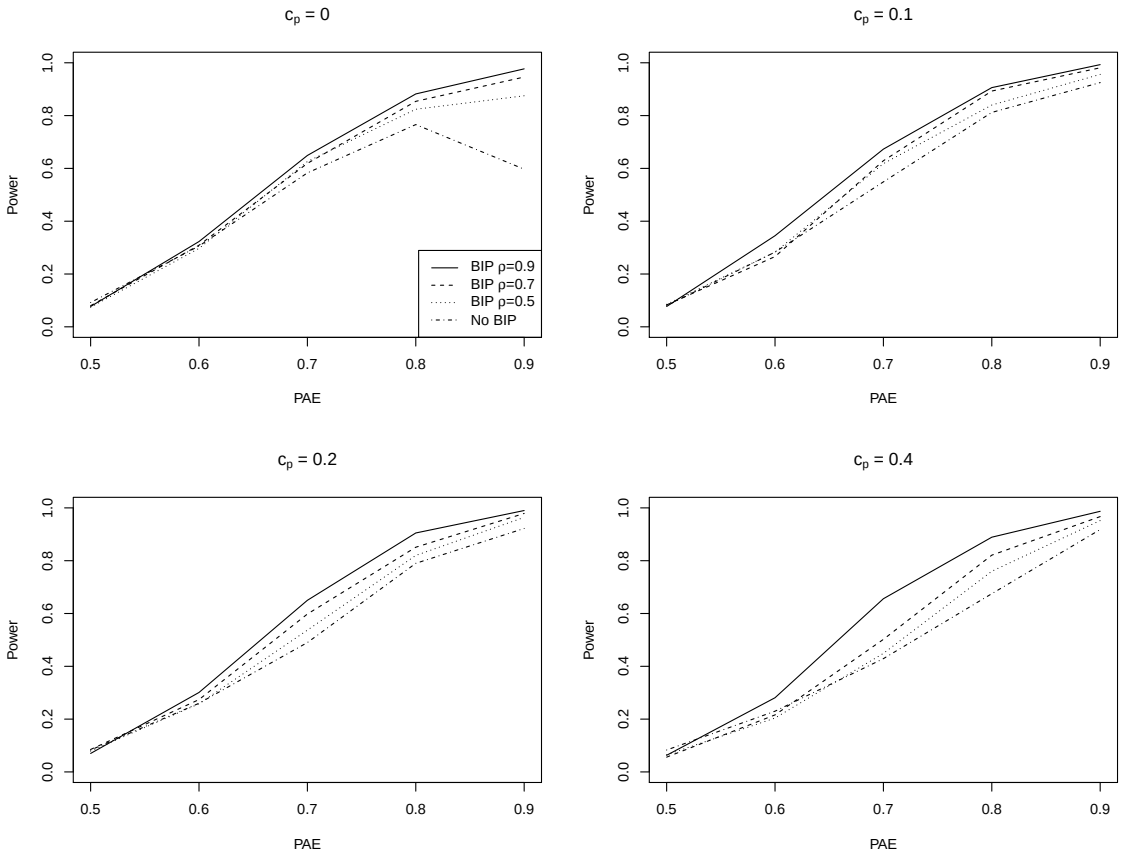

Table 4.1: Results from simulation study described in Section 4.3 with $(S_i(1), W_i)$ binary. Each table entry is based on 500 simulated data sets with m NHPs per arm for $c_p = 0, 0.1$. ρ is the linear correlation. Bias is the median bias and EP is the empirical power for the PBT of $H_0 : PAE = 0.5$ versus $H_A : PAE > 0.5$ at level $\alpha = 0.1$.

			$m = 10$		$m = 15$		$m = 20$		$m = 25$	
c_p	PAE	ρ	Bias	EP	Bias	EP	Bias	EP	Bias	EP
0	0.5	0	0.267	0.067	0.240	0.093	0.242	0.087	0.227	0.092
		0.5	0.259	0.062	0.251	0.084	0.231	0.087	0.230	0.074
		0.9	0.234	0.044	0.238	0.063	0.235	0.083	0.233	0.077
	0.7	0	0.2	0.235	0.055	0.361	0.027	0.462	0.001	0.583
		0.5	0.139	0.247	0.044	0.371	0.020	0.49	-0.001	0.626
		0.9	0.034	0.340	0.032	0.416	0.023	0.546	0.018	0.649
	0.9	0	0.362	0.208	0.221	0.365	0.128	0.522	0.103	0.597
		0.5	0.063	0.360	0.034	0.620	0.015	0.772	0.007	0.875
		0.9	-0.011	0.59	-0.014	0.835	-0.005	0.952	0.003	0.977
0.1	0.5	0	0.219	0.087	0.216	0.092	0.216	0.101	0.211	0.080
		0.5	0.224	0.080	0.220	0.083	0.220	0.081	0.217	0.085
		0.9	0.220	0.064	0.232	0.077	0.235	0.068	0.227	0.077
	0.7	0	0.038	0.314	0.009	0.415	-0.005	0.487	0.002	0.549
		0.5	0.042	0.307	0.016	0.393	0.018	0.483	0.006	0.618
		0.9	0.018	0.314	0.011	0.424	0.012	0.575	0.010	0.673
	0.9	0	0.043	0.439	0.022	0.707	0.009	0.844	0.003	0.925
		0.5	0.025	0.470	0.013	0.771	0.005	0.910	0.010	0.956
		0.9	-0.013	0.616	-0.011	0.857	-0.012	0.956	-0.003	0.993

Table 4.2: Results from simulation study described in Section 4.3 with $(S_i(1), W_i)$ binary. Each table entry is based on 500 simulated data sets with m NHPs per arm for $c_p = 0.2, 0.4$. ρ is the linear correlation. Bias is the median bias and EP is the empirical power for the PBT of $H_0 : PAE = 0.5$ versus $H_A : PAE > 0.5$ at level $\alpha = 0.1$.

			$m = 10$		$m = 15$		$m = 20$		$m = 25$	
c_p	PAE	ρ	Bias	EP	Bias	EP	Bias	EP	Bias	EP
0.2	0.5	0	0.187	0.107	0.208	0.104	0.205	0.088	0.208	0.083
		0.5	0.227	0.068	0.219	0.103	0.215	0.098	0.215	0.079
		0.9	0.211	0.064	0.228	0.065	0.226	0.065	0.229	0.070
	0.7	0	0.034	0.311	0.000	0.394	0.002	0.463	0.012	0.49
		0.5	0.032	0.312	0.004	0.386	0.009	0.468	0.003	0.537
		0.9	-0.015	0.292	-0.012	0.416	0.004	0.547	0.008	0.65
	0.9	0	0.022	0.483	0.023	0.715	0.015	0.856	0.017	0.922
		0.5	0	0.526	0.008	0.784	0.019	0.909	0.014	0.964
		0.9	-0.033	0.609	-0.019	0.871	-0.007	0.958	-0.001	0.990
0.4	0.5	0	0.162	0.089	0.194	0.092	0.208	0.097	0.220	0.083
		0.5	0.158	0.071	0.212	0.097	0.207	0.091	0.216	0.065
		0.9	0	0.06	0.216	0.086	0.217	0.072	0.215	0.063
	0.7	0	0.007	0.252	-0.005	0.386	0.033	0.369	0.023	0.429
		0.5	-0.009	0.259	0.014	0.346	0.027	0.381	0.034	0.449
		0.9	-0.060	0.276	-0.032	0.394	0.000	0.486	-0.005	0.656
	0.9	0	0.027	0.445	0.044	0.658	0.038	0.832	0.046	0.918
		0.5	0.009	0.482	0.024	0.719	0.036	0.879	0.040	0.952
		0.9	-0.041	0.581	-0.014	0.825	-0.014	0.941	0.001	0.987

Table 4.3: Empirical power to reject $H_0 : VE = 0$ where $VE \equiv \{1 - p(1)/p(0)\} \times 100\%$ is the percent reduction in the per-challenge risk of infection due to vaccination. Simulation results assuming 50 NHPs total, with 25 randomized to vaccine initially. The average number of control NHPs that crossover are given in parentheses. The parenthetical numbers in the column headers denote the theoretical expected value of the number of NHPs that crossover.

	Crossover probability (c_p)											
VE	0	(0)	0.1	(2.3)	0.2	(4.2)	0.3	(5.8)	0.4	(7.1)	0.5	(8.3)
0	0.04	(0)	0.06	(2.3)	0.06	(4.1)	0.06	(5.8)	0.06	(7.2)	0.05	(8.4)
0.25	0.26	(0)	0.25	(2.3)	0.24	(4.2)	0.26	(5.8)	0.23	(7.2)	0.22	(8.4)
0.5	0.89	(0)	0.87	(2.3)	0.83	(4.2)	0.82	(5.8)	0.8	(7.2)	0.77	(8.3)
0.75	1.00	(0)	1.00	(2.2)	1.00	(4.2)	1.00	(5.8)	1.00	(7.2)	1.00	(8.3)

Chapter 5

Conclusion

To summarize, assessing the causal effect of interventions to prevent infectious diseases can be difficult in many settings. We have provided methodology to assess the principal effect of treatment on competing risks outcomes. This work can be considered an extension of Shepherd *et al.* (2007), who measured the principal effect on a survival endpoint. Nonparametric bounds and sensitivity methods were created to determine how the potential selection bias affects the estimate of the principal effect. These methods were evaluated by simulation studies and found to be precise and accurate. We then analyzed the BAN study, a recent large study of mother-to-child transmission of HIV.

Because unadjusted bounds of principal effects are frequently uninformative, we have demonstrated methods to calculate adjusted bounds. These bounds, as outlined by Grilli and Mealli (2008), incorporate information from a binary baseline covariate X to create adjusted bounds. Unadjusted bounds are found within strata of X and are averaged to create the overall adjusted bounds. We have shown that the adjusted bounds can improve on the unadjusted bounds, i.e, be narrower than, and have provided the necessary and sufficient conditions when this will occur. Relaxation of the monotonicity assumption and methods for adjusting for multiple covariates are two

possible directions for future research.

Additionally, surrogate evaluation requires consideration of novel study designs. Accordingly, we developed a new design for evaluation of surrogates of vaccine protection in RLC challenge experiments, the crossover vaccination (CrV) design. This new design was an adaptation of the close-out placebo vaccination of Follmann (2006) and also incorporated the use of a baseline covariate to improve estimation. The CrV was investigated under different scenarios using simulation studies and found to be adequate in the estimation and testing of potential surrogates, measured using the proportion associative effect (PAE) and vaccine efficacy (VE). It is believed that the CrV can be used to estimate of other measures of surrogacy, such as the standardized total gain developed by Huang and Gilbert (2011).

Causal inference, more specifically principal stratification, is the thread that binds this dissertation together. Chapters 2 and 3 provide methods for analysis within a principal stratum of interest while the approach addressed in Chapter 4 compares quantities across all strata, i.e., compares effects in the stratum where $S_i(0) = S_i(1)$ to the effects in the strata where $S_i(0) \neq S_i(1)$. For example, for the BAN study addressed in Section 2.5, the causal effect that was estimated is for infants that would never infected under either treatment assignment, $(S_i(0) = S_i(1) = 0)$ whereas the measure of a biomarker's potential as a surrogate of protection is averaged across all possible values of $S_i(1)$ conditional on $S_i(0) = 0$.

In general, principal stratification should be viewed as a tool for researchers to use in settings where traditional methods do not yield causal interpretations. As pointed out by Pearl (2011), one should be careful when using principal stratification to define research questions but instead let it be an aid to answer causal questions when other methods fail. However, in some settings when a proper or well defined question may be lacking, principal stratification can be used to guide future research, i.e, the search

for surrogates of protection, specifically see Gilbert, Hudgens, and Wolfson (2011). Thus, I believe further research in principal stratification methods will allow investigators, specifically those of public health issues, to conduct studies and trials that more efficiently and accurately determine a treatment's casual effect on an outcome.

Appendix I

Asymptotic Results for Chapter 2

Asymptotic Variances of $\hat{F}_1^{NI,up}(t,j)$ and $\hat{F}_1^{NI,low}(t,j)$

To derive the asymptotic variances of $\hat{F}_1^{NI,up}(t,j)$ and $\hat{F}_1^{NI,low}(t,j)$, we first derive the large sample variance of $\hat{\gamma}$. Under monotonicity, it is straightforward to show $\hat{\gamma} - (N_0/n_0)/\ (N_1/n_1) \xrightarrow{p} 0$, implying $\hat{\gamma}$ and $(N_0/n_0)/(N_1/n_1)$ have the same limiting distribution; therefore for the derivation below we can assume $\hat{\gamma} = (N_0/n_0)/(N_1/n_1)$. For $z, s = 0, 1$, define $p_{zs} = \sum I[Z_i = z, S_i = s]/n$ and $\pi_{zs} = \Pr[Z_i = z, S_i = s]$, and let $\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11})'$ and $\boldsymbol{\pi} = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})'$. Define the function g as $g(\boldsymbol{\pi}) = \pi_{00}(\pi_{10} + \pi_{11})/\{\pi_{10}(\pi_{00} + \pi_{01})\}$ and note that $g(\mathbf{p}) = \hat{\gamma}$ and $g(\boldsymbol{\pi}) = \gamma$. Then by the multivariate central limit theorem and the delta method (e.g., see Agresti 2002 , page 580), $\sqrt{n}(\hat{\gamma}-\gamma) = \sqrt{n}\{g(\mathbf{p})-g(\boldsymbol{\pi})\} \xrightarrow{\mathbb{D}} N(0, \sigma_\gamma^2)$ where $\sigma_\gamma^2 = \sum_{z,s=0}^{1} \pi_{zs}(\nabla g_{zs})^2 - (\sum_{z,s=0}^{1} \pi_{zs} \nabla g_{zs})^2$ and $\nabla g_{zs} = \partial g(\boldsymbol{\pi})/\partial \pi_{zs}$. It follows from straightforward algebra that $\sigma_\gamma^2 = \gamma^2[\pi_{01}/\{\pi_{00}(\pi_{00} + \pi_{01})\} + \pi_{11}/\{\pi_{10}(\pi_{10} + \pi_{11})\}]$ for which a consistent estimator is $\hat{\sigma}_\gamma^2 = \hat{\gamma}^2 n(1/N_0 - 1/n_0 + 1/N_1 - 1/n_1)$.

For fixed t and j, let $\boldsymbol{\theta}_{tj} = (F_1(t,j), \gamma)'$ and $\hat{\boldsymbol{\theta}}_{tj} = (\hat{F}_1(t,j), \hat{\gamma})'$. Under the conditions stated in Section 2.3.1 of the main text, in particular assuming equation (2.6), it is straightforward to show $\hat{F}_1^{NI,up}(t,j) - \hat{F}_1(t,j)/\hat{\gamma} \xrightarrow{p} 0$, implying $\hat{F}_1^{NI,up}(t,j)$ and $\hat{F}_1(t,j)/\hat{\gamma}$ have the same limiting distribution. Therefore we can assume $\hat{F}_1^{NI,up}(t,j) = \hat{F}_1(t,j)/\hat{\gamma}$ and, analogously, by equation (2.8) of the main text we can assume $\hat{F}_1^{NI,low}(t,j) = \{\hat{F}_1(t,j)-(1-\hat{\gamma})\}/\hat{\gamma}$. Define the vector of functions $\boldsymbol{h}(x,y) = (x/y, \{x-(1-y)\}/y)'$ such that $\boldsymbol{h}(\hat{\boldsymbol{\theta}}_{tj}) = (\hat{F}_1^{NI,up}(t,j), \hat{F}_1^{NI,low}(t,j))'$. Because $\hat{F}_1(t,j)$ and $\hat{\gamma}$ are

consistent and asymptotically normal, by the delta method $\sqrt{n}\{\boldsymbol{h}(\hat{\boldsymbol{\theta}}_{tj}) - \boldsymbol{h}(\boldsymbol{\theta}_{tj})\} \xrightarrow{\mathbb{D}} N(0, \nabla\boldsymbol{h}(\boldsymbol{\theta}_{tj})\Sigma_{tj}\nabla\boldsymbol{h}(\boldsymbol{\theta}_{tj})')$ where

$$\nabla\boldsymbol{h}(x,y) = \begin{bmatrix} 1/y & -x/y^2 \\ 1/y & (1-x)/y^2 \end{bmatrix}, \qquad \Sigma_{tj} = \begin{bmatrix} \sigma_{tj}^2 & 0 \\ 0 & \sigma_\gamma^2 \end{bmatrix},$$

and σ_{tj}^2 is the asymptotic variance of $\sqrt{n}\{\hat{F}_1(t,j) - F_1(t,j)\}$ such that in large samples $\text{var}\{\hat{F}_1(t,j)\} = \sigma_{tj}^2/n$. It follows that $\hat{F}_1^{NI,up}(t,j)$ and $\hat{F}_1^{NI,low}(t,j)$ are asymptotically normal with variances

$$\text{var}\{\hat{F}_1^{NI,up}(t,j)\} = \frac{\text{var}\{\hat{F}_1(t,j)\}}{\gamma^2} + \frac{F_1(t,j)^2\sigma_\gamma^2}{n\gamma^4}, \tag{5.1}$$

and

$$\text{var}\{\hat{F}_1^{NI,low}(t,j)\} = \frac{\text{var}\{\hat{F}_1(t,j)\}}{\gamma^2} + \frac{\{1-F_1(t,j)\}^2\sigma_\gamma^2}{n\gamma^4}. \tag{5.2}$$

Replacing $\text{var}\{\hat{F}_1(t,j)\}$, γ, $F_1(t,j)$, and σ_γ^2 in (A.1) and (A.2) with $\widehat{\text{var}}\{\hat{F}_1(t,j)\}$, $\hat{\gamma}$, $\hat{F}_1(t,j)$, and $\hat{\sigma}_\gamma^2$ yields equations (2.7) and (2.9) from the main text.

Appendix II

Proofs from Chapter 3

Proof of Proposition 1

Note

$$\begin{aligned}\theta^u_{100X} &= \sum_x \theta^u_{100x}\phi_x = \sum_x \min\left\{\frac{\pi_{1x}}{\gamma_x}, 1\right\}\phi_x \leq \min\left\{\sum_x \frac{\pi_{1x}}{\gamma_x}\phi_x, \sum_x \phi_x\right\} \\ &= \min\left\{\frac{\pi_1}{\gamma}, 1\right\} = \theta^u_{100},\end{aligned}$$

where the inequality holds since $\min\{a_1, b_1\} + \min\{a_2, b_2\} \leq \min\{a_1 + a_2, b_1 + b_2\}$ and the third equality holds because

$$\begin{aligned}\sum_x \frac{\pi_{1x}}{\gamma_x}\phi_x &= \sum_x \frac{\Pr[Y_i(1) = 1|S_i(1) = 0, X_i = x]}{\Pr[S_i(0) = 0|S_i(1) = 0, X_i = x]} \Pr[X_i = x|S_i(0) = S_i(1) = 0] \\ &= \sum_x \frac{\Pr[Y_i(1) = 1, S_i(1) = 0, X_i = x] \Pr[S_i(1) = 0, X_i = x] \Pr[X_i = x, S_i(0) = S_i(1) = 0]}{\Pr[S_i(1) = 0, X_i = x] \Pr[S_i(0) = S_i(1) = 0, X_i = x] \Pr[S_i(0) = S_i(1) = 0]} \\ &= \sum_x \frac{\Pr[Y_i(1) = 1, S_i(1) = 0, X_i = x]}{\Pr[S_i(0) = S_i(1) = 0]} = \sum_x \frac{\Pr[Y_i(1) = 1, X_i = x|S_i(1) = 0]}{\Pr[S_i(0) = 0|S_i(1) = 0]} \\ &= \frac{\pi_1}{\gamma}. \qquad (5.3)\end{aligned}$$

Similarly for the lower bound,

$$\begin{aligned}\theta^l_{100X} &= \sum_x \theta^l_{100x}\phi_x = \sum_x \max\left\{\frac{\pi_{1x} - (1 - \gamma_x)}{\gamma_x}, 0\right\}\phi_x \\ &\geq \max\left\{\sum_x \frac{\pi_{1x} - (1 - \gamma_x)}{\gamma_x}\phi_x, 0\right\} = \max\left\{\frac{\pi_1 - (1 - \gamma)}{\gamma}, 0\right\} = \theta^l_{100},\end{aligned}$$

where the inequality holds because $\max\{a_1, 0\} + \max\{a_2, 0\} \geq \max\{a_1 + a_2, 0\}$ and the

third equality holds because of (5.3) and

$$\begin{aligned}\sum_x \frac{1-\gamma_x}{\gamma_x}\phi_x &= \frac{\Pr[S_i(0)=1|S_i(1)=0,X_i=x]}{\Pr[S_i(0)=0|S_i(1)=0,X_i=x]}\Pr[X_i=x|S_i(0)=S_i(1)=0] \\ &= \sum_x \frac{\Pr[S_i(0)=1,S_i(1)=0,X_i=x]}{\Pr[S_i(1)=0,X_i=x]}\frac{\Pr[S_i(1)=0,X_i=x]\Pr[X_i=x,S_i(0)=S_i(1)=0]}{\Pr[S_i(0)=S_i(1)=0,X_i=x]\Pr[S_i(0)=S_i(1)=0]} \\ &= \sum_x \frac{\Pr[S_i(0)=1,S_i(1)=0,X_i=x]}{\Pr[S_i(0)=S_i(1)=0]} \;=\; \sum_x \frac{\Pr[S_i(0)=1,X_i=x|S_i(1)=0]}{\Pr[S_i(0)=0|S_i(1)=0]} \\ &= \frac{1-\gamma}{\gamma}. \;\square\end{aligned}$$

Proof of Proposition 2

First, suppose (3.8) holds and, without loss of generality, assume $\pi_{10} < \gamma_0$ and $\pi_{11} > \gamma_1$ which implies that $\theta^u_{1000} = \pi_{10}/\gamma_0$ and $\theta^u_{1001} = 1$. If $\theta^u_{100} = \pi_1/\gamma$ then,

$$\theta^u_{100X} \;=\; \sum_x \theta^u_{100x}\phi_x \;=\; \frac{\pi_{10}}{\gamma_0}\phi_0 + \phi_1 \;<\; \frac{\pi_{10}}{\gamma_0}\phi_0 + \frac{\pi_{11}}{\gamma_1}\phi_1 \;=\; \frac{\pi_1}{\gamma} \;=\; \theta^u_{100},$$

where the inequality holds because $\pi_{11}/\gamma_1 > 1$. Likewise, if $\theta^u_{100} = 1$ then,

$$\theta^u_{100X} \;=\; \sum_x \theta^u_{100x}\phi_x \;=\; \frac{\pi_{10}}{\gamma_0}\phi_0 + \phi_1 \;<\; \phi_0 + \phi_1 \;=\; 1 \;=\; \theta^u_{100},$$

where the inequality holds since $\pi_{10}/\gamma_0 < 1$ Thus, if (3.8) is satisfied by X then $\theta^u_{100X} < \theta^u_{100}$.

Now suppose that (3.8) is not satisfied. Suppose that $\pi_1 < \gamma$, which implies that $\theta^u_{100} = \pi_1/\gamma$. Furthermore suppose $\pi_{1x} > \gamma_x$ for $x = 0, 1$. Thus, $\lambda_0\pi_{10} > \lambda_0\gamma_0$ and $\lambda_1\pi_{11} > \lambda_1\gamma_1$ implying that

$$\begin{aligned}\lambda_0\pi_{10} + \lambda_1\pi_{11} &> \lambda_0\gamma_0 + \lambda_1\gamma_1 \\ \pi_1 &> \gamma,\end{aligned}$$

which is a contradiction. Thus, $\pi_{1x} < \gamma_x$ for $x = 0, 1$ and $\theta^u_{100x} = \pi_{1x}/\gamma_x$ which gives

$$\theta^u_{100X} \;=\; \sum_x \theta^u_{100x}\phi_x \;=\; \pi_{10}/\gamma_0\phi_0 + \pi_{11}/\gamma_1\phi_1 \;=\; \pi_1/\gamma \;=\; \theta^u_{100}.$$

A analogous argument exists when $\pi_1 > \gamma$ and combined with the result above we conclude that when (3.8) is not satisfied $\theta^u_{100X} = \theta^u_{100}$. Therefore, $\theta^u_{100X} < \theta^u_{100}$ if and only if X satisfies (3.8). □

Proof of Proposition 3

Without loss of generality, assume $\pi_{10} > (1-\gamma_0)$ and $\pi_{11} < (1-\gamma_1)$ which implies that $\theta^l_{1000} = \{\pi_{10} - (1-\gamma_0)\}/\gamma_0$ and $\theta^l_{1001} = 0$. If $\theta^l_{100} = \{\pi_1 - (1-\gamma)\}/\gamma$ then,

$$\begin{aligned}\theta^l_{100X} &= \sum_x \theta^l_{100x}\phi_x = \frac{\pi_{10} - (1-\gamma_0)}{\gamma_0}\phi_0 > \frac{\pi_{10} - (1-\gamma_0)}{\gamma_0}\phi_0 + \frac{\pi_{11} - (1-\gamma_1)}{\gamma_1}\phi_1 \\ &= \frac{\pi_1 - (1-\gamma)}{\gamma} = \theta^l_{100},\end{aligned}$$

where the inequality holds since $\{\pi_{11} - (1-\gamma_1)\}/\gamma_1 < 0$. Likewise, if $\theta^l_{100} = 0$ then,

$$\theta^l_{100X} \;=\; \sum_x \theta^l_{100x}\phi_x \;=\; \frac{\pi_{10} - (1-\gamma_0)}{\gamma_0}\phi_0 \;>\; 0 \;=\; \theta^l_{100}.$$

where the inequality holds since $\{\pi_{10} - (1-\gamma_0)\}/\gamma_0 > 0$. Thus, if (3.9) is satisfied by X then $\theta^l_{100X} > \theta^l_{100}$.

Suppose that $\pi_1 > 1-\gamma$, which implies that $\theta^l_{100} = \{\pi_1 - (1-\gamma)\}/\gamma$. Now suppose that (3.9) is not satisfied and furthermore suppose $\pi_{1x} < 1-\gamma_x$ for $x = 0, 1$. For $\lambda_x > 0$, $\lambda_0\pi_{10} < \lambda_0(1-\gamma_0)$ and $\lambda_1\pi_{11} < \lambda_1(1-\gamma_1)$ implying that

$$\begin{aligned}\lambda_0\pi_{10} + \lambda_1\pi_{11} &< \lambda_0(1-\gamma_0) + \lambda_1(1-\gamma_1) \\ \pi_1 &< 1-\gamma,\end{aligned}$$

which is a contradiction. Thus, $\pi_{1x} > 1-\gamma_x$ for $x = 0, 1$ and $\theta^l_{100x} = \{\pi_{1x} - (1-\gamma_x)\}/\gamma_x$ which gives

$$\theta^l_{100X} = \sum_x \theta^l_{100x}\phi_x = \frac{\pi_{10} - (1-\gamma_0)}{\gamma_0}\phi_0 + \frac{\pi_{11} - (1-\gamma_1)}{\gamma_1}\phi_1 = \{\pi_1 - (1-\gamma)\}/\gamma = \theta^l_{100}.$$

A analogous argument exists when $\pi_1 < 1 - \gamma$ and combined with the previous result we conclude that when (3.9) is not satisfied $\theta^l_{100X} = \theta^l_{100}$. Therefore, $\theta^l_{100X} > \theta^l_{100}$ if and only if X satisfies (3.9). $\square$

Appendix III

Identifiability Results from Chapter 4

Here we show $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ is not identifiable in the usual RLC challenge study design (i.e., without CrV or BIP) assuming $S_i(1)$ is binary. Since treatment assignment is random and $S_(1)$ is observed in all NHPs randomized to $Z_i = 1$, $G(s)$ is identifiable and can be regarded as fixed and known. Assume the βs are all finite, so that there is a positive probability of (not) observing an infection under either treatment assignment. Also for now assume $G(s)$ is discrete and the mass is not concentrated on a single point.

Under these assumptions, $\beta_1 + \beta_2$ and $\beta_3 + \beta_4$ are identifiable. To see this, suppose there are two parameterizations $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ and $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3, \tilde{\beta}_4)$ such that

$$\Pr[T_i = t, \delta_i = d, S_i = s | Z_i = 1; \beta] = \Pr[T_i = t, \delta_i = d, S_i = s | Z_i = 1; \tilde{\beta}] \tag{5.4}$$

for all t, d, s. Then, for some $s_1 \neq s_2$, t = 1, d = 1, (5.4) implies

$$\Phi\{\beta_1 + \beta_2 + (\beta_3 + \beta_4)s_1\} = \Phi\{\tilde{\beta}_1 + \tilde{\beta}_2 + (\tilde{\beta}_3 + \tilde{\beta}_4)s_1\}$$

and

$$\Phi\{\beta_1 + \beta_2 + (\beta_3 + \beta_4)s_2\} = \Phi\{\tilde{\beta}_1 + \tilde{\beta}_2 + (\tilde{\beta}_3 + \tilde{\beta}_4)s_2\}.$$

Since Φ is invertible, this implies

$$\beta_1+\beta_2+(\beta_3+\beta_4)s_1 = \tilde{\beta}_1+\tilde{\beta}_2+(\tilde{\beta}_3+\tilde{\beta}_4)s_1 \text{ and } \beta_1+\beta_2+(\beta_3+\beta_4)s_2 = \tilde{\beta}_1+\tilde{\beta}_2+(\tilde{\beta}_3+\tilde{\beta}_4)s_2,$$

implying $\beta_1+\beta_2 = \tilde{\beta}_1+\tilde{\beta}_2$ and $\beta_3+\beta_4 = \tilde{\beta}_3+\tilde{\beta}_4$, thus $\beta_1+\beta_2$ and $\beta_3+\beta_4$ are identifiable

The more interesting issue is whether β_2 and β_4 individually are identifiable from the observable data. To show this we will show that β_1 and β_3 are identifiable. Similar to the argument above, suppose there are two parameterizations $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ and $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\beta}_3, \tilde{\beta}_4)$ such that

$$\Pr[T_i = t, \delta_i = d, S_i = s | Z_i = 0; \beta] = \Pr[T_i = t, \delta_i = d, S_i = s | Z_i = 0; \tilde{\beta}] \tag{5.5}$$

Then for t = 1, d = 1, (5.5) implies

$$\int \Phi(\beta_1 + \beta_3 s) dG(s) = \int \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s) dG(s) \tag{5.6}$$

Now if we were considering a single dose challenge study (i.e., $c_{max} = 1$), then the only other possible pattern of observed data under $Z_i = 0$ would be t = 1, d = 0, in which case (5.5) implies

$$\int \{1 - \Phi(\beta_1 + \beta_3 s)\} dG(s) = \int \{1 - \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s)\} dG(s) \tag{5.7}$$

Now suppose, without loss of generality that $\tilde{\beta}_3 \neq 0$. Then if we let and $\beta_3 = 0$ and $\beta_1 = \Phi^{-1}\{\int \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s) dG(s)\}$ then (5.6) and (5.7) hold, yet $\beta_3 \neq \tilde{\beta}_3$. That is, β_1 and β_3 are not identifiable in a single dose challenge study. Fortunately, in repeated low dose studies ($c_{max} > 1$) we are not limited to the two observed data patterns above. In particular, for t = 2, d = 1, (5.5) implies

$$\int \{1 - \Phi(\beta_1 + \beta_3 s)\} \Phi(\beta_1 + \beta_3 s) dG(s) = \int \{1 - \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s)\} \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s) dG(s) \tag{5.8}$$

Together (5.6) and (5.8) imply

$$\int \Phi(\beta_1 + \beta_3 s)^2 dG(s) = \int \Phi(\tilde{\beta}_1 + \tilde{\beta}_3 s)^2 dG(s). \tag{5.9}$$

Now the question becomes whether (5.6) and (5.9) together imply $\beta_1 = \tilde{\beta}_1$ and $\beta_3 = \tilde{\beta}_3$. This seems plausible since we have two equations and two unknowns. Below we provide proof that β_1 and β_3 are identifiable if $S_i(1)$ is binary.

Assume $S_i(1)$ is binary with $\Pr[S_i(1) = 1] = \theta$, $\Pr[S_i(1) = 0] = 1 - \theta$, $0 < \theta < 1$. Below we prove if $\theta \neq 1/2$, then β_1 and β_3 are identifiable. If $\theta = 1/2$, only $|\beta_3|$ is identifiable. To begin, note (5.5) is equivalent to

$$(1-\theta)A^{t-d}(1-A)^d + \theta B^{t-d}B^d = (1-\theta)C^{t-d}(1-C)^d + \theta D^{t-d}D^d$$

for all t, d where $A = 1-\Phi(\beta_1)$, $B = 1-\Phi(\beta_1+\beta_3)$, $C = 1-\Phi(\tilde{\beta}_1)$, $D = 1-\Phi(\tilde{\beta}_1+\tilde{\beta}_3)$. Showing $A = C$ and $B = D$ is equivalent to proving β_1 and β_3 are identifiable. Since Φ is one-to-one function and $1 - \Phi(\beta_1) = A = C = 1 - \Phi(\tilde{\beta}_1)$ implies $\beta_1 = \tilde{\beta}_1$ (and similarly $\beta_3 = \tilde{\beta}_3$). For d = 0, we have

$$(1-\theta)A^t + \theta B^t = (1-\theta)C^t + \theta D^t \quad \text{for } t = 1, 2, \ldots \tag{5.10}$$

Now for t = 1 we have

$$(1-\theta)(A-C) = \theta(D-B) \tag{5.11}$$

If $A = C$, then $B = D$ and identifiability is proved. So assume by way of contradiction $A \neq C$, and thus $D \neq B$. Now for t = 2, 3 we have

$$(1-\theta)(A^2-C^2) = \theta(D^2-B^2) \tag{5.12}$$

$$(1-\theta)(A^3-C^3) = \theta(D^3-B^3) \tag{5.13}$$

Dividing (5.13) and (5.13) by (5.11) yields

$$A + C = D + B) \tag{5.14}$$

$$A^2 + AC + C^2 = D^2 + DB + B^2 \tag{5.15}$$

Squaring (5.15) and subtracting it by (5.15) yields

$$AC = BD \tag{5.16}$$

Squaring (5.15) and then subtracting it by four times of (5.16) on both sides gives

$$(A - C)^2 = (B - D)^2 \tag{5.17}$$

Assume without loss of generality that $A > C$. Then (5.11) implies $D > B$ which by (5.17) in turn implies

$$A - C = B - D. \tag{5.18}$$

Note (5.11) implies

$$A - C = \frac{\theta}{1 - \theta}(D - B). \tag{5.19}$$

Together (5.18) and (5.19) imply

$$0 = \frac{1 - 2\theta}{1 - \theta}(D - B).$$

If $\theta \neq 1/2$, this implies $D = B$, a contradiction. Thus if $\theta \neq 1/2$, then β_1 and β_3 are identifiable, and likewise β_2 and β_4.

Now suppose $\theta = 1/2$. Equation (5.5) is equivalent to

$$A^t + B^t = C^t + D^t \text{ for } t = 1, 2, \ldots \tag{5.20}$$

There are at least two sets of solution to (5.20). One is $A = C$ and $B = D$, which implies $\beta_1 = \tilde{\beta}_1$ and $\beta_3 = \tilde{\beta}_3$. Another solution is $A = D$ and $B = C$, or consequently $\beta_1 = \tilde{\beta}_1 + \tilde{\beta}_3$ and $\beta_1 + \beta_3 = \tilde{\beta}_3$, which implies $\beta_3 = -\tilde{\beta}_3$. Thus if $\theta = 1/2$, only $|\beta_3|$ is identifiable implying that β_4 is not identifiable because if $\beta_3 > 0$ then $\beta 4 + \beta_3 - |\beta_3| = \beta_4$ but if $\beta_3 < 0$ then $\beta_4 + \beta_3 - |\beta_3| = \beta_4 + 2\beta_3$.

Bibliography

[1] O. Aalen, O. Borgan, and H. Gjessing. *Suvival and Event History Analysis: A Process Point of View*. Springer, New York, 2008.

[2] O. Aalen and S. Johansen. An empirical transition matrix for non-homogeneous markov chains based on censored observations. *Scandinavian Journal of Statistics*, 5(3):141–150, 1978.

[3] A. Agresti. *Categorical Data Analysis, Second edition*. John Wiley and Sons, New York, 2002.

[4] A. Alioum, F. Dabis, L. Dequae-Merchadou, G. Haverkamp, M. Hudgens, J. Hughes, J. Karon, V. Leroy, M. Newell, B. Richardson, and G. Weverling. Estimating the efficacy of interventions to prevent mother-to-child transmission of HIV in breast-feeding populations: development of a consensus methodology. *Statistics in Medicine*, 20:3539–3556, 2001.

[5] P. Andersen, S. Abildstrom, and S. Rosthøj. Competing risks as a multi-state model. *Statistical Methods in Medical Research*, 11(2):203–215, 2002.

[6] A. Bedri, B. Gudetta, A. Isehak, S. Kumbi, S. Lulseged, Y. Mengistu, et al. Extended-dose nevirapine to 6 weeks of age for infants to prevent HIV transmission via breastfeeding in Ethiopia, India, and Uganda: an analysis of three randomised controlled trials. *Lancet*, 372:300–313, 2008.

[7] M. Bekaert, S. Vansteelandt, and K. Mertens. Adjusting for time-varying confounding in the subdistribution analysis of a competing risk. *Lifetime Data Analysis*, 16:45–70, 2010.

[8] M. Boily, R. Baggaley, L. Wang, B. Masse, R. White, R. Hayes, and M. Alary. Heterosexual risk of HIV-1 infection per sexual act: systematic review and meta-analysis of observational studies. *The Lancet Infectious Diseases*, 9(2):118–129, 2009.

[9] M. Bulterys, S. Ellington, and A. P. Kourtis. HIV-1 and breastfeeding: Biology of transmission and advances in prevention. *Clinics in Perinatology*, 37(4):807 – 824, 2010.

[10] M. Bulterys, M. Fowler, K. Van Rompay, and A. Kourtis. Prevention of mother-to-child transmission of HIV-1 through breast-feeding: past, present, and future. *Journal of Infectious Diseases*, 189(12):2149, 2004.

[11] C. Chasela, M. Hudgens, D. Jamieson, D. Kayira, M. Hosseinipour, A. Kourtis, F. Martinson, G. Tegha, R. Knight, Y. Ahmed, D. Kamwendo, I. Hoffman, S. Ellington, Z. Kacheche, A. Soko, J. Wiener, S. Fiscus, P. Kazembe, I. Mofolo, M. Chigwenembe, D. Sichali, and C. van der Horst. Maternal or infant antiretroviral drugs to reduce HIV-1 transmission. *New England Journal of Medicine*, 362(24):2271–2281, 2010.

[12] S. Cole and C. Frangakis. The consistency statement in causal inference: A definition or an assumption? *Epidemiology*, 20(1):3, 2009.

[13] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Application*. Cambridge University Press, 1997.

[14] K. De Cock, M. Fowler, E. Mercier, I. de Vincenzi, J. Saba, E. Hoff, D. Alnwick, M. Rogers, and N. Shaffer. Prevention of mother-to-child HIV transmission in resource-poor countries. *Journal of the American Medical Association*, 283(9):1175, 2000.

[15] P. Ding, Z. Geng, W. Yan, and X. Zhou. Identifiability and estimation of causal effects by principal stratification with outcomes truncated by death. *Journal of the American Statistical Association*, 106(496):1578–1591, 2011.

[16] D. Ellenberger, R. A. Otten, B. Li, V. Rodriguez, C. A. Sariol, M. Martinez, M. Monsour, L. Wyatt, M. G. Hudgens, E. Kraiselburd, B. Moss, H. Robinson, T. Folks, and S. Butera. HIV-1 DNA/MVA vaccination reduces the per exposure probability of infection during repeated mucosal SHIV challenges. *Virlogy*, 352:216–225, 2006.

[17] L. Emrich and M. Piedmonte. A method for generating high-dimensional multivariate binary variates. *American Statistician*, 45:302–304, 1991.

[18] A. Fauci, M. Johnston, C. Dieffenbach, D. Burton, S. Hammer, J. Hoxie, M. Martin, J. Overbaugh, D. Watkins, A. Mahmoud, and W. Greene. HIV vaccine research: the way forward. *Science*, 321:530–532, 2008.

[19] M. B. Feinberg and J. P. Moore. AIDS vaccine models: challenging challenge viruses. *Nature Medicine*, 8:207–210, Mar 2002.

[20] D. Follmann. On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial. *Journal of the American Statistical Association*, pages 1101–1109, 2000.

[21] D. Follmann. Augmented designs to assess immune response in vaccine trials. *Biometrics*, 62:1161–1169, 2006.

[22] C. E. Frangakis and D. B. Rubin. Principal stratification in causal inference. *Biometrics*, 58:21–29, 2002.

[23] J. García-Lerma, R. Otten, S. Qari, E. Jackson, M. Cong, S. Masciotra, W. Luo, C. Kim, D. Adams, M. Monsour, J. Lipscomb, J. Johnson, D. Delinsky, R. Schinazi, R. Janssen, T. Folks, and W. Heneine. Prevention of rectal SHIV transmission in macaques by daily or intermittent prophylaxis with emtricitabine and tenofovir. *PLoS Med.*, 5:e28, 2008.

[24] P. Gilbert, M. Hudgens, and J. Wolfson. Commentary on "principal stratificationa goal or a tool?" by Judea Pearl. *The International Journal of Biostatistics*, 7(1):36, 2011.

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

[26] P. B. Gilbert and M. G. Hudgens. Evaluating candidate principal surrogate endpoints. *Biometrics*, 64:1146–1154, 2008.

[27] P. B. Gilbert, M. L. Peterson, D. Follmann, M. G. Hudgens, D. P. Francis, M. Gurwith, W. L. Heyward, D. V. Jobes, V. Popovic, S. G. Self, F. Sinangil, D. Burke, and P. W. Berman. Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. *Journal of Infectious Diseases*, 191:666–77, 2005.

[28] P. B. Gilbert, L. Qin, and S. G. Self. Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Statistics in Medicine*, 27:4758–4778, 2008.

[29] M. Girard, S. Osmanov, O. Assossou, and M. Kieny. Human immunodeficiency virus (HIV) immunopathogenesis and vaccine development: A review. *Vaccine*, 29:6191–6218, 2011.

[30] R. H. Gray, M. J. Wawer, R. Brookmeyer, N. K. Sewankambo, D. Serwadda, F. Wabwire-Mangen, T. Lutalo, X. Li, T. vanCott, and T. C. Quinn and the Rakai Project Team. Probability of HIV-1 transmission per coital act in monogamous, heterosexual HIV-1-discordant couples in Rakai, Uganda. *Lancet*, 357:1149–1153, 2001.

[31] S. Greenland and B. Brumback. An overview of relations among causal modelling methods. *International Journal of Epidemiology*, 31(5):1030, 2002.

[32] L. Grilli and F. Mealli. Nonparametric bounds on the causal effect of university studies on job opportunities using principal stratification. *Journal of Educational and Behavioral Statistics*, 33(1):111, 2008.

[33] P. Groeneboom, M. H. Maathuis, and J. A. Wellner. Current status data with competing risks: Consistency and rates of convergence of the MLE. *Annals of Statistics*, 36:1031–1063, 2008.

[34] P. Groeneboom, M. H. Maathuis, and J. A. Wellner. Current status data with competing risks: Limiting distribution of the MLE. *Annals of Statistics*, 36:1064–1089, 2008.

[35] M. Halloran, I. Longini, and C. Struchiner. *Design and Analysis of Vaccine Studies*. Springer Verlag, 2009.

[36] M. E. Halloran, I. M. Longini, and C. J. Struchiner. Design and interpretation of vaccine field studies. *Epidemiological Reviews*, 21:73–88, 1999.

[37] D. Hayden, D. K. Pauler, and D. Schoenfeld. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics*, 61:305–310, 2005.

[38] J. Hill, J. Waldfogel, and J. Brooks-Gunn. Differential effects of high-quality child care. *Journal of Policy Analysis and Management*, 21(4):601–627, 2002.

[39] P. W. Holland. Statistics and causal inference (C/R: p961-970). *Journal of the American Statistical Association*, 81:945–960, 1986.

[40] Y. Huang and P. B. Gilbert. Comparing Biomarkers as Principal Surrogate Endpoints. *Biometrics*, 67(4):1442–1451, 2011.

[41] M. Hudgens, P. Gilbert, J. Mascola, C. Wu, D. Barouch, and S. Self. Power to detect hiv vaccine effects in repeated low-dose challenge experiments. *The Journal of Infectious Diseases*, 200(4):609, 2009.

[42] M. G. Hudgens and P. B. Gilbert. Assessing vaccine effects in repeated low-dose challenge experiments. *Biometrics*, 65:1223–1232, 2009.

[43] M. G. Hudgens and M. E. Halloran. Causal vaccine effects on binary post-infection outcomes. *Journal of the American Statistical Association*, 101:51–64, 2006.

[44] M. G. Hudgens, A. Hoering, and S. G. Self. On the analysis of viral load endpoints in HIV vaccine trials. *Statistics in Medicine*, 22:2281–2298, 2003.

[45] M. G. Hudgens, G. A. Satten, and I. M. Longini. Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics*, 57:74–80, 2001.

[46] Y. Jemiai, A. Rotnitzky, B. E. Shepherd, and P. E. Gilbert. Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a post-randomization event occurs. *Journal of the Royal Statistical Society, Series B*, 69:879–901, 2007.

[47] J. Jeong and J. Fine. Direct parametric inference for the cumulative incidence function. *Journal of the Royal Statistical Society, Series C-Applied Statistics*, 55(2):187–200, 2006.

[48] J. Jeong and J. Fine. Parametric regression on cumulative incidence function. *Biostatistics*, 8(2):184–196, 2007.

[49] B. Jo and E. Stuart. On the use of propensity scores in principal causal effect estimation. *Statistics in Medicine*, 28(23):2857–2875, 2009.

[50] M. Joffe and T. Greene. Related causal frameworks for surrogate outcomes. *Biometrics*, 65(2):530–538, 2009.

[51] KiBS Study Team. Triple-antiretroviral prophylaxis to prevent mother-to-child HIV transmission through breastfeeding - The Kisumu Breastfeeding Study, Kenya: a clinical trial. *PLoS Medicine*, 8(3):e1001015, 2011.

[52] C. Kilewo, K. Karlsson, M. Ngarina, A. Massawe, E. Lyamuya, A. Swai, R. Lipyoga, F. Mhalu, G. Biberfeld, and Mitra Plus Study Team. Prevention of mother-to-child transmission of HIV-1 through breastfeeding by treating mothers with triple antiretroviral therapy in Dar es Salaam, Tanzania: The Mitra Plus Study. *Journal of Acquired Immune Deficiency Syndromes*, 52(3):406–416, 2009.

[53] C. Kim, D. Adams, S. Bashirian, S. Butera, T. Folks, and R. Otten. Repetitive exposures with simian/human immunodeficiency viruses: strategy to study HIV pre-clinical interventions in non-human primates. *Journal of Medical Primatology*, 35(4-5):210–216, 2006.

[54] W. Koff, P. Johnson, D. Watkins, D. Burton, J. Lifson, K. Hasenkrug, A. McDermott, A. Schultz, T. Zamb, R. Boyle, et al. HIV vaccine design: insights from live attenuated SIV vaccines. *Nature Immunology*, 7(1):19–23, 2005.

[55] L. Kuhn, G. Aldrovandi, M. Sinkala, C. Kankasa, K. Semrau, M. Mwiya, P. Kasonde, N. Scott, C. Vwalika, J. Walter, M. Bulterys, W. Tsai, D. Thea, E. Abrams, T. Colton, W. Fawzi, S. Kapiga, E. Chomba, S. Allen, C. Luo, L. Mofenson, E. Piwoz, K. Ryan, J. Simon, Z. Stein, J. Stringer, and S. Vermund. Effects of early, abrupt weaning on HIV-free survival of children in Zambia. *New England Journal of Medicine*, 359:130–141, 2008.

[56] N. Kumwenda, D. Hoover, L. Mofenson, M. Thigpen, G. Kafulafula, Q. Li, L. Mipando, K. Nkanaunena, T. Mebrahtu, M. Bulterys, M. Fowler, and T. Taha. Extended antiretroviral prophylaxis to reduce breast-milk HIV-1 transmission. *New England Journal of Medicine*, 359(2):119–129, 2008.

[57] J. F. Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, 2003.

[58] K. Y. Liang and S. G. Self. On the asymptotic behaviour of the pseudolikelihood ratio test statistic. *Journal of the Royal Statistical Society, Series B: Methodological*, 58:785–796, 1996.

[59] L. Mofenson. Prevention of breast milk transmission of HIV: the time is now. *Journal of Acquired Immune Deficiency Syndromes*, 52(3):305, 2009.

[60] C. Morgan, M. Marthas, C. Miller, A. Duerr, C. Cheng-Mayer, R. Desrosiers, J. Flores, N. Haigwood, S. Hu, R. Johnson, et al. The use of nonhuman primate models in hiv vaccine development. *PLoS medicine*, 5(8):1200–1204, 2008.

[61] B. Nath, K. Schumann, and J. Boyer. The chimpanzee and other non-human-primate models in HIV-1 vaccine research. *Trends in Microbiology*, 8(9):426–431, 2000.

[62] J. Neyman, D. M. Dabrowska, and T. P. Speed. On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (Pkg: P463-480). *Statistical Science*, 5:465–472, 1990.

[63] T. Nolen and M. Hudgens. Randomization-based inference within principal strata. *Journal of the American Statistical Association*, 106(494):581–593, 2011.

[64] R. Otten, D. Adams, C. Kim, E. Jackson, J. Pullium, K. Lee, L. Grohskopf, M. Monsour, S. Butera, and T. Folks. Multiple vaginal exposures to low doses of R5 simian-human immunodeficiency virus: strategy to study HIV preclinical interventions in nonhuman primates. *Journal of Infectious Diseases*, 191:164–173, 2005.

[65] J. Pearl. Principal stratification − a goal or a tool? *The International Journal of Biostatistics*, 7(1):20, 2011.

[66] J. Pearl and E. Bareinboim. Transportability of causal and statistical relations: A formal approach. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 540–547. IEEE, 2011.

[67] M. S. Pepe and T. R. Fleming. A nonparametric method for dealing with mismeasured covariate data. *Journal of the American Statistical Association*, 86:108–113, 1991.

[68] M. S. Pepe and M. Mori. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine*, 12:737–751, 1993.

[69] R. L. Prentice. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8:431–440, 1989.

[70] L. Qin, P. Gilbert, L. Corey, M. McElrath, and S. Self. A framework for assessing immunological correlates of protection in vaccine trials. *Journal of Infectious Diseases*, 196:1304–1312, 2007.

[71] R. R. Regoes, I. M. Longini, M. B. Feinberg, and S. I. Staprans. Preclinical assessment of HIV vaccines and microbicides by repeated low-dose virus challenges. *PLoS Medicine*, 2(8):e249, 2005.

[72] J. M. Robins, A. Rotnitzky, and D. O. Scharfstein. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In M. Halloran and D. Berry, editors, *Statistics in Epidemiology, Environment and Clinical Trials*. Springer-Verlag, 2000.

[73] K. Rothman. Causes. *American Journal of Epidemiology*, 104(6):587, 1976.

[74] D. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

[75] D. B. Rubin. Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by D. Basu. *Journal of the American Statistical Association*, 75:591–593, 1980.

[76] J. Sadoff and J. Wittes. Correlates, surrogates, and vaccines. *Journal of Infectious Diseases*, 196(9):1279, 2007.

[77] D. O. Scharfstein, M. E. Halloran, H. Chu, and M. J. Daniels. On estimation of vaccine efficacy using validation samples with selection bias. *Biostatistics*, 7:615–629, 2006.

[78] D. Shedlock, G. Silvestri, and D. Weiner. Monkeying around with HIV vaccines: using rhesus macaques to define 'gatekeepers' for clinical trials. *Nature Reviews Immunology*, 9(10):717–728, 2009.

[79] B. Shepherd, P. Gilbert, and C. Dupont. Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected postrandomization and relaxing monotonicity. *Biometrics*, 67(3):1100–1110, 2011.

[80] B. Shepherd, P. B. Gilbert, Y. Jemiai, and A. Rotnitzky. Sensitivity analyses comparing outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials. *Biometrics*, 62(2):332–342, 2006.

[81] B. E. Shepherd, P. B. Gilbert, and T. Lumley. Sensitivity analyses comparing time-to-event outcomes existing only in a subset selected postrandomization. *Journal of the American Statistical Association*, 102(478):573–582, 2007.

[82] B. E. Shepherd, P. B. Gilbert, and D. V. Mehrotra. Eliciting a counterfactual sensitivity parameter. *The American Statistician*, 61(1):56–63, 2007.

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

[84] S. Smith. HIV vaccine development in the nonhuman primate model of aids. *Journal of Biomedical Science*, 9(2):100–111, 2002.

[85] S. Subbarao, R. Otten, A. Ramos, E. Jackson, M. Monsour, S. Bashirian, C. Kim, J. Johnson, V. Soriano, M. G. Hudgens, S. Butera, R. Janssen, L. Paxton, A. Greenberg, and T. Folks. Chemoprophylaxis with Tenofovir Disoproxil Fumarate provided partial protection against Simian Human Immunodeficiency Virus infection in macaques given multiple virus challenges. *Journal of Infectious Diseases*, 194:904–11, 2006.

[86] A. A. Tsiatis. Competing risks. In P. Armitage and T. Colton, editors, *Encyclopedia of Biostatistics*, New York, 1998. Wiley. 824-834.

[87] UNAIDS. AIDS epidemic update. Joint United Nations Programme on HIV/AIDS. Geneva, December 2007.

[88] C. van der Horst, C. Chasela, Y. Ahmed, I. Hoffman, M. Hosseinipour, R. Knight, S. Fiscus, M. Hudgens, P. Kazembe, M. Bentley, L. Adair, E. Piwoz, F. Martinson, A. Duerr, A. Kourtis, A. Loeliger, B. Tohill, S. Ellington, and D. Jamieson. Modifications of a large HIV prevention clinical trial to fit changing realities: A case study of the Breastfeeding, Antiretroviral, and Nutrition (BAN) protocol in Lilongwe, Malawi. *Contemporary Clinical Trials*, 30:24–33, 2009.

[89] T. VanderWeele. Principal stratification–uses and limitations. *The International Journal of Biostatistics*, 7(1):1–14, 2011.

[90] T. J. VanderWeele and M. A. Hernan. Causal inference under multiple versions of treatment. *COBRA Preprint Series*, 77, 2011, http://biostats.bepress.com/cobra/ps/art77/.

[91] S. Vansteelandt, E. Goetghebeur, M. Kenward, and G. Molenberghs. Ignorance and uncertainty regions as inferential tools in a sensitivity analysis. *Statistica Sinica*, 16(3):953–979, 2006.

[92] World Health Organization. HIV transmission through breastfeeding: a review of available evidence. Geneva, 2007.

[93] J. Zhang, D. Rubin, and F. Mealli. Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, 104(485):166–176, 2009.

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