

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

---

Year 2007

Paper 225

---

Detailed Version: Analyzing Direct Effects in  
Randomized Trials with Secondary  
Interventions: An Application to HIV  
Prevention Trials

Michael A. Rosenblum\*

Nicholas P. Jewell<sup>†</sup>

Mark J. van der Laan<sup>‡</sup>

Stephen Shiboski\*\*

Ariane van der Straten<sup>††</sup>

Nancy Padian<sup>‡‡</sup>

\*University of California, Berkeley, [mrosenbl@jhsph.edu](mailto:mrosenbl@jhsph.edu)

<sup>†</sup>Division of Biostatistics, School of Public Health, University of California, Berkeley, [jewell@berkeley.edu](mailto:jewell@berkeley.edu)

<sup>‡</sup>Division of Biostatistics, School of Public Health, University of California, Berkeley, [laan@berkeley.edu](mailto:laan@berkeley.edu)

\*\*Department of Epidemiology and Biostatistics, University of California, San Francisco, [steve@biostat.ucsf.edu](mailto:steve@biostat.ucsf.edu)

<sup>††</sup>Women's Global Health Imperative, Department of Obstetrics, Gynecology, and Reproductive Services, University California, San Francisco, [AVDStraten@globalhealth.ucsf.edu](mailto:AVDStraten@globalhealth.ucsf.edu)

<sup>‡‡</sup>Women's Global Health Imperative, Department of Obstetrics, Gynecology, and Reproductive Services, University of California, San Francisco, [npadian@globalhealth.ucsf.edu](mailto:npadian@globalhealth.ucsf.edu)

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<http://biostats.bepress.com/ucbbiostat/paper225>

Copyright ©2007 by the authors.

# Detailed Version: Analyzing Direct Effects in Randomized Trials with Secondary Interventions: An Application to HIV Prevention Trials

Michael A. Rosenblum, Nicholas P. Jewell, Mark J. van der Laan, Stephen  
Shiboski, Ariane van der Straten, and Nancy Padian

## Abstract

This is the detailed technical report that accompanies the paper “Analyzing Direct Effects in Randomized Trials with Secondary Interventions: An Application to HIV Prevention Trials” (an unpublished, technical report version of which is available online at <http://www.bepress.com/ucbbiostat/paper223>).

The version here gives full details of the models for the time-dependent analysis, and presents further results in the data analysis section. The Methods for Improving Reproductive Health in Africa (MIRA) trial is a recently completed randomized trial that investigated the effect of diaphragm and lubricant gel use in reducing HIV infection among susceptible women. 5,045 women were randomly assigned to either the active treatment arm or not. Additionally, all subjects in both arms received intensive condom counselling and provision, the “gold standard” HIV prevention barrier method.

There was much lower reported condom use in the intervention arm than in the control arm, making it difficult to answer important public health questions based solely on the intention-to-treat analysis.

We adapt an analysis technique from causal inference to estimate the “direct effects” of assignment to the diaphragm arm, adjusting for condom use in an appropriate sense.

Issues raised in the MIRA trial apply to other trials of HIV prevention methods, some of which are currently being conducted or designed.

## 1. Introduction

This is the detailed technical report that accompanies the paper “Analyzing Direct Effects in Randomized Trials with Secondary Interventions: An Application to HIV Prevention Trials” (an unpublished, technical report version of which is available online—see Rosenblum et al. (2007) in the bibliography for the web-address). The version here gives full details of the models for the time-dependent analysis, and presents further results in the data analysis section.

Randomized controlled trials play a critical role in the assessment of public health interventions to prevent and treat disease. Standard intention-to-treat analyses, which are straightforward comparisons of aggregate outcomes in the treatment and control groups, are widely assumed to be the most appropriate method for analyzing and reporting trial results. However, in intervention trials, ethical considerations generally require provision of currently accepted disease prevention methods to all participants as a secondary intervention, irrespective of study arm assignment. For example, in randomized controlled trials of new HIV prevention methods, provision of condoms with counselling is required for all study participants to meet human subjects protection requirements. In such cases, the standard intention-to-treat analysis may not adequately answer important questions of public health interest. In particular, in these cases it may not address the effectiveness of the new prevention method in the absence of the secondary intervention. Also, it may give severely biased estimates of study product efficacy, even in blinded trials with relatively high rates of compliance to the primary intervention (Trussell and Dominik, 2005). The situation is even worse in unblinded trials of new HIV prevention methods; for example, in the MIRA trial of latex diaphragms and gel, the frequency of condom use was different in the treatment and control arms, making it quite difficult to draw conclusions about diaphragm and gel efficacy based only on the intention-to-treat analysis (Padian et al., 2007). In this paper, we highlight limitations in relying solely on standard analyses of randomized trials with secondary interventions such as condom counselling, and we propose a supplemental analysis technique. Our intent is to stimulate debate on additional design and analysis approaches. Improved methods tailored to HIV prevention trials are urgently needed, since many of these trials are currently being conducted or planned. We focus on supplemental analysis approaches throughout this paper, but briefly consider design issues in the discussion section.

We describe a causal inference technique, called a direct effects analysis, that attempts to isolate the effects of the secondary intervention, thereby providing more information about primary treatment efficacy. This technique can be applied in situations where a secondary intervention affects the primary outcome only through a measured covariate or set of covariates. For example, in HIV prevention trials, it is thought that condom provision and counselling affect HIV risk primarily through their impact on condom use, which is often measured in these trials. Unlike standard time-dependent regression techniques to adjust for time-dependent confounders, the direct effects analysis adequately handles time-dependent confounders measured in a trial. We illustrate the benefits and limitations of our direct

effects analysis by applying it to the trial entitled “Methods for Improving Reproductive Health in Africa” (MIRA) (Padian et al., 2007), a randomized controlled trial investigating the effect of providing diaphragm and non-contraceptive, lubricant gel (Replens), on male-to-female HIV transmission.

For simplicity, we consider randomized controlled trials with two arms, in which a primary intervention is provided only to the treatment arm, and a secondary intervention (such as condom provision and counselling) is given to both treatment and control arms. This design is currently used in virtually all trials that evaluate new HIV prevention methods. Examples include trials evaluating the effects of microbicides (Mantell et al., 2005), male circumcision (Auvert et al., 2005; Bailey et al., 2007; Gray et al., 2007), and woman-controlled barrier methods (Padian et al., 2007). In such trials, one risks overlooking and perhaps dismissing effective new HIV prevention methods when interpretations of trial results are based solely on a standard intention-to-treat analysis. It is also possible that an intention-to-treat analysis would overstate the benefits of an intervention, were it deployed in a different setting than that of the study. Direct effects analyses can provide useful, supplemental information to help identify whether a secondary intervention may be having these effects.

We focus primarily on unblinded trials (where participants know which study arm they are in), since here the problems in interpreting the intention-to-treat estimator can be worse than in blinded trials. Unblinding may be a necessary part of a trial design, for example in trials of barrier methods or male circumcision where no placebo is used; some unintended unblinding may also occur in trials with placebo (Friedman et al., 1998). Though we concentrate on unblinded trials with secondary interventions, such as the MIRA trial, we also discuss analogous issues in blinded trials. In particular, we briefly describe problems in interpreting the intention-to-treat analysis in blinded trials, and also explain how a direct effects analysis can be applied to obtain useful information in blinded trials in the discussion section of the paper.

Unblinding leaves open the possibility that differences in average outcomes between the two study arms could be due to factors (such as behaviour change in response to treatment assignment) other than efficacy of the treatment being studied. Also, with unblinding, average outcomes that are equal in the two study arms do not rule out the treatment being efficacious. For example, in the MIRA trial, which is unblinded, the fact that the HIV infection rates in the two study arms were nearly identical does not necessarily mean that latex diaphragms offer no protection against HIV; an alternative explanation, whose plausibility we explore in this paper, is that being assigned to the diaphragm arm resulted in significantly less condom use, which may have cancelled out a protective effect from diaphragms. A direct effects analysis aims to isolate the effect of a suspected alternative causal pathway connecting treatment assignment to outcome, in order to help assess whether such alternative explanations are plausible. Our approach is based on a method of direct effects analysis from Robins and Greenland (1992) and Pearl (2000b). This method is also discussed in Petersen et al. (2006), under a different set of assumptions.

Our paper is organized as follows: We describe the MIRA trial in Section 2. In Section 3 we discuss the limitations of intention-to-treat analyses in the presence of secondary interventions or unblinding. In Section 4 we introduce direct effects, and in Sections 5 and 6 we present estimators of direct effects, and the assumptions needed for these estimators to be valid. In Section 7, we provide summary data from the MIRA trial, and then examine the results of standard intention-to-treat analyses as well as direct effects analyses. Lastly, in Section 8 we interpret the results of the direct effects analysis, discuss its limitations as

well as other settings where it can be applied, and then consider alternative analysis and design approaches.

## 2. The MIRA Trial

We briefly describe the MIRA trial, used throughout the paper to illustrate a direct effects analysis and its interpretation. The MIRA trial is a randomized controlled trial with two arms. In the “diaphragm arm,” each subject was instructed to use a fitted Ortho All Flex diaphragm, Replens non-contraceptive, lubricant gel, and condoms at all sex acts, and were supplied with these products. In the “control arm,” each subject was instructed to use condoms at all sex acts, and was supplied with condoms. Additionally, in both arms, subjects were given intensive condom counselling, as well as diagnosis and treatment of sexually transmitted infections. The eligibility criteria for enrolment in the study included being a woman 18-49 years old, HIV-negative at baseline, sexually active, and non-pregnant. 5,045 women were enrolled at three sites in South Africa and Zimbabwe from September 2003 until September 2005. Participants were assessed at quarterly clinic visits with a goal of following each participant for two years. The median follow-up period was 21 months.

Before we looked at any HIV outcome data, we chose as measure of a subject’s condom use over the period since her last visit to be the participant’s response to the question: “Did you use a condom at your last sexual intercourse?” This choice was based in part on a growing literature regarding accurate measures of condom use based on self-reporting questionnaires (Anderson et al., 1998). Reported condom use at last sex, recorded at each follow-up visit, averaged 53.5% across all visits in the diaphragm arm and 85.1% in the control arm. (Note that other data show an increase in overall condom uptake after recruitment in both arms—see Padian et al. (2007)).

Diaphragm use, based on self-reports of use at last sex at each clinic visit, averaged 73% in the diaphragm arm. Diaphragm use at last sex was reported at only 0.15% of visits by participants in the control arm. For a detailed description of subject recruitment and retention, eligibility, survey instruments and data collection, see Padian et al. (2007).

A key public health question is how much, if any, protection diaphragms provide against male-to-female HIV infection in a community—as compared to a research—setting. Another important question is how much protection consistent diaphragm and gel use (but no condom use) provide, compared to unprotected sex. As further discussed in Section 3, neither of these public health questions is answered adequately by a standard intention-to-treat comparison of HIV infection rates between study arms.

## 3. Intention-To-Treat Analysis

The simplest intention-to-treat analysis of a randomized controlled trial compares the average outcome for the treatment group to that for the control group. For example, in the MIRA trial, this would be a comparison of HIV incidence between the diaphragm arm and control arm. Such an analysis may also include this comparison stratified by pre-specified covariates measured before randomization. The advantages of an intention-to-treat analysis in many types of randomized controlled trials include:

- (a) An intention-to-treat analysis is completely protected from confounding, since it does not involve any measurements made after randomization (e.g. adherence to treatment) except for the outcomes of interest.

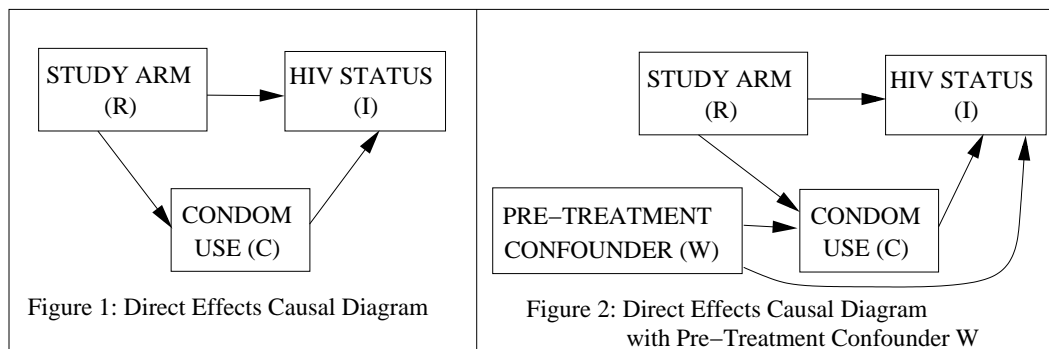
- (b) An intention-to-treat analysis unequivocally answers a specific causal question via its simplest interpretation: specifically, it estimates the effect of the primary intervention in the study setting. We refer to the relative risk of infection between treatment and control arms in a particular setting (e.g. in a community setting, in a study setting) as the treatment “effectiveness.”
- (c) An intention-to-treat analysis gives a conservative approximation of treatment efficacy. Treatment efficacy, the effect one would observe if treatment were given in an ideal setting where full adherence were guaranteed, is likely to be greater than the effect measured under non-ideal conditions of the trial where there is non-compliance.

Before presenting limitations of the usefulness of the intention-to-treat estimator in the setting of trials with secondary interventions, we briefly review the limitations with relying solely on this estimator that have been pointed out in other settings. Sheiner et al. (1995) point out problems in interpreting the intention-to-treat analysis under non-compliance to assigned treatment, and argue that further analyses are needed to understand study product effectiveness and efficacy. Frangakis and Rubin (1999) show that the intention-to-treat estimator can be a biased estimator of product effectiveness in the presence of non-compliance followed by missing outcomes. Hirano et al. (2000) present an example of an unblinded randomized trial with a strongly positive intention-to-treat result, but argue based on further analyses that this result may be more likely due to the encouragement to take the treatment than to the treatment itself.

In prevention trials with a secondary intervention or unblinding, the advantages (b) and (c) listed above may be of little value in answering the most important public health questions related to the primary intervention. First, the simple intention-to-treat comparison may differ substantially from the same comparison in other populations whose members are both subject to different information on the possible efficacy of diaphragms, and who are not provided with free condoms coupled with intensive counselling; this type of limitation of the intention-to-treat estimator has been pointed out by Feinstein (1985), Sheiner et al. (1995) and Friedman et al. (1998). Also, while it is well known that in blinded trials, standard intention-to-treat analyses may underestimate product efficacy (Feinstein, 1985; Sheiner et al., 1995; Friedman et al., 1998), these analyses may be even less informative in the presence of a secondary intervention such as condom provision and counselling (Trussell and Dominik, 2005). This can occur, for example, when the secondary intervention differentially affects adherers and non-adherers to the primary study product. See the Appendix for an example from Trussell and Dominik (2005), except modified so that compliance is roughly at the level observed in the MIRA trial, illustrating this phenomenon in a blinded trial. In short, the question definitively addressed by the standard intention-to-treat analysis may not be of major public health interest.

Second, in unblinded trials, the intention-to-treat analysis may be severely impacted when the primary intervention results in unintended behaviour changes. For example, in the MIRA trial, those randomized to the diaphragm arm reported much less condom use, as compared to the control arm. Without considering the impact of this differential condom use, an intention-to-treat result of no difference in HIV infections between study arms would be consistent with both (i) the diaphragm being not efficacious, but also (ii) the diaphragm providing some protection that may have been cancelled out by additional infections due to decreased condom use in the diaphragm arm. Unintended effects of study arm assignment are illustrated in the causal diagram of Figure 1 which shows schematically how the overall causal effect of study arm assignment can be viewed as having two distinct

causal pathways: (i) a direct effect of treatment on HIV infection (the top arrow), and (ii) an indirect effect through condom use that, in turn, affects the risk of infection (the indirect effect of treatment assignment). Causal diagrams represent sets of assumed causal relationships: for an introduction to causal diagrams and their interpretation see Pearl (2000a) or Chapter 8.2 of Jewell (2004).



Figures 1 and 2

#### 4. Direct Effects Analysis: Overview, Definitions and Causal Interpretation

Given the large difference in reported condom use between the two study arms in the MIRA trial (53.5% in the diaphragm arm vs. 85.1% in the control arm), we would like to better understand the role of condom use in mediating the effect of treatment assignment on HIV outcome. For the MIRA trial, we decided to focus on condom use as a mediator for two reasons. First, the condom provision and intensive counselling intervention was believed primarily to affect condom use, and not to affect other HIV risk factors, and it is of major public health interest what the effect of diaphragms would be in the absence of condom counselling, where condom use is generally quite low. Second, except for diaphragm use, condom use was the only time-dependent variable measured in the MIRA trial that was meaningfully different in the two study arms. Variables such as number of other partners and frequency of sex were measured but each showed less than a 5% difference between the two arms.

##### 4.1. Direct Effects Defined through Hypothetical Randomized Trials

Ideally, we would like to know what the infection rates would be in a hypothetical randomized trial in which the diaphragm intervention were given to one arm, and in which all participants were constrained never to use condoms. Such a trial is highly unethical so cannot be conducted. In our direct effects analysis, we attempt to approximate the results of such a trial, using data from the MIRA study. If we are successful, this would shed light on the efficacy of diaphragms. For example, if we conclude that in this hypothetical trial there would be no benefit of the diaphragm intervention, this would be strong evidence that the diaphragm was not efficacious in preventing HIV infections (unless actual diaphragm use is much lower than what subjects reported). Alternatively, if we would conclude that in such a trial there would be a strong benefit of the diaphragm intervention, then this would be evidence in favour of diaphragm efficacy.



We describe several hypothetical scenarios below, in which we talk about how many new HIV infections would have occurred had all study subjects been constrained to either never use condoms or always use condoms. It has been argued that in order for such hypothetical scenarios to be well defined, it must be possible, at least in principle, to imagine interventions that would lead to such constraints being enforced (Frangakis and Rubin, 2002). We think the constraint on never using condoms could be approximately established by a hypothetical (unethical) intervention. In general, without condom counselling and provision, condom use is quite low and it is possible to imagine an intervention (that of course should never be done) discouraging condom use that would reduce this further and not much affect HIV risk factors in any other way. Similarly, in theory, an intervention is possible that would result in condoms almost always being used. We argue that our estimator based on data collected in the MIRA trial gives consistent, asymptotically normal estimates of the infection rates in these hypothetical scenarios, under a set of strong assumptions described in Sections 5 and 6 below. The validity of these assumptions is discussed in detail in Section 8; there, we describe significant sources of bias in our estimator, including reporting bias in condom use and other data, and unmeasured confounders such as male partner characteristics.

Consider what the infection rate would be in the following two hypothetical scenarios: (1) all subjects in the MIRA trial are assigned to the diaphragm arm and are constrained to use condoms every time they have sex; (2) all subjects in the MIRA trial are assigned to the control arm and are constrained to use condoms every time they have sex. We call the ratio of infection rates in the two scenarios the “direct effects relative risk” of treatment assignment with condom use fixed at “always.” The ratio of infection rates in the same pair of hypothetical scenarios, except with all subjects constrained to never use condoms, gives the direct effects relative risk with condom use fixed at “never.” This definition of direct effects relative risk corresponds to Type I (“controlled”) direct effects in Petersen et al. (2006). The main challenge in a direct effects analysis is to estimate these quantities using data collected from a randomized trial in which subjects themselves (or their partners) decided when to use condoms, and where condom use is not objectively measured but is self-reported.

#### 4.2. Direct Effects in terms of Potential Outcomes

The above definitions can be written in terms of potential outcomes. The potential outcomes viewpoint has a long history related to work by Fisher and Neyman (Angrist et al., 1996), and later extended by Rubin (Rubin, 1974) and Robins (Robins, 1986), though it is not universally accepted (see e.g. Dawid (2000)). In this setting, a potential outcome  $I_{rc}$  is defined for each subject in the population under each of the four possible combinations of  $r \in \{0, 1\}$ ,  $c \in \{0, 1\}$ . Let  $r = 0$  indicate the control arm, and  $r = 1$  indicate the diaphragm arm; similarly,  $c = 0$  indicates never using condoms, and  $c = 1$  indicates always using condoms. For a given subject,  $I_{rc}$  represents whether that subject would have become infected with HIV had she been assigned to study arm  $r$  and been constrained to use condoms at level  $c$ .  $I_{rc}$  can either be 0 indicating no HIV infection, or 1 indicating HIV infection. The direct effects analysis involves estimating the probability of HIV infection,  $p_{rc} := \Pr(I_{rc} = 1)$  for a randomly chosen subject from the MIRA study, would she have been assigned to arm  $r$  and simultaneously constrained to use condoms at fixed level  $c$ , for  $r \in \{0, 1\}$  and  $c \in \{0, 1\}$ . These probabilities can then be combined to make appropriate causal comparisons; for example, we can compute the ratio  $p_{1c}/p_{0c}$  to get the direct effects relative risk of treatment for condom use fixed at level  $c$ .

We now describe how the potential outcomes  $\{I_{rc}\}$  connect to the observed data in the MIRA trial. For each study subject, the study arm assignment she received in the trial is denoted by  $R$ , with  $R = 0$  indicating the control arm and  $R = 1$  indicating the diaphragm arm. For the time being we assume that use of condoms is simply recorded as  $C = 1$  for use and  $C = 0$  for non-use. We discuss more precisely how measurement of  $C$  was done in the MIRA trial, and in particular how it was measured over time, in Section 6. Let  $I$  indicate HIV status at the end of the trial, with  $I = 0$  indicating not infected, and  $I = 1$  indicating infected. We emphasize the distinction between potential (generally unobserved) outcomes  $I_{rc}$  and the observed quantities  $(R, C, I)$  for each subject. The “consistency assumption” connects the observed data (including  $R, C, I$ ) to the set of potential outcomes  $\{I_{rc}\}$ . It states that the observed infection status  $I$  is equal to the potential outcome  $I_{rc}$  when observed study arm assignment  $R = r$  and observed condom use  $C = c$ . We consider the other, unrealized potential outcomes for each subject as missing data, and apply causal inference methodology of van der Laan and Robins (2003) accordingly.

It is important to note the asymmetry here between the two variables  $R$  and  $C$  in that the former refers to assignment to a study arm (but not necessarily use of the study product) and the latter refers to actual use of condoms. Thus, direct effects relative risks do not measure, for example, effectiveness of the diaphragm intervention in a setting in which condom use is merely encouraged; they measure effectiveness of the diaphragm intervention in a setting in which condom use is constrained at a fixed level.

Direct effects analyses provide valuable information for understanding the efficacy of the primary treatment. Though our direct effects relative risks are not targeted at estimating treatment efficacy, and so give biased estimates of efficacy, they do peel away a layer of distortion that results from using the intention-to-treat estimator as an estimate of efficacy. The direct effects analysis removes distortion caused by both (i) the effect of treatment assignment on HIV infection through changing condom use and (ii) the effect of a secondary treatment on HIV infection, insofar as this is captured through changed condom use. The direct effects estimate may give better (though still biased) estimates of efficacy than an intention-to-treat estimator, when the effects in (i) or (ii) are strong, and the remaining effect of treatment assignment on HIV outcome is due primarily to use of the primary study product. The effects of (i) and (ii) are believed to be strong in the case of the MIRA trial, and all data collected indicate that no other measured HIV risk factors (e.g. number of partners, frequency of sex) were affected by treatment assignment. However, it is quite possible that there are HIV risk factors affected by study arm assignment not captured by the study data; this would result in additional bias when direct effects estimates are used as (conservative) estimates of study product efficacy.

Direct effects analyses may also be useful in formulating prevention messages to the community. If justified by the direct effects analysis of the MIRA trial and well supported by evidence from other trials, it would be important to communicate, for example, that for those using condoms consistently, diaphragms offer limited additional protection against HIV infection, but for those never using condoms, diaphragms have an effectiveness of  $x\%$ . Alternatively, findings may indicate that assignment to diaphragms are similarly beneficial across all levels of condom use (i.e. their benefit is additive to condoms on a multiplicative scale, say). Such a finding would have different implications for the promotion of diaphragm use and integration of their value into public health messages about HIV prevention.

## 5. Estimating Direct Effects in a Single Time Point Setting

We describe how to estimate direct effects in what we call the “single time point setting,” for the MIRA trial. We categorize all participants, based on their self-reported condom use at last sex at each of the first 3 visits. “Infrequent users” ( $C = 0$ ) are those who reported condom use at last sex in at most one out of the first three visits, and “frequent users” ( $C = 1$ ) are those who reported condom use at last sex in all of the first three visits. We let  $I$  represent HIV status at the end of the trial (which consists of a total of eight visits, though some women were not followed-up for the full eight visits). Participants who became infected with HIV in the first three visits were not included in this single time point analysis; even though the number of HIV infections in each arm by the end of the first three visits was nearly identical (84 in the diaphragm arm, 81 in the control arm), excluding these early infections still could potentially lead to bias.

A naive analysis that ignores confounding variables is subject to bias since condom use was not constrained by the experimenters, but depends on the individual circumstances and choices of the trial subjects. In order to obtain consistent, asymptotically normal estimates of the direct effect of assignment to the diaphragm arm from observed data, all confounders of the effect of condom use on HIV status must be measured. For example, a possible confounder is number of male sex partners. This is a risk factor for HIV infection and may influence frequency of condom use (for example, if condoms are rarely used in marital relationships but routinely used with other partners).

Potential confounders measured in the MIRA trial can be categorized as either baseline factors (denoted by  $W$ ) or time-dependent measurements (denoted by  $D$ ). Baseline factors measured included (i) personal characteristics (employment status, living with husband, age, ever pregnant, HSV-2 status); (ii) regular partner’s characteristics (partner ever tested HIV-positive, partner’s age and employment status, whether subject or her partner makes condom use decision); (iii) sexual habits (number of lifetime male partners, frequency of intercourse, other contraceptive use, whether had sex under influence of alcohol in past 3 months, condom use at last sex); and (iv) study site. Time-dependent confounders included diaphragm use at last sex, sexual abstinence since last visit, HSV-2 incidence, and whether sexual partners were known to have multiple partners during the previous three months.

In Sections 5.1 and 5.2, we describe how to estimate the HIV infection probabilities  $p_{rc} = \Pr(I_{rc} = 1)$ , for all four combinations of  $r \in \{0, 1\}$ ,  $c \in \{0, 1\}$ ; these estimates can then be combined to estimate direct effects relative risks. To present the simplest case of a direct effects analysis, in Section 5.1 we describe this analysis assuming all potential confounders are measured before randomization. In Section 5.2, we extend the estimation techniques to cover confounding variables that are measured post-randomization and may lie on a causal pathway between study arm assignment and condom use. In Section 6, we deal with time-dependent measurements, including time-dependent confounding. In all of these sections, we present the inverse probability of treatment weighted (IPTW) estimator for the direct effects of study arm assignment ( $R$ ) on HIV status ( $I$ ), with condom use “fixed” at a particular level. In particular, we describe a set of assumptions that guarantee the IPTW estimator will give a consistent and asymptotically normal estimate of these direct effects, and then show how to compute the IPTW estimator. Consistency implies the estimator converges in probability to the correct value, and asymptotic normality allows for asymptotically valid confidence intervals. The proofs that our assumptions imply consistency and asymptotic normality follow directly from arguments in van der Laan and Robins (2003). For all of the discussion below, we assume that the data observed on each subject can be viewed as

random draw from a hypothetical, infinite population of subjects.

### 5.1. Estimation of Direct Effects When All Confounders Can Be Measured Prior to Randomization

We consider the case in which all confounders of the effect of condom use on HIV status are measured before study arm assignment. Such a setting is depicted in Figure 2 of Section 3 above, in which the variable  $W$  denotes pre-treatment (baseline) confounders. The proposed estimator of  $p_{rc}$  is an example of the class of inverse probability of treatment weighted (IPTW) estimators described in Robins and Finkelstein (2000). The IPTW estimator requires a correctly specified model for the probability of condom use  $C$ , given baseline confounders  $W$  and study arm assignment  $R$ . In Section 6 we explain why we chose to use the IPTW estimator instead of other possible estimators (including G-computation estimators (Robins, 1986), doubly robust estimators (van der Laan and Robins, 2003), and targeted maximum likelihood estimators (van der Laan and Rubin, 2006)).

Before presenting the IPTW estimator, we state assumptions under which it gives consistent, asymptotically normal estimates of the direct effects relative risks. First, we make the following conditional independence assumption, sometimes referred to as the “randomization assumption” (van der Laan and Robins, 2003),

$$C \perp\!\!\!\perp \{I_{rc}\} | W, R. \tag{1}$$

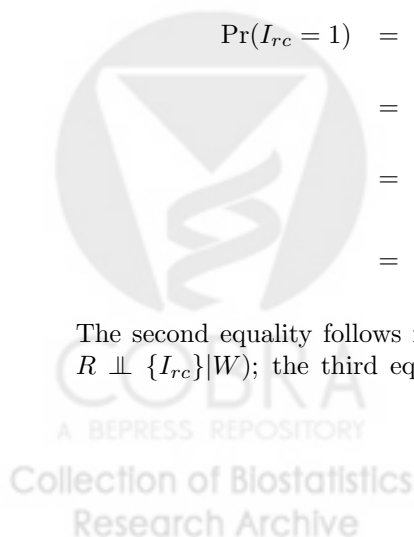
That is, condom use  $C$  is independent of the potential outcomes  $\{I_{rc}\}$ , conditional on baseline confounders  $W$  and study arm assignment  $R$ . Intuitively, this means that, within strata having the same study arm  $R$  and the same baseline measurements  $W$ , the decision to use condoms is not an indicator of underlying HIV risk (e.g. due to having high-risk partners, many partners, or having other STIs). This assumption implies that there are no unmeasured confounders of condom use and HIV infection.

We make the following assumption, called the experimental treatment assignment (ETA) assumption: For all  $r, c, w$ ,  $\Pr(C = c | R = r, W = w) > 0$ . That is, no stratum of treatment assignment and baseline confounders precludes the possibility of using condoms at any specific level. If this assumption is false, direct effects will not be identifiable from the data alone, unless further model assumptions are made. We also make the consistency assumption introduced in Section 4.2.

We can connect the probabilities of specific potential outcomes that we ultimately care about to quantities that can be estimated from the observed data using the following chain of equalities:

$$\begin{aligned} \Pr(I_{rc} = 1) &= \sum_w \Pr(I_{rc} = 1 | W = w) \Pr(W = w) \\ &= \sum_w \Pr(I_{rc} = 1 | W = w, R = r, C = c) \Pr(W = w) \\ &= \sum_w \Pr(I = 1 | W = w, R = r, C = c) \Pr(W = w) \\ &= \sum_w \frac{\Pr(I = 1, R = r, C = c, W = w)}{\Pr(R = r | W = w) \Pr(C = c | W = w, R = r)}. \end{aligned} \tag{2}$$

The second equality follows from Assumption (1) and the fact that  $R$  is randomized (so  $R \perp\!\!\!\perp \{I_{rc}\} | W$ ); the third equality follows from the consistency assumption described in



Section 4 above; and the last equality follows by basic rules of probability. The experimental treatment assignment (ETA) assumption insures that all the terms in (2) are well-defined. Note that once we get estimates for the left hand side of this equation, for all values of  $r, c$ , we can estimate the direct effects relative risks  $\Pr(I_{1c} = 1)/\Pr(I_{0c} = 1)$  for each fixed level of condom use  $c$ . We also point out that estimating  $\Pr(I_{rc} = 1)$  using (2) is equivalent to solving the following estimating equation:

$$E\left(\frac{1}{\Pr(R|W)\Pr(C|W, R)}(I - m(R, C|\beta))\right) = 0, \quad (3)$$

where  $m(r, c|\beta)$  denotes a saturated (non-parametric) model for the unknown probability  $\Pr(I_{rc} = 1)$ .

To estimate (2), we first estimate its numerator using the empirical distribution of the data. The first term in the denominator of (2),  $\Pr(R = r|W = w)$ , equals  $1/2$  since  $R$  was randomized. The second term in the denominator can be estimated by the empirical distribution if  $W$  is low-dimensional, or using a logistic regression model, for example, if  $W$  is high-dimensional. Substituting such estimates for the quantities in the right hand side of (2), and summing over  $w$ , we get what is referred to as the inverse probability of treatment weighted (IPTW) estimator for  $\Pr(I_{rc} = 1)$ . We use the bootstrap, as described in Efron and Tibshirani (1993), to generate confidence intervals for direct effects relative risks based on these estimators. Alternatively, the influence curve for the estimator could be used for statistical inference, as described in van der Laan and Robins (2003).

The argument above assumed that all confounders were measured at baseline, prior to randomization. In the next section, we treat the case in which there are confounders of condom use and HIV status that are themselves affected by study arm assignment.

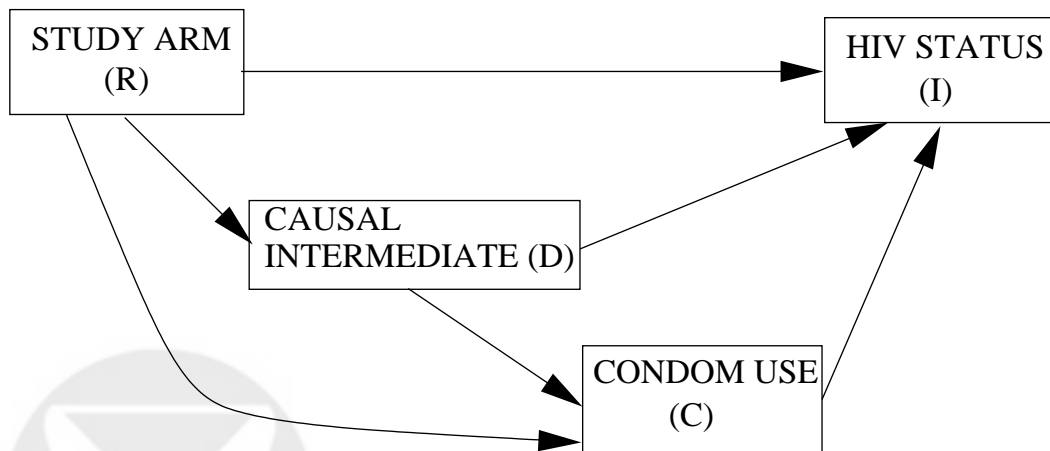


Figure 3: Direct Effects Diagram with Confounder as Causal Intermediate

### 5.2. Estimation of Direct Effects With Confounding by Causal Intermediates

In the MIRA trial, there are potential confounders of condom use and HIV status that are themselves affected by study arm assignment. Such confounders are called “causal

intermediates,” and are denoted by  $D$ . This situation is illustrated in the causal diagram of Figure 3, where for clarity, pre-randomization potential confounders  $W$  are not shown. For an example of a potential confounder that may be a causal intermediate, consider diaphragm use, defined similarly to condom use for the single time point case above—that is, diaphragm use is a dichotomized summary measure, based on self-reports of use at last sex at the initial three visits. Diaphragm use may affect the decision to use condoms and it also may affect HIV infection (by possibly acting as a physical barrier); thus, diaphragm use may be a confounder that is a causal intermediate. In general, the set of confounders that are causal intermediates will be a vector  $D$  including more variables than just diaphragm use (e.g. frequency of sex, number of partners). For concreteness in our explanations, we assume diaphragm use is the only confounder that is a causal intermediate; however in our statistical analyses, we included all the time-dependent variables listed in Section 5. Note that diaphragm use will be “never” ( $D = 0$ ) for virtually all participants in the control arm, except the few who obtained and used diaphragms on their own.

A single regression (e.g. regressing  $I$  on  $R, D, C$ ) cannot be used to estimate the direct effects of  $R$  on  $I$ , fixing  $C$ , in the presence of confounding by causal intermediate  $D$  as in Figure 3 (Rosenbaum, 1984; Robins, 1997; Hernan et al., 2000; Petersen et al., 2006). The problem is that if one controls for such a causal intermediate in the regression analysis, one blocks a causal pathway of interest; in the above diagram, by controlling for diaphragm use  $D$ , one would block the causal pathway from study arm  $R$  through diaphragm use  $D$  to HIV status  $I$ , which is an important part of what we are trying to estimate. Alternatively, by not controlling for such a causal intermediate, one’s estimates may be biased due to confounding from the causal intermediate. The analysis we outline below adequately handles confounding by measured causal intermediates.

In the presence of confounding by causal intermediates, the randomization assumption analogous to assumption (1) of Section 5.1 is more complicated. It involves not only potential HIV outcomes  $I_{rc}$  under different scenarios for study arm assignment  $r$  and condom use  $c$ , but also potential outcomes for diaphragm use under different possible study arm assignments. We let  $D_r$  represent what an individual’s diaphragm use would be, would she have been assigned to study arm  $r$ . The randomization assumption analogous to (1) now takes the form,

$$C \perp\!\!\!\perp \{I_{rc}, D_r\} | W, R, D. \quad (4)$$

Intuitively, given knowledge of a subject’s baseline measurements  $W$ , study arm assignment  $R$ , and diaphragm use  $D$ , her observed condom use gives no additional information as to either (1) her risk of HIV infection would she have been assigned to arm  $r$  and would she have used condoms at fixed level  $c$  or (2) her diaphragm use were she assigned to the other study arm  $R = r$ .

In the case here of confounding by causal intermediates, the experimental treatment assignment assumption is that for each possible  $w, r, c, d$ , we have  $\Pr(C = c | R = r, D = d, W = w) > 0$ . That is, no strata of treatment assignment, diaphragm use, and baseline confounders precludes the possibility of using condoms at any specific level. The consistency assumption in this setting connects the observed data (including  $W, R, D, C, I$ ) to the set of potential outcomes  $\{I_{rc}, D_r\}$ . It states that the observed diaphragm use  $D$  is equal to the potential outcome  $D_r$  when  $R = r$ ; furthermore, it states that infection status  $I$  is equal to the potential outcome  $I_{rc}$  when observed study arm assignment  $R = r$  and observed condom use  $C = c$ .

Analogous to the derivation in Section 5.1, the assumption (4), the consistency assump-

tion, and the experimental treatment assignment assumption imply:

$$\Pr(I_{rc} = 1) = \sum_{w,d} \frac{\Pr(I = 1, R = r, C = c, D = d, W = w)}{\Pr(R = r|W = w)\Pr(C = c|W = w, R = r, D = d)} \quad (5)$$

Estimating  $\Pr(I_{rc} = 1)$  using (5) is equivalent to solving the following estimating equation:

$$E \left( \frac{1}{\Pr(R|W)\Pr(C|W, R, D)} (I - m(R, C|\beta)) \right) = 0 \quad (6)$$

where  $m(r, c|\beta)$  denotes a saturated (non-parametric) model for the unknown probability  $\Pr(I_{rc} = 1)$ .

Here, as in Section 5.1, we estimate the numerator in the right hand side of Equation 5 by the empirical distribution of the data, and the denominator by fitting a logistic regression model for  $\Pr(C = c|W = w, R = r, D = d)$ . We again use the bootstrap to generate confidence intervals for such estimates.

## 6. Time-dependent analysis

We have so far discussed estimating direct effects only in the single time point scenario, in which baseline covariates  $W$  are measured, followed by assignment to a study arm  $R$ , followed by further measurements  $D$  and  $C$ , and lastly outcome  $I$  is measured. One striking disadvantage of this simple approach is that the intermediate variable condom use is based on a single summary whereas actual condom use varies over time. This is also the case for measures of diaphragm use as a confounding variable of condom use as discussed in Section 5.2. In this section, we extend the IPTW estimators of Section 5 to provide a time-dependent analysis of direct effects, again using the MIRA trial as an illustration. The description below is an application of the estimation methodology for causal inference based on longitudinal data given in Section 6.4 of van der Laan and Robins (2003). Note that a single time-dependent regression equation adjusting for causal intermediates would suffer from the problem of confounding by causal intermediates discussed in Section 5.2 at every time point. The analysis presented in this section adequately deals with measured, time-dependent confounders.

In this section, we first describe why we chose to use the IPTW estimator to estimate direct effects in the MIRA trial. Then we describe exactly what time-dependent observations are made for each subject in the MIRA trial. Next, we link the observed data to potential outcomes. In terms of these potential outcomes, we present the main assumptions needed for our estimator to give a consistent estimate of the effect of study arm assignment on HIV infection, for fixed patterns of condom use over time. We then present a model for the hazard of HIV infection at each visit, capitalizing on information about the time lag between HIV infection and when antibodies are detectable. Finally, we give the proposed direct effects estimators with brief details about implementation.

### 6.1. Why We Chose the Inverse Probability of Treatment Weighted (IPTW) Estimator

We chose to use the IPTW estimator rather than other possible estimators (including G-computation estimators (Robins, 1986), doubly robust estimators (van der Laan and Robins, 2003), and targeted maximum likelihood estimators (van der Laan and Rubin, 2006)). The

decision of which estimator is most appropriate for a given data analysis requires knowledge of which parts of the data generating distribution can most reliably be modelled and estimated from the available data. Our decision to use the IPTW estimator was based on the fact that it only requires correct specification of models for the probability of condom use at visit  $v$ , given observed covariates at or before visit  $v$ . In contrast, the G-computation estimator would require correct models for the probability of HIV infection given past covariates and, at each visit  $v$ , models for the joint probability of all time-dependent confounders measured at visit  $v$  given past covariates. The doubly robust estimator and targeted maximum likelihood estimator would require models for all of the aforementioned probabilities (but are robust to incorrect specification of some of these models).

In consultation with the investigators in the MIRA trial, we decided that the models required for the IPTW estimator could be more reliably specified and estimated than the other models mentioned above. One reason is that in our data, among individuals with a given covariate history, there often are many subject-visits at which condoms were used, as well as many subject-visits where condoms were not used; this provided adequate data for estimating the probability of condom use given covariate history. In contrast, HIV infections were relatively rare, as were some values of time-dependent confounders (such as having multiple partners), so data for estimating their probabilities was sparse.

## 6.2. Observed Data

At enrolment, each subject's baseline covariates  $W$  (as listed in Section 5) were measured, prior to study arm assignment. At each of up to 8 follow-up visits  $v$ , the following potential time-dependent confounders (as listed in Section 5) were measured:

- $S(v)$ , the self-report of whether subject had sex at all over last three months (1 indicates had sex at least once, 0 indicates was abstinent).
- $D(v)$ , the vector of indicators recording diaphragm use at last sex, HSV-2 incidence, and whether sexual partners were known to have multiple partners over last three months.
- $C(v)$ , condom use at last sex.

Ignoring missing data for the moment, data collected on a single participant from enrolment ( $v = 0$ ) through the last follow-up visit ( $v = 8$ ) is then represented as:

$$O = (W, R, S(0), D(0), C(0), I(0), \dots, S(8), D(8), C(8), I(8)). \quad (7)$$

Note that since HIV infection is permanent, we have that if someone is uninfected at visit  $v$  ( $I(v) = 0$ ) then she must have been uninfected at all previous visits ( $I(v-1), I(v-2), \dots, I(0)$  all equal 0). Based on this data, the goal is estimation of the effect of study arm assignment on the hazard of HIV infection, under fixed patterns of condom use over time.

The time-ordering implied by our representation (7) of each participant's data have important consequences. If these variables do not actually occur in this order, our analysis will be biased. For example, if a participant's decision to use a diaphragm at last sex before visit 1 (represented by  $D(1)$ ) occurred after the decision to use a condom (represented by  $C(1)$ ), then our ordering would be incorrect leading to bias in our estimator. Our assumption that the decision to use a diaphragm at a given sex act precedes the decision to



use a condom is based on the fact that the diaphragm can be inserted up to 24 hours prior to sex, while a condom must be put on immediately before sex; however, this is no guarantee that our assumption is correct. As a sensitivity analysis to our time-ordering assumption, we also ran our direct effect analyses using a time-ordering in which  $C(i)$  precedes  $D(i)$ , and got nearly identical results to the analysis using the time-ordering (7). However, in reality the decision to use condoms and/or diaphragms will be more complex; whether such a decision is adequately captured by our time ordering assumption is, unfortunately, not testable from the data collected.

Also, we need to lag all measurements except our HIV status measure by one visit, since HIV infection is detectable by the tests used in the study within a 3 month window, and usually takes at least 3 weeks to be detectable. Though this lag was included in our data analyses, we omit it in explanations of our estimator below for clarity.

We record  $C(v)$  as 2, to represent “undefined,” if the participant did not have sex in the previous three months  $S(v) = 0$ .

We represent the set of observations of condom use over all visits  $C(0), C(1), \dots, C(8)$  by the more compact notation  $\bar{C}$ , and similarly for  $\bar{S}$ ,  $\bar{D}$  and  $\bar{I}$ . We use the notational convention that the first  $k$  components of a vector  $\bar{A}$  is represented by  $\bar{A}(k)$ ; that is,  $\bar{A}(k) = (A(1), \dots, A(k))$ . We denote the vector of length 8 of all 0's by  $\bar{0}$ , and the corresponding vector of all 1's by  $\bar{1}$ .

### 6.3. Potential Outcomes

As done for the single time point case in Section 4.2, we define direct effects in terms of potential outcomes. For each possible combination of fixed treatment assignment,  $r$ , and condom use over time  $\bar{c}$ , we let  $I_{r\bar{s}\bar{c}}(v)$  represent whether a subject would be HIV-infected by visit  $v$  were her study arm, abstinence or not, and condom use constrained to  $r$ ,  $\bar{s}$  and  $\bar{c}$ , respectively. This is only defined for  $\bar{s}, \bar{c}$  such that  $s(v) = 0$  implies  $c(v) = 2$  (the coding for “undefined”), to avoid nonsensical potential outcomes, such as  $s(1) = 0, c(1) = 1$ , which would mean abstinence and condom use simultaneously. Similarly, we define  $D_{r\bar{s}\bar{c}}(v)$  whether a subject would use a diaphragm at last sex before visit  $v$  were her study arm, abstinence or not, and condom use constrained to  $r$ ,  $\bar{s}$  and  $\bar{c}$ , respectively. Note that  $D_{r\bar{s}\bar{c}}(v)$  is only defined for  $s(v) = 1$ , which is reflected in our model in Section 6.5.

We can formulate our parameters of interest in terms of potential outcomes. We are aiming to estimate the probabilities of HIV infection by the end of the trial for each study arm, were all subjects constrained to have sex at least once between quarterly visits, and to either (1) always use condoms at last sex before each quarterly visit or (2) never use condoms at last sex before each quarterly visit. That is, we want to estimate  $p_{r\bar{c}} := \Pr(I_{r\bar{s}\bar{c}}(8) = 1)$  for  $\bar{s} = \bar{1}$ ,  $r \in \{0, 1\}$  and  $\bar{c} \in \{\bar{0}, \bar{1}\}$ . We can then take the appropriate ratios of these quantities to get the relevant direct effects relative risks. Note that it is not possible to directly identify from the data the probabilities of the simpler potential outcomes that ignore sexual abstinence:  $\Pr(I_{r\bar{c}}(8) = 1)$ , since we never observe these potential outcomes when subjects abstain from sex between visits; our estimators of this simpler quantity would violate the ETA assumption, so we incorporated  $\bar{s}$  into our model to avoid this problem.

### 6.4. Main Assumptions: Consistency, No Unmeasured Confounders and ETA

We make the following assumptions, which together imply that our direct effects estimator will be consistent and asymptotically normal. First, the consistency assumption is that

at each visit  $v$ , (1) observed infection status  $I(v)$  equals the potential outcome  $I_{r\bar{s}\bar{c}}(v)$  whenever a subject's arm assignment  $R = r$ , abstinence or not  $\bar{S} = \bar{s}$  and condom use over time  $\bar{C} = \bar{c}$  and (2) observed diaphragm use  $D(v)$  equals the potential outcome  $D_{r\bar{s}\bar{c}}(v)$  whenever  $R = r$ ,  $\bar{S} = \bar{s}$  and  $\bar{C} = \bar{c}$ .

Second, we make a conditional independence assumption that implies there are no unmeasured confounders of the relationship between abstinence and HIV infection over the visits  $v$ . We assume that

$$S(v) \perp\!\!\!\perp \{I_{r\bar{s}\bar{c}}, D_{r\bar{s}\bar{c}}\} | W, R, \bar{S}(v-1), \bar{D}(v-1), \bar{C}(v-1), I(v-1) = 0. \quad (8)$$

Third, we make the following conditional independence assumption that implies there are no unmeasured confounders of the relationship between condom use and HIV infection over the visits  $v$ . We assume

$$C(v) \perp\!\!\!\perp \{I_{r\bar{s}\bar{c}}, D_{r\bar{s}\bar{c}}\} | W, R, \bar{S}(v-1), S(v) = 1, \bar{D}(v), \bar{C}(v-1), I(v-1) = 0. \quad (9)$$

We include diaphragm use at visit  $v$  in our set of observations that precede condom use at visit  $v$  since we think diaphragm use is determined before condom use, as discussed in Section 6.2 above. Also, we only have to assume the conditional independence given  $S(v) = 1$ , since  $C(v)$  is automatically 2 (“undefined”) if  $S(v) = 0$ . (9) is a generalization of the assumptions (1) and (4) from the single time point case. The same is true for (8).

Fourth, we make the so-called experimental treatment assignment (ETA) assumption that the probabilities  $\Pr(C(v) = 1 | W, R, \bar{S}(v-1), S(v) = 1, \bar{D}(v), \bar{C}(v-1), I(v-1) = 0)$  do not equal 0 or 1, almost surely, for any values of  $W, R, \bar{S}(v-1), \bar{D}(v), \bar{C}(v-1)$ .

Finally, we make a time-ordering assumption for our potential outcomes. This assumption is that for each visit  $v$ , the potential outcomes for that visit do not depend on the prescribed abstinence or condom use at future visits; only the past and present constraints affect the present HIV outcome. Formally, the assumption is that  $I_{r\bar{s}\bar{c}}(v)$  and  $D_{r\bar{s}\bar{c}}(v)$  only depend on  $r, \bar{s}(v)$ , and  $\bar{c}(v)$ , and not on  $s(v+1), \dots, s(8)$  or  $c(v+1), \dots, c(8)$ .

### 6.5. Model for the Hazard of HIV Infection

Some type of modelling is necessary to avoid extremely high variance in our estimator of the direct effects of treatment assignment fixing condom use at “never.” This is due to there being only 488 subjects who reported not using condoms at last sex, at all visits. Assuming a model, as we do below, will lead to some bias due to model misspecification. However, we are able to take advantage of important subject matter knowledge in creating the model; in particular, our model incorporates the known time lag between when an infection occurs and when it is detectable.

We now present the complete details of the model we used for the hazard of infection in our analyses, which we call  $m_{\text{full}}$ . This model incorporates the following information: visit number, study arm assignment, abstinence, condom use, and study site. There were three study sites (two in South Africa and one in Zimbabwe), which we represent in our model by two indicator variables:  $W_1, W_2$ , indicating the first two study sites, respectively. Note that these variables are a subset of the baseline covariates  $W$ . Subjects were asked at each quarterly visit if they abstained from sex over the last three months. Their response was recorded in the variable  $S(v)$ , for each visit  $v$ ; 1 indicates sex at least once over the last three months, and 0 indicates abstinence from sex over the last three months.

In our analysis of direct effects, we only use subject-visits for which the subject had sex at least once over the previous three months (that is,  $s(v) = 1$ ), since otherwise the subject would not have been exposed to HIV. We model the following conditional probability of potential outcomes, for  $s(v) = 1$ :

$$\Pr(I_{r\bar{c}\bar{s}}(v) = 1 | I_{r\bar{c}\bar{s}}(v-1) = 0, W_1 = w_1, W_2 = w_2) = m_{\text{full}}(v, r, \bar{c}(v), \bar{s}(v), w_1, w_2 | \beta^0), \quad (10)$$

for all  $v, r, \bar{c}(v), \bar{s}(v-1), s(v) = 1, w_1, w_2$ . We call the model  $m_{\text{full}}$  to distinguish it from the simpler model  $m$  in a different version of this paper.

In our particular choice below for a low-dimensional model  $m_{\text{full}}$ , we exploit the following information about the test for HIV infection: HIV infection will, with high probability, be detectable within a 3 month window, and usually takes at least 3 weeks to be detectable. Therefore, (1) condom use measured at the current visit should not have an effect on the outcome of the HIV test at that visit and (2) condom use in the previous 3 months may have a large effect on whether HIV is detected at a given visit, compared to condom use farther in the past. Condom use in the past cannot be completely ignored, since it may impact future behaviour, such as future diaphragm use. We incorporated (1) and (2) into the model by collapsing information about an entire condom use history into just the most recent reported condom use and a single number giving the average condom use before this.

Now we specify precisely how  $m_{\text{full}}$  depends  $v, r, \bar{c}, \bar{s}, w_1, w_2, \beta$ . Let  $1_{x=y}$  be the indicator function that is 1 if  $x = y$ , and is 0 otherwise, for  $x, y$  any mathematical expressions. For example,  $1_{r=1}$  is 1 if  $r = 1$ , and is 0 otherwise. Define the function  $a$  (for ‘‘average’’) that summarizes information in a vector, such as vector of condom use history (or of abstinence from sex), as  $a(\bar{c}(v)) = \sum_{v'=0}^v c(v') / (v + 1)$ . Define  $\text{expit}(x) = \exp(x) / (1 + \exp(x))$ .

Since we stated above that our model for visit  $v$  is only defined for  $s(v) = 1$  (since otherwise there is no exposure to HIV) we need only specify the model  $m_{\text{full}}(v, r, \bar{c}(v), \bar{s}(v), w_1, w_2 | \beta)$  for  $s(v) = 1$ . In this case, we define

$$m_{\text{full}}(v, r, \bar{c}(v), \bar{s}(v), w_1, w_2 | \beta) :=$$

$$\begin{aligned} & 1_{c(v)=0, r=0} \text{expit} \{ \beta_{00} + \beta_{01}v + \beta_{02}w_1 + \beta_{03}w_2 + \beta_{04}a(\bar{c}(v-1)) + \beta_{05}a(\bar{s}(v-1)) \} \\ + & 1_{c(v)=1, r=0} \text{expit} \{ \beta_{10} + \beta_{11}v + \beta_{12}w_1 + \beta_{13}w_2 + \beta_{14}a(\bar{c}(v-1)) + \beta_{15}a(\bar{s}(v-1)) \} \\ + & 1_{c(v)=0, r=1} \text{expit} \{ \beta_{20} + \beta_{21}v + \beta_{22}w_1 + \beta_{23}w_2 + \beta_{24}a(\bar{c}(v-1)) + \beta_{25}a(\bar{s}(v-1)) \} \\ + & 1_{c(v)=1, r=1} \text{expit} \{ \beta_{30} + \beta_{31}v + \beta_{32}w_1 + \beta_{33}w_2 + \beta_{34}a(\bar{c}(v-1)) + \beta_{35}a(\bar{s}(v-1)) \} \end{aligned} \quad (11)$$

where  $\beta$  is an unknown, time-invariant, population parameter vector. This is an example of a marginal structural model (Robins, 1998). There are 24 parameters (the different  $\beta_{ij}$ ’s) in this model. We denote the vector of variables appearing as arguments in the functions  $\text{expit}$  above as  $H(v, \bar{s}(v-1), \bar{c}(v-1), w_1, w_2) = (1, v, w_1, w_2, a(\bar{c}(v-1)), a(\bar{s}(v-1)))$ .

The model we used, in effect, stratifies by condom use over the previous 3 months and by study arm. For example, we only use data from participants who used condoms at visit 1 and were in the diaphragm arm, to estimate the hazard of infection at visit 1 for condom use set to 1 and arm set to 1; that is, we don’t extrapolate from non-condom users at a particular visit to estimate hazard of infection for condom-users. The resulting model is less dependent on model assumptions than a simple logistic model  $\text{expit}(\beta X)$  that incorporates arm  $r$  and condom use at visit  $v$  as terms in  $X$ ; however, our larger model yields wider confidence intervals than this simpler model.

Once we estimate the parameter  $\beta$ , as described in the next section, we can use it to estimate the probabilities  $p_{r\bar{c}}$  using the product integral (Andersen et al., 1993) for calculating the cumulative survival probability from discrete hazards, as follows (where  $\bar{s} = \bar{1}$ ):

$$\begin{aligned} p_{r\bar{c}} &:= \Pr(I_{r\bar{s}\bar{c}}(8) = 1) \\ &= \sum_{w_1, w_2} P(W_1 = w_1, W_2 = w_2) \Pr(I_{r\bar{s}\bar{c}}(8) = 1 | W_1 = w_1, W_2 = w_2). \\ &= \sum_{w_1, w_2} P(W_1 = w_1, W_2 = w_2) \left[ 1 - \prod_{v=1}^8 \{1 - m_{\text{full}}(v, r, \bar{s}, \bar{c}(v), w_1, w_2 | \beta^0)\} \right] \end{aligned} \quad (12)$$

where  $\beta^0$  is the true value of the model parameter. Here, we are combining hazards over all visits for each site separately, and then weighting each site-specific, estimated probability of infection by the size of that site.

### 6.6. Direct Effects Estimator in the Time-Dependent Setting

In this section, we present the estimating equation for the time-dependent setting that we use to estimate the parameter  $\beta$  of the model in (10), which is then used to estimate direct effects relative risks. The estimating equation uses “stabilized inverse probability of treatment weights (IPTW),” (which we refer to more concisely as “stabilized weights”), which are a generalization of standard inverse probability of treatment weights. In our estimating equation, individual subjects contribute information for all visits except those occurring after HIV seroconversion.

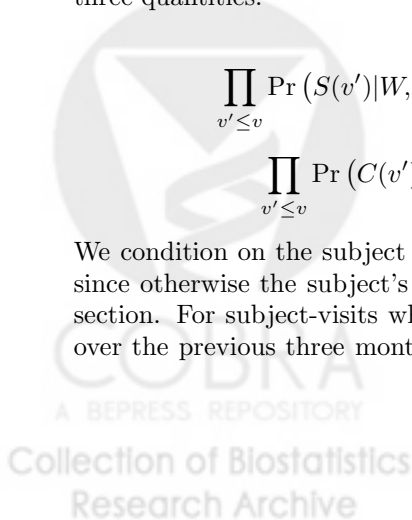
It turns out that solving this estimating equation is equivalent to solving a weighted logistic regression, which can be done with standard statistical software. The details of how to solve this type of estimating equation using weighted logistic regression are given in Section 6.4 of van der Laan and Robins (2003).

#### 6.6.1. Stabilized Weights

Before giving the estimating equation used to estimate  $\beta$ , we describe the stabilized weights that will be used in this estimating equation. First, consider what the standard (non-stabilized) inverse probability of treatment weights would be in the time-dependent setting. For each visit  $v$ , the standard IPTW weight is the reciprocal of the product of the following three quantities:

$$\begin{aligned} &\Pr(R|W), \\ &\prod_{v' \leq v} \Pr(S(v') | W, R, \bar{S}(v' - 1), \bar{D}(v' - 1), \bar{C}(v' - 1), I(v' - 1) = 0), \\ &\prod_{v' \leq v} \Pr(C(v') | W, R, \bar{S}(v'), \bar{D}(v'), \bar{C}(v' - 1), I(v' - 1) = 0). \end{aligned}$$

We condition on the subject being not HIV infected at the previous visit ( $I(v' - 1) = 0$ ), since otherwise the subject’s visit is excluded from the analysis, as explained in the next section. For subject-visits where  $S(v) = 0$ , that is, where the subject abstained from sex over the previous three months, the last quantity above equals 1; this follows since if the



participant did not have sex, then condom use at last sex is coded as “undefined” ( $C(v) = 2$ ), which occurs with probability 1 given no sex over last three months. Note that the IPTW weights used for the single time point estimating equations (3) and (6) in Sections 5.1 and 5.2 are special cases (with  $v = 1$ ) of this general formula.

Stabilized weights are a modification of the standard IPTW weights. Stabilized weights include a numerator term that can only depend on the terms included in the model (10), which include  $v$ ,  $R$ ,  $\bar{S}(v)$ , and  $\bar{C}(v)$ ,  $W_1, W_2$  (but not, for example,  $\bar{D}$ ). They are called “stabilized” since the hope is that for large values of the standard IPTW weights, which correspond to small values in the denominator, the numerator terms will be small; in this case, one would prevent single observations from having a disproportionate amount of weight in the estimation procedure. The stabilized weights are similar to the standard IPTW weights given above, except they have the product of the following three functions in the numerator, for each visit  $v$ :

$$\Pr(R|W_1, W_2),$$

$$\prod_{v' \leq v} \Pr(S(v')|W_1, W_2, R, \bar{S}(v' - 1), \bar{C}(v' - 1), I(v' - 1) = 0),$$

$$\prod_{v' \leq v} \Pr(C(v')|W_1, W_2, R, \bar{S}(v'), \bar{C}(v' - 1), I(v' - 1) = 0).$$

Thus, the stabilized weights  $sw(v, W, R, \bar{S}(v), \bar{D}(v), \bar{C}(v), \bar{I}(v - 1))$  for each subject-visit  $v$  at which  $I(v - 1) = 0$  are defined as the product of following two fractions:

$$\frac{\prod_{v' \leq v} \Pr(S(v')|W_1, W_2, R, \bar{S}(v' - 1), \bar{C}(v' - 1), I(v' - 1) = 0)}{\prod_{v' \leq v} \Pr(S(v')|W, R, \bar{S}(v' - 1), \bar{D}(v' - 1), \bar{C}(v' - 1), I(v' - 1) = 0)} \quad (13)$$

$$\frac{\prod_{v' \leq v} \Pr(C(v')|W_1, W_2, R, \bar{S}(v'), \bar{C}(v' - 1), I(v' - 1) = 0)}{\prod_{v' \leq v} \Pr(C(v')|W, R, \bar{S}(v'), \bar{D}(v'), \bar{C}(v' - 1), I(v' - 1) = 0)}. \quad (14)$$

Note that by randomization,  $\Pr(R|W) = \Pr(R|W_1, W_2) = 1/2$ , which cancelled out in the numerator and denominator above.

Just as in Sections 5.1 and 5.2, we use logistic regression models to estimate the IPTW weights. We estimate each term in the product in the numerators of (13) and (14) by fitting a separate logistic regression model; we do the same for each term in the product in the denominator. Since for each subject, their weights at a particular visit  $v$  are only used in the estimating equation we give below if they are not HIV infected at the previous visits ( $I(v - 1) = 0$ ), we only have to estimate the above weights at each visit  $v$  for the subpopulation who have  $I(v - 1) = 0$ . Also, we only need to model the probability that  $C(v) = 1$  conditional on the event  $S(v) = 1$ , since otherwise  $C(v)$  is deterministic (coded as 2 to represent “undefined”). In deciding which variables to put in the logistic regressions models for estimating the stabilized weights, we use the assumption that abstinence at visit  $v$  and condom use at visit  $v$  are more influenced by variables measured at the previous visit than by variables measured further in the past. This model assumption is incorporated into the logistic regression models by including all variables measured in the past visit, but only summary measures of the variables measured at earlier visits.

To estimate the stabilized weights for each subject, we need models for the terms in the products in the numerators and denominators of (13) and (14). We construct a separate

set of models for each study arm, since we think the mechanism connecting past behavior to current abstinence and condom use may be quite different depending on whether one received the diaphragm intervention or not; thus, each model below is fit first using only subject-visits from the diaphragm arm, and then separately using only subject-visits from the control arm. We suppress the dependence of the model on study arm for clarity. First consider the terms in the numerators above. For each visit  $v$ , we model

$$\Pr(S(v)|W_1, W_2, R, \bar{S}(v-1), \bar{C}(v-1), I(v-1) = 0)$$

by a logistic regression model  $g_{1v}$ , with only main terms; these main terms are  $W_1, W_2, S(v-1), a(\bar{S}(v-2)), C(v-1), a(\bar{C}(v-2))$ . Similarly, we model

$$\Pr(C(v)|W_1, W_2, R, S(v) = 1, \bar{S}(v-1), \bar{C}(v-1), I(v-1) = 0)$$

by a logistic regression model  $g_{2v}$ , with only main terms; these main terms are the following:  $W_1, W_2, S(v-1), a(\bar{S}(v-2)), C(v-1), a(\bar{C}(v-2))$ ; since only subject-visits for which  $S(v) = 1$  are used in the estimating equation below, when fitting this model we restrict to the subpopulation for which  $S(v) = 1$ .

Next, consider the terms in the denominators of (13) and (14) above. We only need models for the case in which  $I(v-1) = 0$ , since we only use subject-visits in the estimating equation below for which infection has not yet occurred. We model

$$\Pr(S(v)|W, R, \bar{S}(v-1), \bar{D}(v-1), \bar{C}(v-1), I(v-1) = 0)$$

by a logistic regression model  $g'_{1v}$ , with only main terms; these main terms are  $W, S(v-1), a(\bar{S}(v-2)), D(v-1), a(\bar{D}(v-2)), C(v-1), a(\bar{C}(v-2))$ , where  $W$  is the list of all baseline variables (including site:  $W_1, W_2$ ) given in Section 5. Note that in the control arm,  $D(v) = 0$  at almost all subject-visits, so is redundant in the model. Similarly, we model

$$\Pr(C(v)|W, R, S(v) = 1, \bar{S}(v-1), \bar{D}(v), \bar{C}(v-1), I(v-1) = 0)$$

by a logistic regression model  $g'_{2v}$ , with only main terms; these main terms are  $W, a(\bar{S}(v-1)), D(v), a(\bar{D}(v-1)), C(v-1), a(\bar{C}(v-2))$ ; since only subject-visits for which  $S(v) = 1$  are used in the estimating equation below, when fitting this model we restrict to the subpopulation for which  $S(v) = 1$ .

Denote the above fitted logistic regression models  $g_{1v}, g_{2v}, g'_{1v}, g'_{2v}$  by  $\hat{g}_{1v}, \hat{g}_{2v}, \hat{g}'_{1v}, \hat{g}'_{2v}$  respectively.

The estimated stabilized weight for each subject-visit, which is a function of an individual's history at and before that visit, is then computed as:

$$\begin{aligned} \hat{sw}(v, W, R, \bar{S}(v), \bar{D}(v), \bar{C}(v), \bar{I}(v-1)) \\ = \frac{\prod_{v'=1}^v \hat{g}_{1v'}(v', W, R, \bar{S}(v'), \bar{C}(v'-1)) \hat{g}_{2v'}(v', W, R, \bar{S}(v'), \bar{D}(v'), \bar{C}(v'))}{\prod_{v'=1}^v \hat{g}'_{1v'}(v', W, R, \bar{S}(v'), \bar{D}(v'), \bar{C}(v'-1)) \hat{g}'_{2v'}(v', W, R, \bar{S}(v'), \bar{D}(v'), \bar{C}(v'))}, \end{aligned} \quad (15)$$

where each occurrence of a fitted model in the above equation denotes the value obtained by substituting the values recorded for the individual subject  $(v, W, R, \bar{S}(v), \bar{D}(v), \bar{C}(v), \bar{I}(v-1))$  into the fit model; thus, each occurrence of a fitted model represents a number between 0 and 1, indicating the corresponding probability of abstinence/condom use given that individual's past covariates.

An additional model assumption we made is that no stabilized weight is larger than 25, which we incorporated by truncating all estimated stabilized weights from (15) at 25; this assumption, qualitatively, is that there are no sets of covariates for which the probability of condom use at last sex is very close to 0 or to 1.

### 6.6.2. Estimating Equation

Having defined the stabilized weights for each subject-visit, we now can construct the estimating equation we use to fit the model  $m_{\text{full}}$  in (10). Let  $O_i$  denote the entire set of data collected on the  $i$ th subject enrolled in the study, and let  $n$  denote the number of subjects in the study. Recall that  $H(v, \bar{S}(v-1), \bar{C}(v-1), W_1, W_2)^t$ , which we denote by  $H^t$  below, denotes the column vector of basis functions used in the model  $m_{\text{full}}$  in (10) above. Below we omit the subscript “full” on  $m_{\text{full}}$  so the equation will fit on a single line. The estimating equation corresponding to the IPTW estimator with stabilized weights is given by

$$\sum_{i=1}^n \sum_{v=1}^8 1_{S(v)=1, I(v-1)=0} \text{sw}(v, W, R, \bar{S}(v), \bar{D}(v), \bar{C}(v), \bar{I}(v-1)) \times \\ \{I(v) - m(v, R, \bar{C}(v), \bar{S}(v), W_1, W_2, \beta)\} H^t = \mathbf{0}$$

Note the similarity in form to the simpler estimating equations (3) and (6). Details and proof that this estimating equation indeed has mean 0 at the true  $\beta$  under the assumptions given in Sections 6.2-6.5 are given in Section 6.4 of van der Laan and Robins (2003).

The above estimating equation is solved in practice by substituting the estimated stabilized weights  $\hat{sw}$  from (15) for  $sw$  for each subject-visit in the equation above, and then solving this estimating equation for  $\beta$ .

Once we obtain a solution  $\hat{\beta}$  to the above estimating equation, we can use it to generate estimates  $\hat{p}_{r\bar{c}}$  for the probability of HIV infection by the end of the trial for a randomly chosen subject in the study, would she have been assigned to study arm  $r$  and have had condom use profile set to  $\bar{c}$ , for  $\bar{s} = \bar{1}$ . Following (12) above, we have

$$\hat{p}_{r\bar{c}} = \sum_{w_1, w_2} \hat{P}r(W_1 = w_1, W_2 = w_2) \left[ 1 - \prod_{v=1}^8 \left( 1 - m_{\text{full}}(v, r, \bar{c}(v), \bar{s}(v), w_1, w_2 | \hat{\beta}) \right) \right], \quad (16)$$

where  $\hat{P}r(W_1 = w_1, W_2 = w_2)$  denote the empirical probabilities of being at each of the three sites; these are just the proportion of subjects at each site.

### 6.6.3. Solving the Estimating Equation with Standard Statistical Software

We now discuss how this estimating equation is equivalent to fitting a weighted logistic regression model, that can be solved with standard statistical software. First, we fit the models  $g_{1v}, g_{2v}, g'_{1v}, g'_{2v}$  defined above, and use them to calculate estimated stabilized weights according to (15) for each subject-visit at which the subject is at risk for HIV infection (that is, for which  $I(v-1) = 0$  and  $S(v) = 1$ ). Solving the above estimating equation is then equivalent to doing maximum likelihood estimation separately on four weighted logistic regression models, as we now explain. The first weighted logistic regression model will correspond to  $r = 0, c(v) = 0$  (the first line in the definition of the model  $m_{\text{full}}$  in (11). Consider just the subject-visits in the control arm ( $R = 0$ ) in which  $S(v) = 1$  and  $I(v-1) = 0$ , but condoms were not used at last sex ( $C(v) = 0$ ). We fit a weighted logistic regression model

$$\Pr(I(v) = 1 | R = 0, W_1, W_2, I(v-1) = 0, S(v) = 1, \bar{S}(v-1), C(v) = 0, \bar{C}(v-1)) \\ = \text{expit} \{ \beta_{00} + \beta_{01}v + \beta_{02}w_1 + \beta_{03}w_2 + \beta_{04}a(\bar{c}(v-1)) + \beta_{05}a(\bar{s}(v-1)) \},$$

with weights for each subject-visit the stabilized weights (15), and restricted to subject-visits at which  $R = 0, S(v) = 1, I(v - 1) = 0$ , and  $C(v) = 0$ . This can be done with standard logistic regression software, using a pooled sample in which each subject contributes multiple observations, and assigning each subject's  $v$ -specific observation the weight

$$1_{S(v)=1, I(v-1)=0} s\hat{w}(v, W, R, \bar{S}(v), \bar{D}(v), \bar{C}(v), \bar{I}(v - 1)) \tag{17}$$

In other words, we first create a data set with a single record for each subject-visit; each subject will contribute up to eight records, one for each visit, where a record contains the visit number  $v$ , and the subject's study arm, HIV status, abstinence history and condom use history through visit  $v$ . If a subject becomes HIV positive during the trial, we only use her subject-visits through the visit at which she seroconverted; also we only use subject-visits at which the subject had sex at least once over the previous three 3 months.

We repeat the procedure above three more times, corresponding to the last three lines in the definition of the model  $m_{full}$  in (11). That is, we do it for the set of subject-visits for which  $R = 1, C(v) = 0$ , then the set of subject-visits for which  $R = 0, C(v) = 1$ , then the set of subject-visits for which  $R = 1, C(v) = 1$ . At the end we have fit four weighted logistic regression models, giving estimates for all the parameters  $\beta$  in  $m_{full}$ . We use the product integral in (16) above to then separately construct estimates of  $p_{r\bar{c}}$  for  $r \in \{0, 1\}$  and  $\bar{c} \in \{\bar{0}, \bar{1}\}$ .

We can then construct estimates of direct effects relative risks by taking ratios of the estimates  $\hat{p}_{r\bar{c}}$ . The direct effects relative risk estimate with condom use fixed at "never" is  $\hat{p}_{1\bar{0}}/\hat{p}_{0\bar{0}}$ . (Recall that  $\bar{0}$  denotes the vector of length 8 of all 0's.) The direct direct effects relative risk estimate with condom use fixed at "always" is  $\hat{p}_{1\bar{1}}/\hat{p}_{0\bar{1}}$ . We point out that the estimators in Sections 5.1 and 5.2 differ from the estimator for the time-dependent case described in this section in that the former do not assume a model on potential outcomes. Here, we assumed the model  $m_{full}$  in (10), to deal with the high-dimensional nature of the time-dependent data.

We used the bootstrap percentile method for constructing confidence intervals for these direct effects relative risk estimates, where replicates were selected using subjects (not subject-visits) as the experimental unit. Each bootstrap replicate requires not only refitting the model  $m_{full}$ , but also the models  $g_{1v}, g_{2v}, g'_{1v}, g'_{2v}$ , using the resampled population.

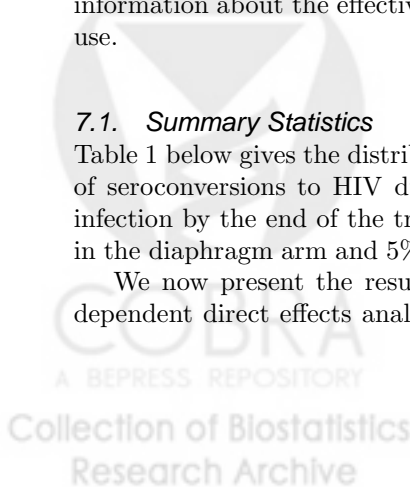
## 7. Analysis of HIV Infection in the MIRA Trial

In this section, we give summary statistics for the MIRA trial and then the results of the single time point analysis, followed by the time-dependent analysis. In both types of analysis, there are no significant results, and confidence intervals are too wide to provide any information about the effectiveness of the diaphragm intervention at fixed levels of condom use.

### 7.1. Summary Statistics

Table 1 below gives the distribution of the two arms of the study with regard to the number of seroconversions to HIV during follow-up. The intention-to-treat relative risk of HIV infection by the end of the trial is  $158/151 = 1.05$  (95% CI: 0.84,1.30). 6% of participants in the diaphragm arm and 5% of the participants in the control arm were lost to follow-up.

We now present the results of the time-dependent direct effects analysis. The time-dependent direct effects analysis was based on a discrete time scale made up of the eight





**Table 1. SUMMARY OF HIV SEROCONVERSIONS BY STUDY ARM**

	Number of Participants	Seroconversions (percent of group)	Woman-years of Follow-up	Incidence Rate (per 100 woman-years)
Total	4948	309 (6.2%)	7702	4.0
Intervention	2472	158 (6.4%)	3835	4.1
Control	2476	151 (6.1%)	3867	3.9

quarterly visits by each participant, in addition to an initial enrolment visit (visit 0). For individuals who seroconverted, the (discrete) time to seroconversion was defined as the visit number at the time of the first positive HIV test. For participants who missed one or more visits between the last negative and first positive test, the seroconversion time was defined as the visit number covering the midpoint between the two discordant tests. Covariates were imputed by carrying forward the most recently observed value; for example, if a subject came to visit 3 but missed visits 4 and 5, her covariates for visits 4 and 5 were imputed to be what was recorded at visit 3.

Before calculating the direct effects relative risks of study arm assignment on HIV infection with condom use fixed, we looked at whether the MIRA trial data were consistent with reported condom use being causally related to HIV infections. If it were not, then a direct effects analysis fixing reported condom use would probably not be of much use. We did a causal analysis of the effect of condom use at last sex on subsequent HIV seroconversion, within each study arm separately. We did this by estimating  $p_{r\bar{c}}$  for  $r \in \{0, 1\}$ ,  $\bar{c} \in \{\bar{0}, \bar{1}\}$  as described in Section 6. But instead of looking at the direct effects relative risks holding condom use fixed as described in that section, we looked at the following relative risk:  $p_{r\bar{c}}/p_{r\bar{c}'}$ , first for  $r = 0$ ,  $\bar{c} = \bar{1}$ ,  $\bar{c}' = \bar{0}$ , which is the relative risk of HIV infection comparing always to never condom use in the control arm. We then calculated this relative risk  $p_{r\bar{c}}/p_{r\bar{c}'}$ , except setting  $r = 1$ ,  $\bar{c} = \bar{1}$ ,  $\bar{c}' = \bar{0}$ , which gives the relative risk for the diaphragm arm. The estimates for these measures of condom efficacy are: 0.33 (95% CI 0.16, 2.33) for the control arm and 0.54 (95% CI: 0.30, 0.98) for the diaphragm arm, indicating a statistically significant protective effect in the diaphragm arm but no statistically significant result in the control arm; here, confidence intervals were calculated using non-parametric bootstrap percentile method. One reason for the larger confidence intervals in the control arm is that there were fewer instances of non-condom use in this arm, so the denominator of the relative risk was more variable. The relative protection corresponding to the point estimate for each arm is  $1 - 0.33 = 67\%$  for the control arm and  $1 - 0.54 = 46\%$  for the diaphragm arm. This protection against HIV infection is less than the approximately 80% relative protection normally ascribed to condom use (Weller and Davis-Beaty, 2002). The lower protection reflected in the above estimates may be due to overreporting of condom use due to social desirability bias, which would dilute the measured effectiveness of condoms.

Consider the following set of facts: (i) infection rates are similar in the two arms, (ii) reported condom use is much lower in the diaphragm arm compared to the control arm, and (iii) the data is consistent with condom use at last sex (as measured by self-report)

protecting against HIV infection. Together, these facts seem to suggest that the direct effect of assignment to the diaphragm arm that is not mediated through condom use patterns, may turn out to be positive.

### 7.2. Single Time Point Data Analysis

We start, for simplicity, with a simple time-fixed measure of condom use ( $C$ ) to apply the methods of Section 5, and then turn to the more informative setting where  $C$  is allowed to vary over time (visits). We use a fixed condom use measure that describes average use over the first three visits in the trial, and look at time to seroconversion for those who were uninfected by the end of the third visit. In this point-treatment analysis, we ignore condom use after the third visit. We will use condom measurements at all visits when we consider the time-dependent analysis later in this section.

Trial participants were divided into three categories based on the frequency of their positive response to the “condom use at last sex” question at each of the first three clinic visits. Some participants missed one or more of their initial visits, and so we categorized them based on the percent of unmissed visits among the first three visits at which they said they used condoms at last sex. Low condom users were defined to be those who responded positively less than 50% of the times questioned; medium condom users responded positively between 50–99% of the time, and high condom users responded positively 100% of the time. Among participants in the diaphragm arm, 49%, 28%, and 23% were low, medium and high condom users, respectively. In contrast, among those in the control arm, 11%, 32%, and 57% were low, medium and high condom users, respectively.

To estimate  $P(I_{rc} = 1)$  for a particular choice of  $r, c$  using (5) we need a model for  $P(C = c | W = w, R = r, D = d)$ . Here, we used multinomial logistic regression models given the three categories for  $C$ . Baseline confounders  $W$  were entered additively and included all baseline factors listed in Section 5. The set of confounders  $D$  were the following: (1) average diaphragm use (a continuous variable, defined as the fraction of visits among the first three at which a subject said she used diaphragms at last sex), (2) fraction of visits among first three at which a subject said she had sex at least once over past three months, and (3) fraction of visits at which a subject said her partner had had multiple partners over the past three months.

The identity (5), in particular, allows estimation of both  $P(I_{1c} = 1)$  and  $P(I_{0c} = 1)$  separately with their ratio yielding an estimated direct effects relative risk of assignment to the treatment arm for each of the three values of  $c$ . 10,000 bootstrap samples were used to provide estimates of the variability of these estimated direct effects relative risks, yielding 95% confidence intervals using the percentile method. The results are shown in Table 2. We point out that a similar analysis, except where participants were classified into low, medium, and high condom users based not on reported condom use at the first three visits but on condom use at all visits, yields qualitatively similar results.

### 7.3. Time Dependent Data Analysis

We now present the results of the time-dependent direct effects analysis described in Section 6. The list of baseline and time-dependent confounders included in the analyses are those listed in Section 5. Using the methods from Section 6, we can generate an estimate of  $p_{r\bar{c}}$

**Table 2.** (INTENT TO TREAT) ESTIMATES OF THE DIRECT EFFECTS OF ASSIGNMENT TO TREATMENT ARM ON HIV INFECTION FOR VARIOUS FIXED CATEGORIES OF CONDOM USE

Condom Use	Relative Risk of HIV Infection ( $R = 1$ vs. $R = 0$ )	95% Confidence Interval
Low	1.32	(0.63, 2.86)
Medium	0.93	(0.46, 1.48)
High	0.68	(0.43, 1.48)

for  $r = 0, 1$  and any specific pattern of condom use reported over the 8 possible visits, which is denoted by  $\bar{c}$ . Here we focus on two extreme patterns of condom use: (i) condoms used at last sex at all visits, so that  $c(v) = 1$  for all  $v$ , and (ii) condoms not used at last sex at all visits, so that  $c(v) = 0$  for all  $v$ . The methods can be used to estimate the direct effects for other condom use patterns as well.

Tables 2 and 4 give the results for these two extreme patterns of condom use. Table 2 gives the estimated direct effects of assignment to the diaphragm arm vs. control arm, for condom use set to “no” at all 8 visits. First, the direct effects relative risk 0.59(95% CI: 0.26, 4.56) is given; this is an estimate of  $p_{1\bar{c}}/p_{0\bar{c}}$  for  $\bar{c} = \bar{0}$ . We also give the direct effects cumulative survival difference 0.10(95% CI: -0.13, 0.33), which is an estimate of  $(1 - p_{1\bar{c}}) - (1 - p_{0\bar{c}})$  for  $\bar{c} = \bar{0}$ . Table 3 gives the estimated direct effect of assignment to the diaphragm arm vs. control arm, for condom use set to “yes” at all 8 visits, using the same format as Table 3. All confidence intervals were calculated with the non-parametric bootstrap percentile method, using 10,000 iterations.

Though both point estimates are in the direction of assignment to the diaphragm arm being protective in these scenarios, neither is close to statistical significance. Also, based on the bootstrap estimates of standard errors, both of the direct effects relative risk point estimates are within one standard error of the value corresponding to no effect of assignment to the diaphragm arm. We therefore conclude that the direct effects analysis provides no evidence in support of, nor in refutation of, diaphragms providing a protective effect against HIV infection.

As a sensitivity analysis to determine if the large confidence intervals in Tables 3 and 4 were a result of our stabilized weights, we ran our time-dependent analysis with no weights at all (that is, all weights = 1). The confidence intervals were still quite large; for example, the 95% confidence intervals for the direct effects relative risks were (0.348, 2.12) for condom use set to “never” and (0.55, 1.18) for condom use set to “always.” We conclude that the large confidence intervals are not due primarily to the stabilized weights we used.

## 8. Discussion

The analysis above, due to large confidence intervals, does not provide information as to whether providing latex diaphragm and gel is protective against HIV, at the fixed levels of condom use we considered. At first glance, this seems irreconcilable with the following

**Table 3.** ESTIMATES OF THE DIRECT EFFECT OF ASSIGNMENT TO STUDY ARM ON HIV INFECTION USING A TIME-DEPENDENT ANALYSIS, FOR CONDOM USE SET TO “No” FOR ALL VISITS

Relative Risk of HIV Infection by Visit 8	95% CI	Cumulative Survival Difference through Visit 8	95% CI
0.59	(0.26, 4.56)	0.10	(-0.13, 0.33)

**Table 4.** ESTIMATES OF THE DIRECT EFFECT OF ASSIGNMENT TO STUDY ARM ON HIV INFECTION USING A TIME-DEPENDENT ANALYSIS, FOR CONDOM USE SET TO “Yes” FOR ALL VISITS

Relative Risk of HIV Infection by Visit 8	95% CI	Cumulative Survival Difference through Visit 8	95% CI
0.96	(0.59, 1.45)	0.00	(-0.03, 0.04)

facts: (i) infection rates are similar in the two arms, (ii) reported condom use is much lower in the diaphragm arm compared to the control arm, and (iii) the data is consistent with condom use at last sex (as measured by self-report) protecting against HIV infection. However, the direct effects analysis and these facts are perfectly consistent with each other; the illusory contradiction disappears once we consider more closely how many infections the differential condom use would likely have prevented, compared to the difference in HIV infections between arms required to achieve statistical significance in a trial of this size.

We present a simple, crude adjustment for the differential reported condom use of 53.5% in the diaphragm arm and 85% in the control arm. We emphasize that this crude analysis is highly susceptible to confounding (which is why we did the direct effects analysis), but it sheds light on the apparent contradiction in the previous paragraph. We estimate, quite roughly, the number of infections that would have been prevented had condom use been increased in the diaphragm arm to the level observed in the control arm. That is, we estimate the number of infections that would have been prevented, had an additional  $(85\% - 53.5\%) = 31.6\%$  of the 2472 subjects in the diaphragm arm used condoms consistently; 31.6% of 2472 subjects is about 781 subjects. First, we calculate how many infections we would expect among 781 randomly chosen subjects among those in the diaphragm arm who never used condoms. Based on the observed infection rate in the diaphragm arm among subject-visits in which women reported no condom use at last sex (4.2% new infections per woman-year), and the mean length of time subjects were followed-up (1.6 years), we would expect  $(781 \times 1.6) \text{ woman-years} \times 4.2\% \text{ new infections per woman-year} \approx 51$  of these 781 subjects to become infected with HIV by the end of the trial. Next, we estimate how many of these 51 infections would have been prevented had these 51 subjects instead used condoms consistently. Assuming that reported condom use offers  $(100\% - 54\%) = 46\%$

protection against HIV infection, as was estimated from the data in the diaphragm arm and explained in Section 7 above, a crude estimate of the number of infections that would have been prevented is 46% of  $51 \approx 23$  infections prevented. However, this is much lower than the number of prevented infections required to achieve a statistically significant difference between the two study arms, at significance level  $\alpha = 0.05$  (two-tailed); given that there were 151 infections in the control arm and 158 infections in the diaphragm arm, the required difference in infections is 39. Thus, based on this crude analysis, adjusting for the number of infections attributed to the observed difference in reported condom use between arms does not lead to a statistically significant difference in infections between the arms. The direct effects analysis can be viewed as an attempt to more carefully carry out the above crude analysis, adequately dealing with confounders and the time-dependent nature of the data.

The direct effects estimator we used, based on the inverse probability of treatment weighted (IPTW) estimator, could be biased if (1) there are unmeasured confounders (e.g. characteristics of male partners associated with their condom use and HIV status); (2) there is measurement error in reported condom use (for example, due to social desirability bias, or if quarterly reported condom use at last sex is not sufficiently informative about overall condom use) or measurement error in confounders; (3) the models for condom use or hazard of HIV infection are not correctly specified; (4) missing data values are very different from observed values; (5) the experimental treatment assignment assumption is violated (see Section 6.4); or (6) the consistency assumption or time-ordering assumption is violated. All of these biases may exist to some degree, with (1) and (2) being of particular concern in the context of the MIRA trial. (1) is important since partners, who are believed to control condom use much of the time, may be associated with important unmeasured confounding. (2) is important since condom use by self-report is likely to be inflated upward, and this is consistent with our causal analysis of condom use in Section 7. We treat the point estimates for direct effects relative risks with a good deal of scepticism for these reasons.

However, the wide confidence intervals in the direct effects analysis allow us to make the following valuable conclusion: The data in the MIRA trial do not contain enough information to determine whether differential condom use may have masked a possibly protective effect of study arm assignment on HIV infection. This was not evident from a first look at the intention-to-treat results and the large difference in reported condom use, which on their face seemed, to the contrary, to imply that there might be enough information in the MIRA trial data to adjust for condom use and show a positive effect of diaphragms. The crude analysis at the beginning of this section, though suggestive, does not account for confounding or the time-dependent nature of the data; an analysis taking these into account (such as our direct effects analysis) was required to determine if there was enough information in the MIRA trial data to adjust for condom use and possibly uncover a positive effect of diaphragms. Without such an analysis, readers of the Lancet article reporting only the MIRA trial intention-to-treat results and differential condom use (Padian et al., 2007) may continue to wonder whether there was evidence in the trial data for a benefit of the diaphragm intervention.

We briefly discuss some alternative analyses that could be done to answer some important public health questions related to the MIRA trial. For example, one could try to estimate diaphragm efficacy at a fixed level of condom use, that is, the effect of diaphragm use  $D$  on HIV status  $I$ , fixing condom use  $C$ . We expect this would differ from what we estimated using direct effects, which involved the effect of being assigned to the diaphragm arm. An analysis of the effect of diaphragm use  $D$  on  $I$ , fixing condom use  $C$  would require

additional assumptions than our direct effects analysis, since one would have to deal with time-dependent confounding of the effect of diaphragm use on HIV status. This highlights an advantage of doing a direct effects analysis of treatment assignment in a randomized trial, as we did in this paper: there cannot be confounders of (randomized) treatment assignment and primary outcome.

Another type of causal analysis that could be applied to the MIRA trial is principal stratification (Frangakis and Rubin, 2002). It could be used, for example to estimate the effect of study arm assignment for just the group of women who would not have used condoms regardless of which study arm they had been assigned to. This parameter may shed light on the important public health question of whether a diaphragm intervention would provide protection against HIV for women who cannot get their partners to consistently use male condoms, or sustain high levels of condom use over time. A different set of assumptions than for the direct effects analysis would be necessary to estimate this parameter (Frangakis and Rubin, 2002); also,  $R$  cannot be assumed to be an instrumental variable, since  $R$  potentially influences  $I$  through the separate causal pathway involving  $C$ .

The direct effects analysis can also be applied in trials with blinding, to try to remove distortion caused by a secondary intervention given to both arms. For example, consider microbicide trials in which a placebo (an inactive gel) is used, and participants cannot distinguish whether they are using the active microbicide or placebo, as discussed in Mantell et al. (2005) and Trussell and Dominik (2005). In a truly blinded trial, assignment to different study arms cannot affect the outcome through systematically modifying subjects' risk behaviours (such as condom use). But even in blinded trials, secondary interventions can affect the intention-to-treat estimator, pushing it either toward or away from the efficacy of the study product, as shown in Trussell and Dominik (2005). A direct effects analysis can remove some of this distortion caused by a secondary intervention, if this secondary intervention influences the outcome only through a measured, potentially controllable behaviour (e.g. condom use). A direct effects analysis produces an estimate of what the effect of treatment assignment would have been, had subjects been constrained in this behaviour. This information could be useful, for example, in a blinded microbicide trial with a strongly positive intention-to-treat result; here, a direct effects analysis could try to uncover whether intervention effectiveness might be very different in a setting without condom counselling in which condom use were quite low. Similarly, in a blinded microbicide trial that resulted in a null intention-to-treat result, the direct effects analysis could elucidate whether condom counselling could have pushed the intention-to-treat result far away from actual microbicide efficacy.

Changes to the MIRA trial design could help alleviate some of the difficulties discussed in this paper. For example, more resources could be put into recruiting participants whose partners refuse to use condoms. This is the population for which the diaphragm intervention was initially conceived, but in practice is difficult to recruit. Such a population would have low condom use in both arms in the presence of intensive condom counselling, obviating some of the problems discussed in this paper. To identify such subjects, recruiters could screen for women having covariates that in the MIRA trial were highly predictive of low condom use. There is much need for trial designs that are better able to target efficacy and effectiveness of new HIV prevention methods.

## 9. Appendix: Example of Intention-to-Treat Estimator Severely Underestimating Efficacy of Primary Treatment in a Blinded Trial, Due to a Secondary Intervention

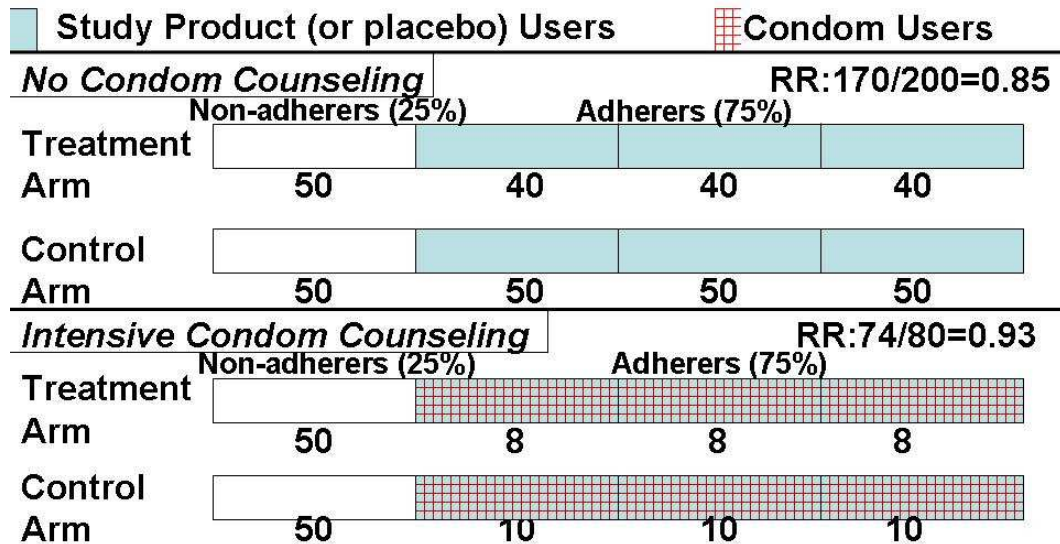
Here we give an example illustrating how a study product could be less effective in an environment of intensive condom counselling compared to an environment of no condom counselling, in a blinded trial with relatively high adherence (75%) to the assigned product. This example is derived from Trussell and Dominik (2005), who gives very similar examples. We will use the example of a microbicide being the study product, where the control arm gets a placebo (an inactive gel). Since the trial is blinded, subjects do not know which arm they are in. We use the term “assigned product” to denote the microbicide for subjects in the treatment arm, and the inactive gel for those in the control arm.

Consider the following two hypothetical randomized trials: The first trial is carried out in an environment of no condom counselling, and the second in the presence of intensive condom counselling. For simplicity, we make the following assumptions:

- (a) In both arms of both hypothetical trials, 75% of the subjects adhere to the assigned product.
- (b) The microbicide offers 20% relative protection against HIV infection for all subjects who use it—that is, the chance of infection for someone using the microbicide is  $(100\% - 20\% = 80\%)$  of what it would have been had they not used the microbicide; the placebo (inactive gel) gives no protection.
- (c) Condoms offer 80% relative protection against HIV (independent of microbicide use).
- (d) In the first trial (with no condom counselling) no one uses condoms, and in the second trial (with intensive condom counselling) all and only the adherers to assigned product (microbicide or placebo) also use condoms.

This information is depicted in the figure below, where the set of subjects in each arm is broken into fourths, each represented by a rectangle. Shading represents subjects who adhere to the assigned product, and hashing represents subjects who use condoms. (Shading and hashing that overlap indicate subjects who use both the assigned product and condoms.) In the diagram, numbers below the rectangles indicate the number of infections corresponding to the particular group of subjects. We first arbitrarily set the number of infections for those represented by a single rectangle, and who use neither the assigned product nor condoms, to be 50. By assumption, there are 20% fewer infections (so 40 infections) among the same number of people using the microbicide. Also by assumption, there are 80% fewer infections (so 10 infections) among the same number of people when they use only condoms (or condoms plus inactive gel); there are  $20\% * 80\% * 50 = 8$  infections among the same number of people using both condoms and the microbicide. In the first hypothetical trial (that is, with no condom counselling), the relative risk of infection is then  $170/200 = 0.85$ . In the second hypothetical randomized trial (that is, with intensive condom counselling), the relative risk of HIV infection is  $RR=74/80 = 0.93$ . Thus, in our artificial example, the study product is about half as effective in the environment of intensive condom counselling as in the environment of no condom counselling. In a similar scenario, but in which intensive condom counselling influences primarily those who don't use the assigned product to use condoms, the study product would be more effective in an environment of intensive condom counselling.

**Assume Condoms 80% protective, Study Product 20% efficacy**



**References**

Andersen, P., O. Borgan, R. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. NY, USA: Springer.

Anderson, J., C. Rietmeijer, and R. Wilson (1998). Asking about condom use: is there a standard approach that should be adopted across surveys? *Annual meeting of the American Statistical Association, St. Louis*.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996, June). Identification of causal effects using instrumental variables. *J. Am. Statist. Ass.* 91(434), 444–455.

Auvert, B., D. Taljaard, E. Lagarde, J. Sobngwi-Tambekou, R. Sitta, and A. Puren (2005). Randomized, controlled intervention trial of male circumcision for reduction of hiv infection risk: the anrs 1265 trial. *PLoS Med* 2: e298.

Bailey, R. C., S. Moses, C. B. Parker, K. Agot, I. Maclean, J. N. Krieger, C. F. Williams, R. T. Campbell, and J. O. Ndinya-Achola (2007, February). Male circumcision for hiv prevention in young men in kisumu, kenya: a randomised controlled trial. *The Lancet* 369(9562), 643–656.

Dawid, A. P. (2000, June). Causal inference without counterfactuals. *J. Am. Statist. Ass.* 95(450), 407–424.



- Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. NY, USA: Chapman and Hall.
- Feinstein, A. R. (1985). *The architecture of clinical research*. Philadelphia: W.B. Saunders.
- Frangakis, C. and D. Rubin (2002). Principal stratification in causal inference. *Biometrics* 58, 21–29.
- Frangakis, C. E. and D. B. Rubin (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 86(2), 365–379.
- Friedman, L., C. Furberg, and D. DeMets (1998). *Fundamentals of Clinical Trials, 3rd Ed.* NY, USA: Springer.
- Gray, R. H., G. Kigozi, D. Serwadda, F. Makumbi, S. Watya, F. Nalugoda, N. Kiwanuka, L. H. Moulton, M. A. Chaudhary, M. Z. Chen, N. K. S. F. Wabwire-Mangen, M. C. Bacon, C. F. Williams, P. Opendi, S. J. Reynolds, O. Laeyendecker, T. C. Quinn, and M. J. Wawer (2007, February). Male circumcision for hiv prevention in men in rakai, uganda: a randomised trial. *The Lancet* 369(9562), 657–666.
- Hernan, M., B. Brumback, and J. Robins (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology* 11(5), 561–570.
- Hirano, K., G. W. Imbens, D. Rubin, and X.-H. Zhou (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics* 1(1), 69–88.
- Jewell, N. P. (2004). *Statistics for Epidemiology*. Boca Raton, Florida, USA: Chapman and Hall/CRC.
- Mantell, J., L. Myer, and A. C.-D. et al. (2005). Microbicide acceptability research: current approaches and future directions. *Soc Sci Med* 60, 319–330.
- Padian, N., A. van der Straten, G. Ramjee, T. Chipato, D. de Bruyn, K. Blanchard, S. Shiboski, E. Montgomery, H. Fancher, H. Cheng, M. Rosenblum, M. van der Laan, N. Jewell, J. McIntyre, and the Mira team (2007, July). Diaphragm and lubricant gel for prevention of hiv acquisition in southern african women: a randomised controlled trial. *The Lancet* 370(9583), 251–261.
- Pearl, J. (2000a). *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.
- Pearl, J. (2000b). Direct and indirect effects. in M. Kaufmann, editor, *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 411–420.
- Petersen, M., S. Sinisi, and M. van der Laan (2006). Estimation of direct causal effects. *Epidemiology* 17, 276–284.
- Robins, J. (1986). A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect. (with errata). *Mathematical Modelling* 7, 1393–1512.

- Robins, J. (1998). Marginal structural models. *1997 Proceedings of the American Statistical Association*, 1–10.
- Robins, J. and D. Finkelstein (2000). Correcting for non-compliance and dependent censoring in an aids clinical trial with inverse probability of censoring weighted (ipcw) log-rank tests. *Biometrics* 56(3), 779–788.
- Robins, J. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3, 143–155.
- Robins, J. M. (1997). *Causal Inference from Complex Longitudinal Data. Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics (120)*. M. Berkane, Editor. NY: Springer Verlag.
- Rosenbaum, P. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *J.R. Statist. Soc. A* 147(5), 656–666.
- Rosenblum, M., N. Jewell, M. van der Laan, S. Shiboski, A. van der Straten, and N. Padian (2007, September). Analyzing direct effects of treatments in randomized trials with secondary interventions: An application to hiv prevention trials. *U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 223*. <http://www.bepress.com/ucbbiostat/paper223>.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 688–701.
- Sheiner, L. B., , and D. B. Rubin (1995). Intention-to-treat analysis and the goals of clinical trials. *Clin Pharmacol Ther* 57, 6–15.
- Trussell, J. and R. Dominik (2005). Will microbicide trials yield unbiased estimates of microbicide efficacy? *Contraception* 72, 408–413.
- van der Laan, M. and M. Robins (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer-Verlag.
- van der Laan, M. and D. Rubin (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics. (Article 11)* Available at: <http://www.bepress.com/ijb/vol2/iss1/11> 2(1).
- Weller, S. C. and K. Davis-Beaty (2002). Condom effectiveness in reducing heterosexual hiv transmission. art. no.: Cd003255. doi: 10.1002/14651858.cd003255. *Cochrane Database of Systematic Reviews* (1).

