Published in final edited form as:

J Am Stat Assoc. 2012 March 1; 107(497): 52–65. doi:10.1080/01621459.2011.643743.

Partially hidden Markov model for time-varying principal stratification in HIV prevention trials

James Y. Dai,

Vaccine and Infectious Disease Institute Fred Hutchinson Cancer Research Center, Seattle, WA
98109 jdai@fhcrc.org

Peter B. Gilbert, and

Vaccine and Infectious Disease Institute Fred Hutchinson Cancer Research Center

Benoît R. Mâsse

CHU Sainte-Justine Research Center University of Montreal Vaccine and Infectious Disease
Institute Fred Hutchinson Cancer Research Center

Abstract

It is frequently of interest to estimate the intervention effect that adjusts for post-randomization variables in clinical trials. In the recently completed HPTN 035 trial, there is differential condom use between the three microbicide gel arms and the No Gel control arm, so that intention to treat (ITT) analyses only assess the net treatment effect that includes the indirect treatment effect mediated through differential condom use. Various statistical methods in causal inference have been developed to adjust for post-randomization variables. We extend the principal stratification framework to time-varying behavioral variables in HIV prevention trials with a time-to-event endpoint, using a partially hidden Markov model (pHMM). We formulate the causal estimand of interest, establish assumptions that enable identifiability of the causal parameters, and develop maximum likelihood methods for estimation. Application of our model on the HPTN 035 trial reveals an interesting pattern of prevention effectiveness among different condom-use principal strata.

Keywords

microbicide; causal inference; posttreatment variables; direct effect

1. INTRODUCTION

Strategies to curb and eventually eliminate the HIV/AIDS epidemic are of major interest to global public health. Ongoing efforts include vaccines, microbicides, antiretroviral medications, circumcision and behavioral risk reduction counseling. Topical microbicides offer a female initiated prevention method, particularly for women who cannot negotiate use of condoms by their male partners (Stone and Hitchcock 1994; Elias and Coggins 1996). The HIV Prevention Trials Network (HPTN) 035 trial was a randomized, four-arm, placebo-controlled, phase II/IIB trial to evaluate the safety and effectiveness of two vaginal microbicide products, BufferGel and PRO2000 Gel, in preventing acquisition of HIV infection among at-risk women (Fleming and Richardson 2004; Karim et al. 2011). Two control arms were designated: one arm with a placebo gel blinded with the other two active gels; one arm without any gel product, therefore unblinded. The rationale for a “No Gel” unblinded arm was partially driven by a lack of information on the potential protective effect of the placebo gel. One attribute of the placebo gel is that it may serve as a physical barrier

to HIV infection, even though there is no active biochemical ingredient (Fleming and Richardson 2004).

Primary intention to treat (ITT) analyses of the HPTN 035 trial suggest that one of the two active gels, the PRO2000 Gel, reduces the risk of HIV infection by 33% when compared to the No Gel arm (p -value=0.06), by 30% when compared to the Placebo Gel arm (p -value=0.1). The other active product, the BufferGel, does not provide protection as compared to either the No Gel arm or the Placebo Gel arm (Karim et al. 2011). HIV incidence was similar in the No Gel arm and the Placebo Gel arm (hazard ratio=0.97; p -value=0.89).

An interesting observation is that women in the No Gel control arm reported a higher rate of condom use than women in the three gel arms, even though women in all four arms were provided condoms and condom promotion counseling. Figure 1 shows the rates of quarterly self-reported condom use in the last sex act in the largest site of HPTN 035 - Durban, South Africa. A consistent 10-15% difference in reported condom use was observed at each quarterly visit after enrollment in the No Gel arm as compared to the three gel arms (p -value < 0.001). No significant differences were found among the three gel arms. This pattern is consistent across all seven sites. One explanation is behavioral disinhibition, or risk compensation (Cassel et al. 2006); that is, couples may have increased their sexual risk taking, e.g., unprotected sex, when they perceive that female partners have some protection by using gel products even though the efficacy has not been established.

Since consistent and correct use of condom offers over 90% protection against HIV infection (Centers for Disease Control 1993), any reduction of condom use can offset protection attributable to an effective vaginal microbicide. As such, null results from the ITT comparison between one of the gel arms and the No Gel arm may not rule out the biologic efficacy of the gel. Certainly the ITT results address the net effect of the gel assignment on the acquisition of HIV infection; however, it is also of interest to adjust the intervention effect for the differential condom use. Specifically, since condom use is an intermediate variable in the causal pathway from the treatment assignment to HIV infection, we are interested in the direct effect of treatment assignment on HIV infection, not mediated through differential condom use.

Risk compensation has been considered as the Achilles' heel in innovations in HIV prevention, e.g., circumcision, microbicides, and pre-exposure prophylaxis (Cassel et al. 2006), that potentially jeopardises the benefits of these new methods. For any HIV prevention trial to evaluate a new intervention strategy, condom counselling and provision are provided to all participants as a standard, secondary intervention. If the treatment is unblinded to participants, e.g., the HPTN 035 trial, participants may adapt the condom-uptake behaviour upon their perception of protection from infection. Similar phenomena has also been reported in the recent MIRA (Methods for Improving Reproductive health in Africa) trial (Rosenblum et al. 2009). On top of the typical compliance issue, that is, the participants may not consistently use the new prevention method, risk compensation further complicates the analysis on the efficacy of the new intervention.

The fundamental statistical problem arising in this example is how to properly adjust for post-randomization variables, since, if the treatment has an effect on post-randomization variables, standard adjustment by regression modelling is susceptible to selection bias (Rosenbaum 1984). The counterfactual or potential outcome framework, first introduced in randomized experiments by Neyman (1923) and later generalized to observational studies (Rubin 1978; Holland 1986), has been increasingly used to draw inference conditional on post-randomization variables that are intermediates in a causal pathway. The central

property of this framework is that causal inference is based on probability distributions of potential outcomes, which are separated from the treatment assignment mechanism. One way to estimate direct effects in causal inference requires a hypothetical intervention on the intermediate variable so that it can be “forced” to a fixed level (Robins and Greenland 1992; Petersen et al. 2006). Such intervention is feasible for some biological variables, that are well understood and therefore can be controlled by investigators, for example hyperlipidemia which can be controlled by diet or drugs (Robins and Greenland 1992). Identifiability of direct effects relies on the assumption that there are no unmeasured confounders between the intermediate variable and the outcome, conditional on the observed covariates, also known as “sequential ignorability”. Joffe and Greene (2009) referred to this framework as the “Causal-Effects” paradigm, as it models the causal effect of the intermediate variable on the outcome. An array of estimation procedures have been developed for intermediate variables measured either at a single post-randomization time point, or longitudinally (Robins, Hernan and Brumback 2000; Robins, Rotnitzky and Scharfstein 2000; van der Laan and Robins 2003). Notably, direct effects based on Inverse Probability of Treatment Weighted (IPTW) estimators have been presented in the analysis of the MIRA trial (Rosenblum et al. 2009).

In HIV prevention trials, however, we argue that application of the “Causal-Effects” paradigm to estimating causal treatment effect adjusting for post-randomization behavioral variables may not be sensible. Contrary to intermediates of biological or experimental nature, the ability to manipulate behavioral intermediates, such as condom use, is limited. Indeed, the precise motivation for developing new prevention strategies for different risk populations, for instance, microbicides for women, is that consistent condom use can be counseled, but can not be “forced” or “manipulated”, even though it is well known that condom use is very effective in preventing HIV infection. Therefore the interpretation of direct effects as “the treatment effect in a hypothetical trial where everyone uses or does not use condoms” is not realistic. Moreover, the assumption of no unmeasured confounders of condom use and HIV infection is not very tenable. Predictors of behavioral patterns, for example sexual risk taking, are not well understood by prevailing behavioral theories (Halperin 2007). Though we often collect a number of baseline demographic and behavioral covariates, few post-randomization covariates other than adherence and sexual behaviors are routinely collected. In the absence of post-randomization predictors, it is overreaching to assume the behavioral pattern is a “statistically exogenous or ancillary” process (Hernan et al. 2001).

Alternatively, causal inference adjusting for post-randomization variables can be made within population-inherent strata defined by the joint potential outcomes of the intermediate variable under both treatment assignments, namely principal stratification (Frangakis and Rubin 2002; Rubin 2004). Joffe and Greene (2009) called this framework the “Causal-Association” paradigm, as it does not address the causal effect of the intermediate variable on the outcome. Note that principal strata are not observed, hence various assumptions have to be exploited to identify the treatment effect within principal strata (Angrist et al. 1996; Imbens and Rubin 1997), or sensitivity analyses can be employed to examine the causal effect given a plausible range of the sensitivity parameter (Gilbert et al. 2003).

In the setting of HIV prevention, we believe principal stratification is more interpretable and useful because there appear to be subgroups existing in the study population with different risk behaviors; for example, women who always use condoms, and women who never use condoms, regardless of which treatment assignment they received. Estimating causal effect in these subgroups provides a natural way to adjust for differential condom use. In particular, the motivation for developing microbicides is to protect women who could not negotiate with their partner to use condoms, therefore estimating treatment effects within the

subgroup with low-condom users addresses the key scientific interest. The methodological challenge, however, is how to handle the longitudinal post-treatment variables on condom use.

There have been extensions of the principal stratification framework to trials with longitudinal treatment compliance, though the settings are quite different from ours. Frangakis et al. (2004) considered studies with a time-varying treatment, time-varying compliance, and a time-to-event outcome. The identifiability of principal causal effect relies on several assumptions, e.g., multilevel monotonicity and compound exclusion restriction. For trials with randomization administered at enrollment and longitudinal continuous outcomes, a nested latent class model was proposed to account for time-varying compliance (Lin et al. 2008; Lin et al. 2009). Longitudinal principal strata on compliance were collapsed to compliance superclasses, within which principal effects were defined and estimated by the Bayesian Markov chain Monte Carlo (MCMC) method. Our setting differs from those papers in that the endpoint is time-to-event, which is common in HIV prevention trials, and that randomization to an intervention is fixed at enrollment.

In this article, we develop a principal stratification framework for disentangling prevention effect from risk compensation, i.e., the differential condom use induced by the gel assignment in the HPTN 035 trial. Inspection of data leads us to a partially hidden Markov model (pHMM) for time-varying principal strata on condom use, and the time-to-infection endpoint in HIV prevention trials. A hidden Markov model (HMM) is a doubly stochastic process in which a hidden Markov chain dictates the probability distributions of a sequence of observed random variables. The rich structure imposed by the HMM facilitates identifiability and likelihood maximization.

HMMs have during the last two decades been widely used in areas such as speech recognition (Rabiner 1989), biology (Leroux 1992), genetic linkage mapping (Lander and Green 1987), and microarray differential expression (Yuan and Kendziorski 2006). In randomized clinical trials, HMMs have recently been used to model weakly dependent health outcomes, such as movement disorder (Scott et al. 2005), sclerosis (Altman and Petkau 2005), and alcoholism (Shirley et al. 2010). Differing from a standard HMM, the Markov chain we propose, essentially principal strata on condom use, is partially observed because the condom use in the assigned study arm is observed, but not the condom use were the participant assigned to the other arm. The observable process, the HIV infection status at each study visit, depends on the condom-use state at that visit and the treatment assignment, assuming a transient behavioral effect on HIV infection. We formulate two estimands, cumulative risk ratio and discrete hazard ratio, that describe the cumulative and the transient treatment effect respectively.

The remainder of this article is organized as follows. In Section 2, we define the notation, the pHMM model and the estimands. We discuss the assumptions that are required to identify the estimands. The proof of identifiability and the modified Baum-Welch algorithm for maximizing the likelihood are put in the Appendix. In Section 3, we evaluate the small sample properties of the proposed estimators. In Section 4, we present the results of applying the pHMM to the HPTN 035 trial. We then proceed to discussion and future work in Section 5.

2. METHODOLOGY

2.1 Notation

We use the HPTN 035 trial as an example for our notation. A total of n HIV-seronegative subjects were randomly assigned to either the gel (treatment) arm or the No Gel (control)

arm. Let $Z_i = 0$ if subject i , $i = 1, \dots, n$, is assigned to the No Gel arm and $Z_i = 1$ if subject i is assigned to a gel arm. After enrollment participants are evaluated regularly at a series of discrete times (i.e., quarterly visits), denoted as a_j , $j = (1, 2, \dots, J)$, with evaluations including interviewer-administered sexual behavior questionnaires and HIV testing performed each quarter. We consider an indicator variable for self-reported recent condom use, X_{ij} , and let $\bar{X}_{ij} = \{X_{i1}, \dots, X_{ij}\}$ be the history of condom use up to time a_j . With this notation, we suppose only two possible level of condom use. The study endpoint is time to HIV detection, therefore subjects exit the follow-up window once they are detected HIV positive or are censored due to loss or end of follow-up. Let T_i denote the time when subject i tests HIV positive, and let C_i denote the time when subject i is censored due to loss or end of follow-up. Both T_i and C_i have discrete distributions with mass points at $0 < a_1 < \dots < a_J$. We observe $Y_i = \min(T_i, C_i)$ and $\Delta_i = 1_{[Y_i = T_i]}$. Let J_i denote the last visit for the i^{th} participant, $J_i = j: Y_i = T_i$. X_{ij} is only meaningful and therefore observed up to J_i . In addition, a number of baseline demographic and behavioral variables are collected, denoted as a vector of random variables, W_i . The observed data are therefore, i.i.d.

$$(Z_i, \bar{X}_{ij}, Y_i, \Delta_i, W_i).$$

We use the potential outcomes framework to define the estimands of interest. Let $\mathbf{Z}$ be a n -dimensional vector of assignments with i^{th} element Z_i , and let

$\bar{X}_{ij}(\mathbf{Z}) = \{X_{ij}(\mathbf{Z}), a_j = a_1, \dots, a_J\}$ be a vector of counterfactual indicators of whether subject i used a condom for the last sex act at the time a_j if participants are given the randomly assigned treatment vector $\mathbf{Z}$. We first consider the hypothetical scenario where there is no HIV testing, no drop-out, and no censoring, so that participants would have been followed for behavioral assessments in all J scheduled visits. Then $\bar{\mathbf{X}}(\mathbf{Z})$ is the $n \times J$ matrix of potential outcomes for condom use of n subjects in J visits, if the assignment vector is $\mathbf{Z}$.

Define $T_i(\mathbf{Z}, \bar{\mathbf{X}}(\mathbf{Z}))$ to be the time-to-event outcome given the assignment vector $\mathbf{Z}$, and

the condom use matrix $\bar{\mathbf{X}}(\mathbf{Z})$. The potential outcomes for $C_i(\mathbf{Z}, \bar{\mathbf{X}}(\mathbf{Z}))$, $Y_i(\mathbf{Z}, \bar{\mathbf{X}}(\mathbf{Z}))$,

and $\Delta_i(\mathbf{Z}, \bar{\mathbf{X}}(\mathbf{Z}))$ are similarly defined. We assume that the potential outcomes for each individual i are independent of the treatment assignments of other individuals, i.e., there is no interference between units (Cox 1958). This assumption is plausible in the HPTN 035 trial since the number of women enrolled in each site is relatively small compared to the size of local community, and most of the women enrolled are in a relatively stable partnership. Formally this assumption is expressed as follows.

Assumption 1: Stable unit treatment value assumption (SUTVA)—(Rubin 1978; Angrist et al. 1996). For any two treatment assignments, $\mathbf{Z} = (Z_1, \dots, Z_n)$ and $\mathbf{Z}' = (Z'_1, \dots,$

$Z'_n)$, the following hold: a. If $Z_i = Z'_i$, then $\bar{X}_{ij}(\mathbf{Z}) = \bar{X}_{ij}(\mathbf{Z}')$; b. If $Z_i = Z'_i$, $\bar{X}_{ij} = \bar{X}'_{ij}$, then

$$T_i(\mathbf{Z}, \bar{\mathbf{X}}) = T_i(\mathbf{Z}', \bar{\mathbf{X}}'), C_i(\mathbf{Z}, \bar{\mathbf{X}}) = C_i(\mathbf{Z}', \bar{\mathbf{X}}'), Y_i(\mathbf{Z}, \bar{\mathbf{X}}) = Y_i(\mathbf{Z}', \bar{\mathbf{X}}'), \Delta_i(\mathbf{Z}, \bar{\mathbf{X}}) = \Delta_i(\mathbf{Z}', \bar{\mathbf{X}}')$$

. The SUTVA assumption allows us simplify the notations and write potential outcomes of interest, $\bar{X}_{ij}(z)$, $T_i(z)$, $C_i(z)$, $Y_i(z)$, $\Delta_i(z)$ for $z = 0, 1$.

For a post-randomization variable measured at a single time point, say X , Frangakis and Rubin (2002) defined the *basic principal stratification* P_0 as the partition of units such that within any set of P_0 , all units have the same vector $(X(1), X(0))$. A *principal stratification* P is a partition of the units whose sets are unions of sets in P_0 . For a vector of time-varying

post-randomization variables such as $\bar{X}_{ij}$, the condom use in the last sex act at each quarterly visit, principal strata are defined upon the sets in which units have the same $2 \times J$ matrix of potential condom use under either assignment, $\bar{U}_{ij}$,

$$\bar{U}_{ij} = \begin{pmatrix} X_{i1}(0) & X_{i2}(0) & \cdots & X_{ij}(0) \\ X_{i1}(1) & X_{i2}(1) & \cdots & X_{ij}(1) \end{pmatrix}.$$

Let $U_{ij} = (X_{ij}(0), X_{ij}(1))$, for easy exposition $\bar{U}_{ij}$ is reduced to a vector of U_{ij} . U_{ij} takes the following values,

$$U_{ij} = \begin{cases} \text{N} & \text{condom use under neither assignment} & \text{if } X_{ij}(0) = X_{ij}(1) = 0, \\ \text{C} & \text{condom use under control only} & \text{if } X_{ij}(0) = 1, X_{ij}(1) = 0, \\ \text{B} & \text{condom use under both assignments} & \text{if } X_{ij}(0) = X_{ij}(1) = 1, \\ \text{T} & \text{condom use under treatment only} & \text{if } X_{ij}(0) = 0, X_{ij}(1) = 1, \end{cases}$$

Henceforth we call U_{ij} the “principal state”, with 4 possible states: N, C, B, and T defined above. Principal strata are formed by a sequence of principal states. These condom use principal states resemble those used to describe the compliance states in Imbens and Rubin (1997), and are of great interest to public health: there are men who always use condoms for protection regardless of whether their partners are using the potentially protective intervention (B), men who never use condoms regardless (N), and others who assess their risk of HIV acquisition and adapt their condom use accordingly (C and T). These risk groups may exhibit heterogeneity in HIV susceptibility and prevention efficacy. In particular, the second state C characterizes behavioral disinhibition or risk compensation/gamble.

The fourth state T refers to the behavior of men who use condoms when their female partners are assigned to the gel arm, but not when assigned to the condom-only arm. This state is highly unlikely given the data we observed in the trial and the HIV/AIDS literature on risk compensation. For instance, meta-analysis on antiretroviral treatment has shown that those who perceived that treatment might reduce the risk of HIV transmission had significantly higher rates risky sex (Crepaz et al. 2004). There is not yet empirical evidence to suggest people decrease risk-taking when they receive preventative treatment. Henceforth we assume that the fourth stratum is empty. Formally this can be expressed as:

Assumption 2: Monotonicity of condom use— $X_{ij}(1) \geq X_{ij}(0)$ for all i, j .

The consequence of assuming condom use monotonicity is that the condom use principal state is unambiguous for women who reported no condom use in the last sex act in the No Gel arm, and for women who reported condom use in the gel arm (Table 1). Presumably the risk gambling, is driving the observed difference of rates of condom use between the gel arms and the No Gel arm.

Note that our model is built for a time-to-event outcome with a longitudinal post-randomization variable X_{ij} . X_{ij} is only of interest and observed when a participant remains in the risk set. The principal strata, however, is conceived as a pre-existing baseline characteristic that should not depend on the censored time-to-event outcome. Therefore we have defined U_{ij} , the latent process governing X_{ij} , for all visits $j = 1, \dots, J$, as if X_{ij} will be measured regardless whether an infection event or censoring has occurred. As we describe

later, the censored time-to-event outcome process acts upon the underlying condom-use principal states, stopping the condom-use process if an event or censoring has occurred.

Before we proceed, we explicitly state that because of randomization, the treatment assignment is independent of the joint distribution of potential post-randomization outcomes, though it may depend on baseline covariates, e.g. study sites. In addition, we will assume that the censoring is independent of time to HIV infection, and principal strata $\bar{U}_{ij}$, conditional on baseline covariates. We use the symbol $\perp$ to denote statistical independence; the conditional independence is denoted by $A \perp B|C$, for instance. These two assumptions are expressed as

Assumption 3: Randomization of assignments—

$$Z \perp (\bar{U}_{ij}, T_i(0), T_i(1), C_i(0), C_i(1), Y_i(0), Y_i(1), \Delta_i(0), \Delta_i(1)) | W_i$$

Assumption 4: Independent censoring—

$$(C_i(0), C_i(1)) \perp (\bar{U}_{ij}, T_i(0), T_i(1)) | W_i$$

The censoring mechanism in the HPTN 035 trial is, for most of participants, termination due to end of the study period. There are a few participants who were lost to follow-up. Thus the independent censoring is a quite reasonable assumption for this trial.

2.2 Causal estimands

The causal estimand of interest is the risk ratio of the counterfactual probabilities of having HIV infection before a_j under either treatment assignment, namely the cumulative risk ratio (CRR), for women who have a particular pattern of $\bar{U}_{ij}$. In particular, interest centers on women who would report consistent no condom-use (stratum $\bar{N}$), or women who would report consistent condom-use (stratum $\bar{B}$), up to the j^{th} visit (Table 1),

$$\text{CRR}(a_j | \bar{U}_{ij} = \bar{N}_j) = \frac{1 - \Pr(T_i(1) > a_j | \bar{U}_{ij} = \bar{N}_j)}{1 - \Pr(T_i(0) > a_j | \bar{U}_{ij} = \bar{N}_j)}, \quad (1)$$

$$\text{CRR}(a_j | \bar{U}_{ij} = \bar{B}_j) = \frac{1 - \Pr(T_i(1) > a_j | \bar{U}_{ij} = \bar{B}_j)}{1 - \Pr(T_i(0) > a_j | \bar{U}_{ij} = \bar{B}_j)}, \quad (2)$$

where $\bar{N}_j$ and $\bar{B}_j$ are vectors of N and B with length j . These estimands depict the effects of the treatment assignment directly on the HIV acquisition endpoint, disentangled from the effect mediated through the differential condom use, since within principal stratum N and B there should be no effect from condom use. For those who would have not used condom consistently under either assignment, one may expect to observe a larger treatment effect than the overall ITT effect. On the other hand, for those who would have consistently used condoms under both assignments, one may expect less protection effect from the gel treatment since effective condom use may mask the gel effect. For the third condom-use

state C, the contrast between $\Pr(T_i(1) > a_j | \bar{U}_{ij} = \bar{C}_j)$ and $\Pr(T_i(0) > a_j | \bar{U}_{ij} = \bar{C}_j)$ in risk gamblers is less interpretable as it mixes both the gel effect and the condom effect together. Note that throughout this paper, we focus on the ITT effect of the treatment assignment among the principal strata on condom use, not the true efficacy of the gel women actually used.

Difficulties arise when the causal estimand is based on a contrast of survival probabilities under two assignments given a longitudinal principal stratum. Since U_{ij} varies over time, and there are at most 10 visits in the HPTN 035 trial, there could be as many as 3^{10} principal strata on longitudinal condom use. The identifiability of these strata with longitudinal states becomes almost insurmountable without strong parametric assumptions. Even if they are identified, for any given stratum, there could be very few participants, if any, rendering estimation of survival probabilities instable. For trials with longitudinal, continuous outcomes and longitudinal compliance, Lin et al. (2008) address these challenges by collapsing longitudinal compliance patterns to a few superclass, and put estimation in a Bayesian framework. The procedure does not apply to our settings with a censored time-to-event outcome. Next we describe a partially hidden Markov model (pHMM) that exploits features of the condom-use pattern and the risk of HIV acquisition in the HPTN 035 trial and establishes the identifiability and the estimation.

2.3 The partially Hidden Markov model (pHMM)

The main difficulty in estimating CRR (1-2) is to identify the latent principal stratum $\bar{U}_{ij}$, as it is “partially” hidden: for a participant reporting condom-use in the no-gel arm at the j^{th} visit, she could be in the state C or in the state B; for a participant reporting no condom-use in the gel arms at the j^{th} visit, she could be in the state N or in the state C. The rest of principal states, the state N in the no-gel arm and the state B in the gel arm, are directly observed by the monotonicity assumption.

In the regularly monitored study and for a short study period such as the HPTN 035 trial, the observed condom use appears to be relatively stable (Figure 1). We hypothesize that conditional on the baseline covariate W_i , the principal state U_{ij} follows a first-order Markov chain with a homogeneous transition matrix. Assume that the marginal distribution of each U_{ij} is multinomial and the possible states (s) take value from a countable set $S = \{1, 2, 3\}$, representing three states N, C, and B respectively.

Assumption 5a: Markov property of the condom use states—

$$\Pr(U_{ij} = t | U_{ij-1} = s, U_{ij-2}, \dots, U_{i1}, W_j = w) = \Pr(U_{ij} = t | U_{ij-1} = s, W_j = w) = q_{st}(w) \quad (3)$$

for $z = 0, 1$ and $s, t \in S$, where $q_{st}(w)$ refers to the transition probability from s to t for a subject with a covariate value w . Note that the principal strata, $\bar{U}_{ij}$, at the subject level is deterministic. The probabilistic statement on Assumption 5a describes the Markov chain in the population level, that is, when there are many subjects with different $\bar{U}_{ij}$, the Markov property in Assumption 5a simplifies the expression of the joint probability of randomly sampled $\bar{U}_{ij}$.

The first-order Markov property is suggested by the fact that the condom use in a visit is mostly predicted by the condom use in the previous visit. The assumption of the homogeneous transition probabilities across different visits is critical, as it greatly reduces the number of parameters that are needed to describe the underlying state process, yet it is

not usually testable in HMM. In our pHMM, however, part of this 3×3 transition matrix can be computed because the state N in the no-gel arm and the state B in the gel are directly observed, which allows direct assessment of the homogeneity of the transition probabilities.

Figure 3 shows in the largest site - Durban, the two sets of transition probabilities from the j^{th} visit to the $j+1^{\text{th}}$ visit, $j=1, \dots, 9$, and their 95% confidence intervals. The probability of condom use at the $j+1^{\text{th}}$ visit given condom use at the j^{th} visit, computed based on all three gel arms, is fairly high, ranging 0.85 to 0.91. The differences between adjacent transition probabilities are fairly small. The transition probabilities from the N state to the N state in the no-gel arm is also fairly stable across visits, though the confidence intervals are much wider due to much smaller sample sizes compared to those in the gel arms. Taking collectively, these data suggest that at least in part of the transition matrix, there is good evidence of homogeneous transition probabilities.

The outcome process in our pHMM is the indicator of being infected at the j^{th} visit or not. We hypothesize that conditional on baseline covariates W_i , the hazard of being infected at the j^{th} visit is governed by U_{ij} but not the previous principal state.

Assumption 5b: Transient effect of condom use on the hazard—

$$\Pr(T_i(z)=a_j | T_i(z) \geq a_j, U_{ij}=s, \dots, U_{i1}, W_i=w) = \Pr(T_i(z)=a_j | T_i(z) \geq a_j, U_{ij}=s, W_i=w) \\ = \lambda(U_{ij}=s, z, w),$$

where $\lambda(s, z, w)$ is the discrete hazard of acquiring HIV infection at a_k given the condom-use state $U_{ik}=s$, the treatment assignment z and the covariate $W_i=w$. This assumption stipulates that the effect of the condom-use principal strata is “transient”, that is, the risk of HIV acquisition is largely determined by the recent sexual risk-taking. In the context of HIV prevention trials, this is a biologically sensible assumption, since participants return to the clinic quarterly to test HIV infection status and report recent condom use, and the current HIV assay is quite sensitive to determine a recent HIV acquisition event. For instance, a common way to adjust for post-randomization compliance or condom-use is to put recent compliance or condom-use in a Cox regression model as time-varying covariates. Examples can be found in recent papers on several major HIV prevention trials, e.g., compliance in the HPTN 039 trial (Celum et al. 2008) and the iPrEx trial (Grant et al. 2010), condom-use in the HPTN 039 trial (Reid et al. 2010). This assumption allows us to decompose the event-free probability into the product of discrete hazards given U_{ik} and W_i at each visit:

$$\Pr(T_i(Z) > a_j | \bar{U}_{ij}, W_i) = \prod_{k=1}^j [1 - \Pr(T_i(Z) = a_k | T_i(Z) \geq a_k, U_{ik}, W_i)] = \prod_{k=1}^j [1 - \lambda(U_{ik}, z, w)]. \quad (4)$$

To this end, it is useful to quantify the transient treatment effect conditional on a condom-use principal state, for instance, the ratio $\lambda(U_{ij}=s, z=1, w)/\lambda(U_{ij}=s, z=0, w)$. More generally, for a monotone-increasing, twice-differentiable function $h(\cdot)$, we define the estimand β_s as

$$\beta_s = h(\lambda(Z_i=1, U_{ij}=s, W_i=w)) - h(\lambda(Z_i=0, U_{ij}=s, W_i=w)). \quad (5)$$

Choices of h include log, complimentary log-log, or logit (Kalbfleisch and Prentice 2002). We choose the logit function henceforth because of its easy estimation, which is

implemented essentially by fitting a logistic regression to subjects remaining at risk for each discrete time. We call this estimand “the discrete hazard ratio” (DHR).

Note that DHR (5) is not a causal estimand. The hazard, $\Pr(T_i(z) = a_j | T_i(z) \neq a_j, U_{ij}, W_i = w)$, is not conditional on the same set of potential outcomes, i.e., we may compare different populations who survived up to a_j under two assignments. This is different from causal hazard ratios formulated under marginal structural models, as the latter assume sequential randomization at each time point given the history (Hernan et al. 2001), whereas we assume randomization happens at enrollment. Nevertheless, DHR is still a useful estimand in that it captures the marginal contrast of hazards under either treatment assignment, similar to the hazard ratio formulated in a Cox proportional hazard model.

In summary, Figure 2 illustrates the pHMM we have thus derived. Our model differs from a general HMM in three aspects. Firstly, there are two observable processes instead of one outcome process in a typical HMM. U_{ij} is a sequence of hidden Markov states, and $1_{\{Y_i=a_j\}}$ is the observed outcome process. Together with the treatment assignment Z_i , U_{ij} dictates the X_{ij} deterministically and the HIV infection outcome probabilistically. Secondly, because of the monotonicity assumption, U_{ij} can be partially observed from X_{ij} . Thirdly, the outcome is censored time-to-event so that the Markov chain stops once there is HIV infection or censoring.

2.4 Likelihood, identifiability and maximization

Typical parameterization of HMMs includes three elements: the probabilities of the initial states; the transition probabilities; and the emission probabilities of the outcome process (Baum and Petrie 1966). For the pHMM, we have defined the transition probabilities $q_s(w)$, the emission probabilities $\lambda(U_{ij}, z, w)$, and the time-invariant hazard ratio $\beta_s(w)$. We now denote the initiation probabilities for the hidden state process,

$$\Pr(U_{i1}=s|W_i=w)=p_s(w), \quad s \in \mathcal{S} \quad (6)$$

Because of the randomization, the distribution of U_{i1} at the first visit is balanced for both arms, so that

$$\Pr(U_{i1}=s|Z_i=1, W_i=w)=\Pr(U_{i1}=s|Z_i=0, W_i=w)=p_s(w).$$

Let $\lambda_s(w)$ denote the hazard for a participant in the No Gel arm. Parameterization at visit a_j is shown in Table 1. For the parameters defined in the set $\Theta = \{p_s(w), q_s(w), \lambda_s(w), \beta_s\}$, we

derive the likelihood for the observed data, $(Y_i, \Delta_i, \bar{X}_{i,i}|Z_i, W_i)$, as follows:

$$\begin{aligned} & \prod_{i=1}^n \sum_{\bar{U}_{i,i} \in \bar{X}_{i,i}} \Pr(Y_i, \Delta_i, \bar{U}_{i,i}|Z_i, W_i), \\ & \propto \prod_{i=1}^n \left\{ \sum_{\bar{U}_{i,i} \in \bar{X}_{i,i}} \left[\prod_{j=1}^{J_i-1} ((1 - \lambda(U_{ij}, Z_i, W_i)) \lambda(U_{i,i}, Z_i, W_i)) \right]^{\Delta_i} \left[\prod_{j=1}^{J_i} (1 - \lambda(U_{ij}, Z_i, W_i)) \right]^{1-\Delta_i} p_{U_{i1}}(W_i) \prod_{j=2}^{J_i} q_{U_{ij}U_{ij-1}}(W_i) \right\} \end{aligned}$$

where $\bar{U}_{i,i} \in \bar{X}_{i,i}$ denotes the sequence of possible condom use states that conform to the observed history of condom use according to Table 1; $\lambda(U_{ij}, Z_i=0, W_i)$ equals $\lambda_s(w)$ when $Z_i=0$, and equals $h^{-1}(h(\lambda_s(w)) + \beta_s)$ when $Z_i=1$.

The identifiability of parameters in this likelihood is proved in the Appendix. Since there are two observable processes in our pHMM, namely the condom-use process and the censored infection process, the strategy of proving identifiability is to establish the one-to-one mapping from the parameters in the pHMM and the parameters in the two observable processes. For a single hidden Markov chain, the consistency and the asymptotic normality of the maximum likelihood estimators have been established when the number of time points go to infinity (Leroux 1992; Bickel et al. 1998). In our hidden Markov model, however, the sample size increases while the number of time points (visits) remains fixed, the usual large sample theory of the maximum likelihood estimators applies with the standard regularity condition (Cox and Hinkley 1974, chapter 9).

For the initiation probabilities $p_s(w)$ and the transition probabilities $q_{st}(w)$, we fit a polychotomous (multinomial) regression of the current state on the previous state if available, plus baseline variables including site, age, participant's education level, marital status, owning income or not, partner's education level and the number of vaginal sex acts in the last week. For estimating the emission probabilities, we fit a logistic regression of the HIV event at each visit on the condom-use principal state, the treatment assignment, and their interactions plus a number of baseline variables including site, age, participant's education level, marital status, owning income or not, partner's education level, ever having anal sex and the number of vaginal sex acts in the last week. The estimated survival probability for a subject with $\bar{U}_{ij}$ is computed as follows, from which we can compute the CRR estimand,

$$\Pr\left(T_i(Z) > a_j | \bar{U}_{ij}\right) = \sum_i \Pr\left(T_i(Z) > a_j | \bar{U}_{ij}, W_i\right) \frac{\Pr\left(\bar{U}_{ij} | W_i\right)}{\sum_i \Pr\left(\bar{U}_{ij} | W_i\right)},$$

where $\Pr\left(T_i(Z) > a_j | \bar{U}_{ij}, W_i\right)$ is computed by the estimated model parameters.

We adapt the Baum-Welch algorithm to maximize the likelihood. Computation of the observed data likelihood involves summing over a sequence of partially hidden states. The Baum-Welch algorithm exploits the Markov property of the likelihood, and uses the forward-backward procedure to efficiently sum over the sequence of hidden states. The unobserved states are handled as missing data, so the algorithm iterates between an E-step and a M-step until convergence (Baum and Petrie 1966; Baum et al. 1970). The details of the computational machinery are described in the Appendix. For the variance estimates, we use the bootstrap method since direct computation of information matrix of a HMM is cumbersome when the number of parameters involved in the aforementioned logistic regression and polychotomous regression is large.

3. SIMULATIONS

To evaluate the performance of parameter estimates in finite samples, we simulate data from a two-arm randomized clinical trial that follows the pHMM as previously described. Specifically, there are at most 8 visits; the initiation probabilities of the three states is $p_s = (0.25, 0.25, 0.5)$ for $s=(1,2,3)$; and the rows of the transition matrix are $(0.5, 0.3, 0.2)$, $(0.3, 0.3, 0.4)$, and $(0.1, 0.2, 0.7)$. The probabilities of having an event at each visit follow an inverse-logit function,

$$\frac{\exp\left(\beta_0 + \alpha_s 1_{[U_{ij}=s]} + \beta_s 1_{[U_{ij}=s \& Z_i=1]}\right)}{1 + \exp\left(\beta_0 + \alpha_s 1_{[U_{ij}=s]} + \beta_s 1_{[U_{ij}=s \& Z_i=1]}\right)},$$

with $\beta_0 = -2$, $\alpha_s = (*, -0.5, -0.2)$ (the N state is the baseline of the comparison) and three sets of treatment effects for $s=(N,C,B)$ respectively: Scenario 1 with $\beta_s = (0.0, 0.0, 0.0)$, Scenario 2 with $\beta_s = (-0.6, 0.2, -0.3)$, and Scenario 3 with $\beta_s = (-1.2, 0.4, -0.6)$, representing no efficacy, medium efficacy, and high efficacy. At each visit, 5% of at-risk participants are randomly censored. The randomization ratio for treatment and control is 1:1. One thousand datasets are generated for each of two sample sizes, 2000 and 8000. We estimate variances of parameter estimates by the bootstrap method using 200 resamples. Table 2 shows the characteristics of the DHR, as defined in (5), in the simulations. Since the CRR is derived directly from the DHR, the performance of CRR estimates is not shown.

Table 2 shows the characteristics of the treatment effect and the stratum effect in 1000 simulations, including bias, sample variance, estimated variance, coverage probability of 95% confidence interval, and the power or the type I error to reject a null effect. Among the three β_s estimators, $\widehat{\beta}_B$ has the lowest variance because this stratum is the most frequent. On the other hand, $\widehat{\beta}_C$ has the largest variance since the stratum C is less frequent and more ambiguous compared to the other two strata. The two α_s shows the similar pattern. All five estimates appear to be consistent toward their true values as the sample size increases. The bootstrap variance estimates provide good coverage probabilities for 95% confidence intervals for all effect estimates. The empirical type I errors in the first scenario with null treatment effects are close to 0.05. The power to detect the treatment effects in the N and B states is fairly low when the sample size is 2000, and it is much increased when the sample size is 8000. In either sample size, there is essentially no power to detect the treatment effect in the C principal state, again due to the low frequency and the ambiguity of this state in both arms.

4. DATA ANALYSES

4.1 Results from the pHMM

Three sets of analyses were conducted for the comparison between one of the three gel arms, namely PRO2000, BufferGel and PlaceboGel, and the No Gel arm. Table 3 shows the estimated estimands within three principal strata for three comparisons. The DHR describes the transient treatment effect for a principal state, and the CRR describes the causal cumulative risk ratio for women who were in the N, C, B condom use state consistently throughout the trial. Interestingly, there is an estimated 53% reduction of risk of HIV infection when women are assigned to the PRO2000 arm, for women who are consistently low-condom users (N); there is a 28% reduction of risk of HIV infection for women who are consistently high-condom users (B). The former estimate may be one step closer to the true efficacy of the PRO2000 gel, though the actual adherence of gel treatment has not been considered; the latter suggests that some protection offered by the gel may be masked by condom use. The estimate among women who are risk-gambling (C), 5% greater risk in the PRO2000 arm, is comparing the condom effect to the gel effect (Table 1). The gel is perhaps not as effective as condom use, therefore male partners switching to no condom use when women are assigned to the gel could increase the risk of HIV infection for women. None of these estimates has a confidence interval that excludes 1; in particular, the estimate for risk-gamblers (C) shows the widest confidence interval. This is not surprising because the risk-

gamble state has ambiguity in both arms (Table 1), and therefore has the greatest variability. Similar patterns of effect estimates are observed for the comparison between BufferGel and the No Gel arm. There is an estimated 38% reduction of HIV risk for a women who is a consistent low-condom user (N), and 3% reduction for a women who is a consistent high-condom user (B). Again HIV risk increases when participants gamble their risk (C), though none of these effects are statistically significant. The comparison between the placebo gel and the No Gel arm yields a mild effect among the three strata, again without any statistical significance.

Correspondingly, Table 4 displays three sets of parameter estimates for the Markov chain: initial probabilities of the states and the transition probabilities, that are averaged over all study participants to demonstrate the general pattern. Since the three gel arms are blinded, these estimates should be ideally the same for the three analyses. Indeed, they are very close numerically, confirming that blinding is largely effective. We note that these estimates vary across sites so that site-specific estimates would be more informative. At baseline, there are an estimated 15% low-condom users (N), 15% risk gamblers, and 70% high-condom users at enrollment. The transition probability from the current high-condom state to the next high-condom state is 0.826 in the comparison between PRO2000 and No Gel, the largest among all transition probabilities, whereas the transition probability from the current risk-gambling state to the next risk-gambling state is almost zero.

4.2 Model diagnostics

Compared to likelihood computation, maximization and asymptotics, model diagnostics for HMM are less well developed, mostly involving a qualitative comparison of observed frequencies and expected frequencies from the estimated HMM (Albert 1991; Leroux 1992; Hughes and Guttorp 1994). For our proposed pHMM, one special feature is that some of the states are actually observed, so we are able to directly compute part of the transition matrix, as shown in Figure 3. The comparison of transition probabilities at different visits suggests little variation. One of the key assumptions in our pHMM, the homogeneous transition probabilities, is therefore supported by empirical evidence.

We also evaluate the model fitting by plotting the observed rate of condom use and HIV infection during study visits along with the estimated rates from the fitted HMM (Figure 4). Because the condom use varies greatly among sites, we show the rates of condom use in the largest site - Durban, South Africa. The estimated condom use was computed by first estimating the probabilities of the condom use states at each visit through estimated parameters in the HMM, then averaging over the cohort in the site. Clearly, the fitted HMM largely captures the general trend of differential condom use. In the 9th and 10th visits, however, there is a lowered condom use which is not captured by the HMM. This is because the number of participants having the 9th and 10th visits are fairly small. We also plot the estimated cumulative HIV infection rates and observed ones. The HIV incidence rate at each visit varies greatly with many zero and thus was not plotted. The cumulative HIV infection rate is computed as 1 minus the survival curve. The estimated curve is the average of individual predicted curves based on the pHMM. It appears that the estimated cumulative HIV infection probabilities agree with the observed HIV infection probabilities fairly well, suggesting that the model fitting is likely adequate.

5. DISCUSSION

In this article, we address estimation of causal effect adjusting for time-varying post-treatment variables, such as sexual behaviors, in HIV prevention trials with a time-to-event endpoint. For longitudinal endpoints in clinical trials, latent class models have been proposed with Bayesian inference (Lin et al. 2008; Lin et al. 2009), though those methods

do not apply to our setting with time-to-event outcomes. Recognizing that the pattern of human behaviors for a participant is relatively stable during a short study period, we model the sequence of principal states and the HIV infection process by a HMM, which greatly facilitates the identifiability and likelihood maximization. The same rationale and procedure could potentially be applied to adjusting for treatment adherence, as compliance states can be defined as always-complier, never-complier, and complier (Imbens and Rubin 1997).

Our analysis of the HPTN 035 trial is exploratory and only serves as a supplement to the primary ITT analyses. The purpose is to reveal true effectiveness of the gel treatment disentangled from differential condom use observed in the HPTN 035 trial. We study the treatment effect within the latent population strata that characterize the condom use under both treatment assignments, for examples, people who would not use condom under either treatment assignment. The correlation of the latent states of longitudinal condom use was modeled by a hidden Markov model. We found suggestive evidence that within participants who would not consistently use the condom under either assignment, the gel offers a larger protection effect against HIV infection relative to the average ITT treatment effect.

The major limitation in interpreting the results of our analysis is that the condom-use is self-reported for the last sex act. Because the no-gel assignment is unblinded to participants, women may over-report the condom-use just because they know they do not have the gel protection. Furthermore, even if there is no differential reporting between the gel and no-gel arms, the condom-use in the last sex act does not cover the entire condom-use history during the 3-month interval, nor does it provide an accurate measure of actual condom-use. Most likely, no condom-use is more accurately reported than the use of condom, therefore the treatment effect estimated among those who reported no condom-use consistently is perhaps more believable than other principal effects.

The central issue in causal inference is identifiability. The principal strata are not observed, and therefore not identifiable in the study population, since we only get to observe one of two sets of potential outcomes. With a “scientifically informed” maneuver of assumptions and constraints, the causal estimands within these strata can be identified and estimated. There are concerns on making inferences on unidentified subgroups, and it is certainly of interest to develop sensitivity analysis methods that consider departures from the HMM assumptions that were needed to achieve identifiability. Nonetheless these subgroups define different patterns of behaviors and do exist in the study population. They are merely not observable under the current study design. Alternative designs, such as a blinded cross-over experiment, could collect data on both sets of potential outcomes, thereby informing which stratum a participant belong to.

Ultimately, the estimands within these subgroups provide useful estimates of product efficacy to scientists who develop the treatment product, inform the general public about the degree of treatment effect one would receive given her behavior pattern, and finally make it possible to extend the results from a clinical trial to the targeted population. The study participants in a clinical trial are selective and can be very different from the general population. For example, self-reported condom use in HIV prevention trials can achieve 80% after intense condom counseling (Figure 1), while in African communities where the microbicide gels are primarily targeted for, the rate of condom use is much lower (MEASURE Demographic and Health Surveys, <http://www.measuredhs.com/hivdata/surveys/start.cfm>). The causal estimand in the stratum with low condom use may better represent the treatment effect in the African population.

The HPTN 035 trial is a phase II/2B trial, a screening trial with roughly a quarter of the size of a Phase 3 confirmatory trial. The sample size does not permit sufficient power for testing

treatment effect within a latent population stratum, as shown in the simulation studies, thus the confidence intervals we got are generally not able to exclude 0. Nonetheless, the value of our paper is to illustrate a novel method to deal with time-varying principal stratification and time-to-event outcomes. The method can be applied to dealing with compliance in other HIV prevention trials, for example, recently completed iPrEX trial (Grant et al. 2010). We expect a phase 3 trial might provide adequate power to infer principal effect, at least for the no condom-use stratum. In future work, we will explore ways to improve efficiency of estimators within principal strata. For example, in this work we assume there are only baseline predictors of condom use states. Sometimes there are post-randomization predictors of principal strata available. Incorporating the post-randomization predictors could enhance the precision of parameter estimates.

Acknowledgments

Author's Footnote: James Y. Dai is Assistant Member in Fred Hutchinson Cancer Research Center, Seattle, WA, 98109 (jdai@fhcrc.org). Peter B. Gilbert is Member in Fred Hutchinson Cancer Research Center, Seattle, WA, 98109 (pgilbert@scharp.org). Benoît R. Mâsse is Associate Professor in University of Montreal and Associate Member in Fred Hutchinson Cancer Research Center, Seattle, WA, 98109 (ben@scharp.org). JYD and BRM were supported by the NIH grant 5 U01 AI068615-03. PBG was supported by the NIH grant 2 R37 AI054165-08. The authors thank James Hughes for helpful comments. HPTN 035 was funded by the US National Institutes of Health (NIH). The study was designed and implemented by the HIV Prevention Trials Network (HPTN) and the Microbicide Trials Network (MTN). The HPTN (U01AI46749) has been funded by the National Institute of Allergy and Infectious Diseases (NIAID), National Institute of Child Health and Development (NICHD), National Institute of Drug Abuse (NIDA), and National Institute of Mental Health (NIMH). The MTN (U01AI068633) has been funded by NIAID, NICHD, and NIMH. The study products were provided free of charge by Indevus Pharmaceuticals, Inc. and ReProtect, Inc. The US Agency for International Development (USAID) provided funding for manufacturing of BufferGel for this study.

APPENDIX A: IDENTIFIABILITY OF PARAMETERS IN

Since we assume independent censoring (Assumption 4), the censoring probabilities at each a_j can be identified from the observed data, irrespective of whether the rest of model parameters are identifiable. Also, identifiability of targeted parameters does not depend on the baseline covariates W , though W could help in the precision of the estimation. Therefore in what follows, we prove the identifiability of Θ in the scenario where there is no censoring and no baseline covariates. We reduce parameters to $\Theta = \{p_s, q_{st}, \lambda_s, \beta_s\}$: p_s is the initiation probabilities for state s ; q_{st} is the transition probability from state s to t ; λ_s is the probability of having an event in the s state at the a_j visit when the assignment is the control; β_s is the treatment effect for the s state, so that the probability of having an event in the s state in the treatment arm is $\lambda_s \exp(\beta_s)$. Let P_Θ denote the probability distribution of the HMM indexed by the parameter set Θ . For easy exposition, we assume an exponential link function for the probability of HIV infection here, but the proof applies to other link functions.

The typical strategy to prove identifiability in HMM is to establish one-to-one mapping from the parameters in the latent process to the parameters in the observed process (Petrie 1969).

Specifically, let $\Omega = \{\rho_{zj}, \bar{x}_j, \tau_{zj}, \bar{x}_j\}$ denote the parameter set that indexes the probability of the observable condom-use process $\bar{X}_j$ under the z treatment assignment, at the j^{th} visit, and the corresponding outcome process $1_{[Y_j=a_j]}$ conditional on that condom-use pattern. For example, let $\rho_{01,1}$ denote the probability of the reported condom use for the control arm at time a_1 ; $\rho_{12,11}$ denote the probability of having $\bar{X}_j = (1, 1)$ at a_2 in the treatment arm. Ω is a reparameterization of Θ and Ω is identifiable as it describes observables. Hence to prove Θ is identifiable, we simply show that Θ is a one-to-one mapping from Ω .

We next write down the expression of $\{\rho, \tau\}$ as functions of $\{p_s, q_{st}, \lambda_{sj}, \beta_s\}$. Note that four sets of constraints are imposed for $\{p_s, q_{st}\}$: $\sum_s P_s = 1$ and $\sum_t q_{st} = 1$, for $s = 1, 2, 3$ and $t = 1, 2, 3$. Observe that at the first visit a_1 , because of the randomization and the monotonicity assumption

$$\begin{aligned}\rho_{01,0} &= p_1, \\ \rho_{01,1} &= p_2 + p_3,\end{aligned}\quad (\text{A.1})$$

$$\begin{aligned}\rho_{11,0} &= p_1 + p_2, \\ \rho_{11,1} &= p_3,\end{aligned}\quad (\text{A.2})$$

With (A.1), (A.2) and the constraint $p_1 + p_2 + p_3 = 1$, p_s is uniquely determined. For the HIV infection probabilities at the 2th visit, observe that

$$\tau_{00,1} = \lambda_1, \quad (\text{A.3})$$

$$\tau_{01,1} = p_2 \lambda_2 + p_3 \lambda_3, \quad (\text{A.4})$$

$$\tau_{10,1} = p_1 \lambda_1 \exp(\beta_1) + p_2 \lambda_2 \exp(\beta_2), \quad (\text{A.5})$$

$$\tau_{11,1} = p_3 \lambda_3 \exp(\beta_3). \quad (\text{A.6})$$

From (A.3) and (A.6), λ_1 and $\lambda_3 \exp(\beta_3)$ are identified. After the first visit, those participants who were having HIV infection will not advance to the second visit, therefore at the second visit a_2 , four independent equations can be written for the observed condom-use probabilities:

$$\rho_{02,10} = p_3 (1 - \lambda_3) q_{31} + p_2 (1 - \lambda_2) q_{21}, \quad (\text{A.7})$$

$$\rho_{02,00} = p_1 (1 - \lambda_1) q_{11}, \quad (\text{A.8})$$

$$\rho_{12,11} = p_3 [1 - \lambda_3 \exp(\beta_3)] q_{33}, \quad (\text{A.9})$$

$$\rho_{12,01} = p_1 [1 - \lambda_1 \exp(\beta_1)] q_{13} + p_2 [1 - \lambda_2 \exp(\beta_2)] q_{23}. \quad (\text{A.10})$$

For the observed HIV infection probabilities at a_2 ,

$$\tau_{02,01} = p_1 (1 - \lambda_1) q_{12} \lambda_2 + p_1 (1 - \lambda_1) q_{13} \lambda_3 \quad (\text{A.11})$$

$$\tau_{12,10} = p_3 [1 - \lambda_3 \exp(\beta_3)] q_{31} \lambda_1 \exp(\beta_1) + p_3 [1 - \lambda_3 \exp(\beta_3)] q_{32} \lambda_2 \exp(\beta_2). \quad (\text{A.12})$$

From (A.8) and (A.9), q_{11} and q_{33} are uniquely determined. For the condom-use probabilities in the no-gel arm at a_3 and a_4 , the following expressions can be written:

$$\rho_{03,101} = p_3 (1 - \lambda_3) q_{31} (1 - \lambda_1) q_{12} + p_2 (1 - \lambda_2) q_{21} (1 - \lambda_1) q_{13}, \quad (\text{A.13})$$

$$\rho_{04,101} = p_3 (1 - \lambda_3) q_{31} (1 - \lambda_1) q_{12} (1 - \lambda_2) q_{21} + p_2 (1 - \lambda_2) q_{21} (1 - \lambda_1) q_{13} (1 - \lambda_3) q_{31} \quad (\text{A.14})$$

Let $A = p_3(1-\lambda_3)q_{31}$, $B = p_2(1-\lambda_2)q_{12}$, $C = (1-\lambda_1)q_{12}$, and $D = (1-\lambda_1)q_{13}$, the following equations can be written based on (A.7), (A.8), (A.13), and (A.14):

$$\begin{aligned} A+B &= \rho_{02,10} \\ C+D &= 1 - \rho_{02,00} \\ AC+BD &= \frac{\rho_{03,101}}{p_3} \\ ACB+BDA &= \rho_{04,101} \end{aligned}$$

Thus A, B, C and D are all identified from these equations. Now (A.11) can be expressed as

$$C\lambda_2 + D\lambda_3 = \frac{\tau_{02,01}}{p_1}$$

Together with (A.4), λ_2 and λ_3 are identified. This leads to identification of q_{31} , q_{21} , q_{32} , q_{12} , q_{13} . Using the identified parameters, the identifiability of β_1 and β_2 is determined from the expressions from the gel arm, (A.5) and (A.12). The identifiability of q_{23} and q_{22} follows from (A.10).

APPENDIX B: THE MODIFIED BAUM-WELCH ALGORITHM

For easy exposition, we describe a modified Baum-Welch (backward-forward) algorithm for a subject in the control arm who has HIV infection in the J^{th} visit. The observed data are

therefore $(Z_i=0, Y_i, \Delta_i, \bar{X}_{ij}, W_i)$ for this subject. The computation for the likelihood of the rest of subjects follows easily. For ease of exposition, we suppress the subscript i henceforth and omit W_i until we address the update of p_s , q_{sb} and λ_s in the EM algorithm. To conform to the standard notation in the HMM literature, we re-express the data in the form of longitudinal outcomes, $(Z, D_{1:j}, X_{1:j})$, where $D_{1:j}$ is a vector of length J , with the j^{th} element $1_{[Y=j]}$ and $X_{1:j}$ represents a vector with the j^{th} element X_j .

Consider the forward variable $\alpha_j(s) = \Pr(D_{1:j}, X_{1:j} | U_j = s)$, where j is an integer $1 \leq j \leq J$, which represents the joint probabilities of the condom-use principal state at a_j and the observed data up to a_j . Let s takes values in $(1,2,3)$, representing the three states (N,C,B). Since (U_j, Z) completely determines X_j , we denote $X_j \in U_j$ that the observed condom use is compatible with the underlying condom-use principal state. Hence we have

$$\alpha_j(s) = \Pr(D_{1:j}, X_{1:j}, U_j = s) 1_{[X_j \in U_j]}.$$

The recursive factorization of $\alpha_j(s)$ is given inductively: **Initialization**

$$\alpha_1(s) = \Pr(U_1 = s) \Pr(D_1 | U_1 = s) 1_{[X_1 \in U_1]} = p_s (1 - \lambda_s)$$

Induction

$$\begin{aligned}
\alpha_{j+1}(t) &= \Pr(D_{1:j+1}, X_{1:j+1}, U_{t+1}=t) 1_{[X_{j+1} \in U_{j+1}]} \\
&= \sum_{s=1}^3 \alpha_j(s) \Pr(D_{j+1}, X_{j+1}, U_{j+1}=t | D_{1:j}, X_{1:j}, U_j=s) 1_{[X_{j+1} \in U_{j+1}]} \\
&= \sum_{s=1}^3 \alpha_j(s) \Pr(D_{j+1} | D_{1:j}, U_{j+1}=t) \Pr(U_{j+1}=t | U_t=s) 1_{[X_{j+1} \in U_{j+1}]} \\
&= \sum_{s=1}^3 \alpha_j(s) \lambda_t^{D_{j+1}} (1 - \lambda_t)^{1-D_{j+1}} q_{st} 1_{[X_{j+1} \in U_{j+1}]}
\end{aligned}$$

In a similar manner, we consider a backward variable $\gamma_j(s) = \Pr(D_{j+1:J}, X_{j+1:J} | U_j = s)$ that denotes the probability of the observed data after a_j given the condom use state at a_j . Again we can solve for $\gamma_j(s)$ inductively, as follows: **Initialization**

$$\gamma_J(s) = 1, 1 \leq s \leq 3$$

Induction

$$\begin{aligned}
\gamma_j(s) &= \frac{\Pr(D_{j+1:J}, X_{j+1:J}, U_j=s)}{\Pr(U_j=s)} \\
&= \frac{\sum_{t=1}^3 \Pr(D_{j+1}, X_{j+1}, U_j=s | D_{j+2:J}, X_{j+2:J}, U_{j+1}=t) \gamma_{j+1}(t) \Pr(U_{j+1}=t) 1_{[X_{j+1} \in U_{j+1}]}}{\Pr(U_j=s)} \\
&= \sum_{t=1}^3 \lambda_t^{D_{j+1}} (1 - \lambda_t)^{1-D_{j+1}} q_{st} \gamma_{j+1}(t) 1_{[X_{j+1} \in U_{j+1}]}
\end{aligned}$$

The EM algorithm for parameter estimation in typical HMMs has been elaborated in Cappé et al. (2005). Following the same notations, we define the conditional probabilities of the hidden states given the observed data to be “smoothing functions”, including both probabilities for one state and probabilities for two consecutive states.

$$\begin{aligned}
\phi_{j|J}(s) &= \Pr(U_j=s | D_{1:J}, X_{1:J}), \quad 1 \leq j \leq J, \\
\phi_{j|J}(st) &= \Pr(U_j=s, U_{j+1}=t | D_{1:J}, X_{1:J}), \quad 1 \leq j \leq J.
\end{aligned}$$

These smoothing functions can be calculated by the aforementioned forward-backward variables. Specifically,

$$\begin{aligned}
\phi_{j|J}(s) &= \frac{\Pr(D_{1:j}, X_{1:j}, U_j=s) \Pr(D_{j+1:J}, X_{j+1:J} | U_j=s)}{\sum_{t=1}^3 \Pr(D_{1:j}, X_{1:j}, U_j=t) \Pr(D_{j+1:J}, X_{j+1:J} | U_j=t)} \\
&= \frac{\alpha_j(s) \gamma_j(s)}{\sum_{t=1}^3 \alpha_j(t) \gamma_j(t)}.
\end{aligned}$$

And similarly,

$$\begin{aligned}
\phi_{j,t}(st) &\propto \frac{\Pr(U_j=s, D_{1:j}, X_{1:j}) \Pr(D_{j+2:t}, X_{j+2:t} | U_{j+1}=t) \Pr(D_{j+1}, X_{j+1} | U_{j+1}=t) \Pr(U_{j+1}=t | U_j=s)}{\sum_s \sum_t \alpha_j(s) \gamma_{j+1}(t) q_{st} \lambda_t^{D_{j+1}} (1-\lambda_t)^{1-D_{j+1}} 1_{[X_{j+1} \in U_{j+1}]}} \\
&= \frac{\alpha_j(s) \gamma_{j+1}(t) q_{st} \lambda_t^{D_{j+1}} (1-\lambda_t)^{1-D_{j+1}} 1_{[X_{j+1} \in U_{j+1}]}}{\sum_s \sum_t \alpha_j(s) \gamma_{j+1}(t) q_{st} \lambda_t^{D_{j+1}} (1-\lambda_t)^{1-D_{j+1}} 1_{[X_{j+1} \in U_{j+1}]}}
\end{aligned}$$

Given the current estimates, a single EM cycle runs the forward-backward algorithm, computes the smoothing probabilities, and updates parameter estimates. The smoothing probabilities serve as weights in the likelihood maximization. New estimates of p_s , q_{st} are obtained by fitting a polychotomous regression of the current state on the previous state if available, and W . New estimates of λ_s and β_s is obtained by fitting a logistic regression of HIV infection status on Z , U_j and W .

REFERENCES

- Albert PS. A two-state Markov mixture model fro a time series of epileptic seizure counts. *Biometrics*. 1991; 47:1371–1381. [PubMed: 1786324]
- Altman RM, Petkau AJ. Application of hidden Markov models to multiple sclerosis lesion count data. *Statistics in Medicine*. 2005; 24:2335–2344. [PubMed: 15909288]
- Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*. 1996; 91:444–455.
- Baum LE, Petrie T. Statistical inferences for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*. 1966; 37:1554–1563.
- Baum LE, Petrie T, Soules G, Weiss N. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*. 1970; 41:164–171.
- Bickel PJ, Ritov Y, Rydén T. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *The Annals of Statistics*. 1998; 26:1614–1635.
- Cappé, O.; Moulines, E.; Ryden, T. *Inference in Hidden Markov Models*. Springer; 2005.
- Cassel MM, Halperin DT, Shelton JD, Stanton D. Risk compensation: the Achilles' heel of innovations in HIV prevention. *Br Med J*. 2006; 332:605–607. [PubMed: 16528088]
- Celum C, Wald A, Hughes J, Sanchez J, Stewart R, Delany-Moretlwe S, Cowan F, Casapia M, Ortiz A, Fuchs J, Buchbinder S, Koblin B, Zwierski S, Rose S, Wang J, Corey L, the HPTN 039 Protocol Team. Effect of aciclovir on HIV-1 acquisition in herpes simplex virus 2 seropositive women and men who have sex with men: a randomised double-blind, placebo-controlled trial. *Lancet*. 2008; 371:2109–2119. [PubMed: 18572080]
- Centers for Disease Control. Update: barrier protection against HIV infection and other sexually transmitted diseases. *MMWR*. 1993; 42:589–597. [PubMed: 8336689]
- Cox, DR. *Planning of Experiments*. Wiley; New York: 1958.
- Cox, DR.; Hinkley, DV. *Theoretical Statistics*. Chapman and Hall; New York: 1974.
- Crepaz N, Hart TA, Marks G. Highly active antiretroviral therapy and sexual risk behavior: a meta-analytic review. *Journal of American Medical Association*. 2004; 292:224–236.
- Elias CJ, Coggins C. Female controlled method to prevent sexual transmission of HIV. *AIDS*. 1996; 10:S43–S51. [PubMed: 8970711]
- Fleming TR, Richardson BA. Some design issues in trials of microbicides for the prevention of HIV infection. *Journal of Infectious Diseases*. 2004; 190:666–74. [PubMed: 15272392]
- Frangakis CE, Brookmeyer RS, Varadhan R, Safaeian M, Vlahov D, Strathdee SA. Methodology for evaluating a partially controlled longitudinal treatment using principal stratification, with application to a needle exchange program. *Journal of the American Statistical Association*. 2004; 99:239–249.

- Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
- Gilbert PB, Bosch RJ, Hudgens MG. Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*. 2003; 59:531–541. [PubMed: 14601754]
- Grant RM, Lama JR, L. AP, McMahan V, Liu AY, Vargas L, Giocochea P, the iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *The New England Journal of Medicine*. 2010; 363:2587–2599. [PubMed: 21091279]
- Halperin, D. What do gay men want?: An essay on sex, risk and subjectivity. University of Michigan Press; Ann Arbor: 2007.
- Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*. 2001; 96:440–448.
- Holland PW. Statistics and causal inference. *Journal of the American Statistical Association*. 1986; 81:945–960.
- Hughes JP, Guttorp P. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resource Research*. 1994; 30:1535–1546.
- Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*. 1997; 25:305–327.
- Joffe MM, Greene T. Related causal frameworks for surrogate outcomes. *Biometrics*. 2009; 65:530–538. [PubMed: 18759836]
- Kalbfleisch, JD.; Prentice, RL. The statistical analysis of failure time data. Wiley; New Jersey: 2002.
- Karim AS, Richardson BA, Ramjee G, Hoffman IF, Chirenje MZ, Taha T, Kapina M, Maslankowski L, Coletti A, Profy A, Moench TR, Piwowar-Manning E, Masse B, Hillier SL, Soto-Torres L. Safety and effectiveness of vaginal microbicides BufferGel and 0.5% PRO 2000/5 gel for the prevention of HIV infection in women. *AIDS*. 2011; 25:957–966. [PubMed: 21330907]
- Lander ES, Green P. Construction of multilocus genetic linkage maps in humans. *Proceedings of the National Academy of Sciences of the United States of America*. 1987; 84(8):2363–2367. [PubMed: 3470801]
- Leroux BG. Maximum-likelihood estimation for hidden Markov models. *Stochastic process and their applications*. 1992; 40:127–143.
- Lin JY, Ten Have TR, Elliott MR. Longitudinal nested compliance class model in the presence of time-varying noncompliance. *Journal of the American Statistical Association*. 2008; 103:462–473.
- Lin JY, Ten Have TR, Elliott MR. Nested Markov compliance class model in the presence of time-varying noncompliance. *Biometrics*. 2009; 65:505–513. [PubMed: 18759831]
- Neyman J. On the application of probability theory to agricultural experiments: essay on principles, Section 9. *Statistical Science*. 1923; 5:465–480. 1990.
- Petersen ML, Sinisi SE, van der Laan M. Estimation of direct causal effect. *Epidemiology*. 2006; 17:276–284. [PubMed: 16617276]
- Petrie T. Probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics*. 1969; 40:97–115.
- Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*. 1989; 77:257–284.
- Reid SE, Dai JY, Wang J, Sichalwe BN, Alkpomiemie G, Cowan FM, Delany-Moretlwe S, Baeten JM, Jughes JP, Wald A, Celum C. Pregnancy, contraceptive use and HIV acquisition in HPTN 039: relevance for HIV prevention trials among African women. *J Acquir Immune Defic Syndr*. 2010; 53:606–613. [PubMed: 19838129]
- Robins J, Greenland S. Identifiability and exchangeability for direct and indirect effect. *Epidemiology*. 1992; 3:143–155. [PubMed: 1576220]
- Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000; 11:550–560. [PubMed: 10955408]
- Robins, JM.; Rotnitzky, A.; Scharfstein, DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran, E.; Berry, D., editors. *Statistical Models in Epidemiology*. Springer-Verlag; New York: 2000.

- Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by treatment. *The Journal of the Royal Statistical Society, Series A*. 1984; 47:656–666.
- Rosenblum M, Jewell NP, van der Laan M, Shiboski S, van der Straten A, Padian N. Analysing direct effects in randomized trials with secondary interventions: an application to human immunodeficiency virus prevention trials. *Journal of the Royal Statistical Society, Series A*. 2009; 172:443–465.
- Rubin DB. Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*. 1978; 6:34–58.
- Rubin DB. Direct and indirect effects via potential outcomes. *Scandinavian Journal of Statistics*. 2004; 31:161–170.
- Scott SL, James GM, Sugar CA. Hidden Markov models for longitudinal comparisons. *Journal of American Statistical Association*. 2005; 100:359–369.
- Shirley KE, Small DS, Lynch KG, Maisto SA, Oslin DW. Hidden Markov models for alcoholism treatment trial data. *Annals of Applied Statistics*. 2010; 3:491–520.
- Stone AB, Hitchcock PJ. Vaginal microbicides for preventing the sexual transmission of HIV. *AIDS*. 1994; 8:S285–S293.
- van der Laan, M.; Robins, JM. Unified methods for censored longitudinal data and causality. Springer-Verlag; Cambridge, UK: 2003.
- Yuan M, Kendziorski C. Hidden Markov models for microarray time course data in multiple biological condition. *Journal of American Statistical Association*. 2006; 101:1323–1340.

Figure 1.

The quarterly self-reported condom use in the last sex act in the largest site of HPTN 035 - Durban, South Africa. The circles represent the percentage of women in the No Gel (control) arm who reported using a condom in the last sex act for the quarterly visit. The triangles represent the percentage of women in the gel arms who reported using a condom in the last sex act. The numbers on top of the x axis are the numbers of women at risk at the visit: the top line of numbers for the No Gel (control) arm and the lower line of numbers for the gel arms. The decreasing numbers after the 4th quarter are due to censoring at the end of study.

Figure 2.

A diagram of the hidden Markov model for subject i with the hidden condom-state U_{ij} , observed condom use X_{ij} and the observed time-to-HIV infection detection or censoring (Y_i). U_{ij} forms a first-order homogeneous Markov chain. At each visit a_j , U_{ij} and the treatment assignment determine the observed condom use X_{ij} and predict the probability of the outcome Y_i .

Figure 3.

Transition probabilities for the observed condom-use states at each visit for the gel arms and the No Gel arm. The left panel shows the transition probabilities from condom-use (the state B) to condom-use (the state B) in the gel arms. The right panel shows the transition probabilities from no condom-use (the state N) to no condom-use (the state N) in the No Gel arm. The number of participants used in the computation was listed on the bottom of the graph. The 95% confidence intervals were also plotted.

Figure 4.

Model diagnostics for the analysis comparing the PRO2000 arm versus the No Gel arm. The left panel shows the observed proportion of condom use and the estimated proportion of condom use in the largest site - Durban, South Africa, based on the HMM, separately for the gel arms and the control arm. The right panel shows the observed event probability curve (one minus a survival curve) for the whole study cohort and the estimated curve based on the HMM.

Table 1

Condom-use principal states and the corresponding hazard of HIV infection at visit a_j , assuming the treatment effect on hazard is additive on the log scale.

condom use state	No Gel ($Z_i = 0$)			Gel ($Z_i = 1$)	
	U_{ij}	$X_{ij}(0)$	hazard	$X_{ij}(1)$	hazard
use under neither assignment	N	0	λ_1	0	$\lambda_1 \exp(\beta_1)$
use under control only (risk gamble)	C	1	λ_2	0	$\lambda_2 \exp(\beta_2)$
use under both assignments	B	1	λ_3	1	$\lambda_3 \exp(\beta_3)$

Table 2

The characteristics of treatment effect $\widehat{\beta}_s$ and the stratum effect $\widehat{\alpha}_s$ in 1000 simulated datasets. The true values for α_s is $(*, -0.5, -0.2)$ and the true values for β_s is $(0.0, 0.0, 0.0)$ in Scenario 1, $(-0.6, 0.2, -0.3)$ in Scenario 2, and $(-1.2, 0.4, -0.6)$ in Scenario 3.

	n=2000					n=8000				
	Bias	Var ^a	Var ^b	C.P. ^c	Power/ Error ^d	Bias	Var ^a	Var ^b	C.P. ^c	Power/ Error ^d
Scenario 1										
$\widehat{\beta}_1$	-0.024	0.127	0.126	0.938	0.056	-0.014	0.021	0.023	0.953	0.047
$\widehat{\beta}_2$	0.028	0.505	0.624	0.965	0.035	0.015	0.071	0.090	0.968	0.032
$\widehat{\beta}_3$	-0.008	0.035	0.039	0.956	0.044	0.003	0.008	0.008	0.945	0.055
$\widehat{\alpha}_2$	0.025	0.150	0.208	0.967	0.067	0.012	0.042	0.043	0.945	0.684
$\widehat{\alpha}_3$	-0.003	0.041	0.043	0.955	0.108	-0.002	0.009	0.009	0.938	0.558
Scenario 2										
$\widehat{\beta}_1$	0.040	0.157	0.183	0.946	0.128	0.009	0.040	0.043	0.952	0.885
$\widehat{\beta}_2$	-0.036	0.308	0.412	0.987	0.018	-0.018	0.077	0.074	0.966	0.071
$\widehat{\beta}_3$	-0.012	0.029	0.035	0.956	0.384	0.005	0.009	0.008	0.949	0.880
$\widehat{\alpha}_2$	-0.023	0.175	0.199	0.965	0.080	0.007	0.046	0.043	0.955	0.693
$\widehat{\alpha}_3$	0.011	0.032	0.039	0.956	0.118	0.001	0.008	0.009	0.956	0.558
Scenario 3										
$\widehat{\beta}_1$	0.084	0.320	0.393	0.924	0.329	0.078	0.082	0.091	0.906	0.986
$\widehat{\beta}_2$	-0.036	0.263	0.317	0.985	0.017	-0.030	0.065	0.068	0.938	0.247
$\widehat{\beta}_3$	-0.005	0.033	0.039	0.960	0.777	-0.006	0.010	0.009	0.954	0.996
$\widehat{\alpha}_2$	-0.011	0.177	0.206	0.975	0.063	0.001	0.051	0.047	0.954	0.660
$\widehat{\alpha}_3$	0.004	0.037	0.039	0.949	0.109	0.007	0.010	0.010	0.952	0.510

^a sample variance of parameter estimates;

^b variance estimate by 200 bootstrap samples;

^c coverage probabilities of 95% confidence interval;

^d empirical power or type I error to test the parameter is zero.

Table 3

The estimated causal estimands and confidence intervals for the comparisons between one of the three gel arms and the No Gel arm.

Comparison	DHR		CRR	
	estimate	95% C.I.	estimate	95% C.I.
PRO2000 vs No Gel				
N	0.47	[0.14,1.59]	0.56	[0.16,2.04]
C	1.05	[0.29,3.80]	1.02	[0.22,4.70]
B	0.68	[0.40,1.15]	0.70	[0.45,1.09]
BufferGel vs No Gel				
N	0.62	[0.19,2.00]	0.68	[0.23,2.01]
C	2.88	[0.90,9.21]	2.80	[0.88,8.90]
B	0.97	[0.59,1.60]	0.99	[0.53,1.53]
Placebo vs No Gel				
N	0.80	[0.25,2.60]	0.80	[0.27,2.36]
C	0.76	[0.28,2.07]	0.91	[0.39,2.13]
B	1.08	[0.66,1.77]	1.05	[0.68,1.62]

Table 4

The parameter estimates for the Markov chain averaged over all subjects. The estimates are averages over all study participants.

Comparison	N	C	B
PRO2000 vs No Gel			
Initial probability	0.146	0.157	0.697
Transition matrix			
N	0.532	0.270	0.198
C	0.282	0.003	0.715
B	0.084	0.090	0.826
BufferGel vs No Gel			
Initial probability	0.145	0.167	0.688
Transition matrix			
N	0.505	0.268	0.227
C	0.335	0.038	0.627
B	0.068	0.099	0.833
Placebo vs No Gel			
Initial probability	0.144	0.236	0.620
Transition matrix			
N	0.518	0.324	0.158
C	0.312	0.056	0.632
B	0.061	0.114	0.825