Published in final edited form as:

Biometrics. 2014 September ; 70(3): 742–750. doi:10.1111/biom.12183.

Estimating the Effectiveness in HIV Prevention Trials by Incorporating the Exposure Process: Application to HPTN 035 Data

Jingyang Zhang^{1,*} and Elizabeth R. Brown^{1,2}

¹Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-C200, Seattle, Washington 98109, U.S.A

²Department of Biostatistics, University of Washington, Seattle, Washington 98109, U.S.A

Summary

Estimating the effectiveness of a new intervention is usually the primary objective for HIV prevention trials. The Cox proportional hazard model is mainly used to estimate effectiveness by assuming that participants share the same risk under the covariates and the risk is always non-zero. In fact, the risk is only non-zero when an exposure event occurs, and participants can have a varying risk to transmit due to varying patterns of exposure events. Therefore, we propose a novel estimate of effectiveness adjusted for the heterogeneity in the magnitude of exposure among the study population, using a latent Poisson process model for the exposure path of each participant. Moreover, our model considers the scenario in which a proportion of participants never experience an exposure event and adopts a zero-inflated distribution for the rate of the exposure process. We employ a Bayesian estimation approach to estimate the exposure-adjusted effectiveness eliciting the priors from the historical information. Simulation studies are carried out to validate the approach and explore the properties of the estimates. An application example is presented from an HIV prevention trial.

Keywords

Hierarchical models; HIV prevention; Intercourse; Markov chain Monte Carlo; Per-exposure effectiveness; Zero-inflated gamma

1. Introduction

Worldwide, 2.5 million [2.2–2.8 million] people became newly infected with HIV in 2011 with women in Sub-Saharan Africa being disproportionately impacted by the epidemic (Joint United Nations Programme on HIV/AIDS (UNAIDS), 2012). There are several completed (Van Damme et al., 2002; National Institute of Allergy and Infectious Diseases (NIAID), 2008; Skoler-Karpoff et al., 2008; McCormack et al., 2010; Abdool Karim et al., 2010, 2011; Thigpen et al., 2012; Baeten et al., 2012) and ongoing (CONRAD, 2011; International Partnership for Microbicides, Inc. and National Institutes of Health (NIH),

*jzhang2@fhcrc.org.

2012; International Partnership for Microbicides, Inc., 2012) clinical trials designed to assess the effectiveness of topical microbicides and oral chemoprophylaxis for prevention of heterosexually transmitted HIV. Unfortunately, several of the completed trials have failed to show consistent results. There are many potential explanations for these differences with a main explanation being differences in compliance to the intervention. Another potential driver of these discrepancies is differences in risk distributions between populations. For example, in a study that recruits serodiscordant couples where one partner is known to be infected, the enrolled HIV-negative participants are all exposed to HIV; while studies that enroll HIV-negative participants from high prevalence settings (10–30%) will likely have a high proportion of participants with no HIV exposure since the majority of the participants will be in a mutually monogamous relationship with an HIV-uninfected partner. Even in serodiscordant partner studies, sex behaviors are extremely different from couple to couple (Blower and Boe, 1993), resulting in variation in the magnitude of exposure due to types and frequency of exposure. In most of these studies, the primary outcome is effectiveness (=1-hazard ratio) from a proportional hazards model, a population measure that describes the reduction in incidence for an entire population. In populations with high heterogeneity in risk, population effectiveness can vary dramatically from an individual measure of effectiveness, the reduction in risk that an individual can expect from an intervention. It is in fact possible for an intervention to appear to be harmful on the population level when in fact it is effective at the individual level due to the heterogeneity (Aalen, 1988; Auvert et al., 2011; O’Hagan et al., 2012). Here, we propose a model that reflects heterogeneity in exposure and aims to estimate the individual effect of the intervention, as opposed to the overall effectiveness for the population.

Models for effectiveness have traditionally been driven by the type of data collected, timing and results of HIV test, rather than the biological/behavioral/physical processes driving infection. For estimating effectiveness, HIV prevention trials have adopted the Cox proportional hazards model (Cox and Oakes, 1984), assuming that participants with a given set of covariates share the same risk and that this risk is always non-zero. Biologically, however, the risk is non-zero only when an exposure (sexual activity with an HIV-infected partner) occurs. Participants in a clinical trial have varying patterns of sexual activity with partners at varying risks to transmit, including some who remain HIV-uninfected. In this article, motivated by the underlying biological process of transmission of HIV and other sexually transmitted diseases (STDs), we formulate a statistical model to estimate the per-exposure infectivity and better assess the effectiveness for the HIV prevention trials by taking some exclusive features of HIV and other STDs into consideration.

Exposure to STDs (including HIV) occurs at discrete and identifiable times. Because these exposures are required for sexual transmission of HIV, the time to HIV infection is dependent on the type and the frequency of exposure events. If every sexual encounter is an exposure event (i.e., in HIV serodiscordant couples) and the exposure over time is easy to track for all participants, the per-exposure infectivity can be estimated through statistical models similar to those derived for fertility and fecundity studies (Barrett and Marshall, 1969; Zhou et al., 1996; Kim et al., 2010; Sundaram, McLain, and Buck Louis, 2012). But in practice, it is impractical, if not impossible, to identify the exposure and the exposure times in HIV and other STD studies. Even if we know the complete sexual activity of a

participant, in non-couple studies we still do not know the infection status of the partner. When partial knowledge of the exposure process is available, statistical models for the per-exposure risk of infection have been discussed in Satten, Mastro, and Longini (1994) and Vittinghoff et al. (1999). Wilson (2010) built upon Barrett and Marshall's model to calculate the overall effectiveness of an intervention for prevention of STDs using the per-act infectivity and the efficacy from previous studies. This approach provides an alternative estimate of the overall effectiveness of an intervention accounting for the magnitude of exposures, but it did not account for the substantial heterogeneity in the magnitude. Instead, we borrow the idea of Rhodes, Halloran, and Longini (1996) and propose a novel model that utilizes a counting process model for the contact rates, so the exposure information for each subject is a stochastic process rather than a fixed covariate matrix.

The rest of the article is organized as follows. In Section 2, we introduce a stochastic model for the time to infection by counting exposure events as a mixed Poisson process. Next in Section 3, we consider the scenario in which there is a group of unexposed participants and present a zero-inflated distribution for the rate of the exposure process. The Bayesian estimation approach is described in Section 4 with constructing the likelihood function and eliciting the priors from the historical information. The model is assessed by simulation studies in Section 5 and illustrated with the data from HPTN 035 in Section 6. We conclude with the discussion in Section 7.

2. Modeling HIV Risk

We begin by introducing notations and assumptions for our model. For the i th participant, ($i = 1, \dots, N$), let T_i denote the time to infection, which is the time when the transmission occurs at one specific exposure to HIV. Unlike the time to detection, T_i is not directly observed and presumes that we know exactly at which sexual exposure the infection occurs. The randomization is denoted by a dichotomous variable G_i , with $G_i = 1$ indicating the intervention arm. $\{N_i(t), t > 0\}$ is the stochastic process representing the process for the exposure events. Here, we assume that $\{N_i(t), t > 0\}$ is a Poisson process with rate λ_i . ρ denotes the per-contact risk of infection for subjects at risk without the intervention and ε is the effectiveness of the intervention per exposure. ρ and ε are shared across the population.

The probability of acquiring HIV at each exposure is $(1 - G_i\varepsilon)\rho$, and the number of exposure events until the infection, X_i , follows a geometric distribution, namely,

$$\Pr(X_i = x_i | G_i, \rho, \varepsilon) = (1 - (1 - G_i\varepsilon)\rho)^{x_i - 1} (1 - G_i\varepsilon)\rho, \text{ for } x_i = 1, 2, 3, \dots$$

This is analogous to a discrete hazard model where time is instead incremented on the exposure scale. That is, each sexual exposure increments time by one unit. Therefore, we could think of the participant i as following a separate time scale according to the exposure process $\{N_i(t), t > 0\}$, which is assumed to be a Poisson process. Based on these assumptions, given λ_i , the time to infection T_i , or equivalently the time to the X_i th exposure (the exposure at which transmission occurs), is distributed as the sum of X_i i.i.d. exponential

distributed variables with rate λ_i , which is equivalent to $\Gamma(X_i, \lambda_i)$ ($\Gamma(a, b)$ represents the gamma distribution with shape a and rate b throughout the article.) (Ross, 1995).

Given ρ , ε and λ_i , the time to infection T_i can be modeled hierarchically,

$$\begin{aligned} T_i | X_i, \lambda_i, \rho &\sim \Gamma(X_i, \lambda_i), \\ X_i | G_i, \rho, \varepsilon &\sim \text{Geometric}((1 - G_i \varepsilon) \rho). \end{aligned}$$

Accordingly, the pdf of T_i given ρ , ε , and λ_i under G_i is

$$\begin{aligned} f(t | G_i, \rho, \varepsilon, \lambda_i) &= \sum_{x_i=1}^{+\infty} f(t | x_i, G_i, \rho, \varepsilon, \lambda_i) p(x_i | G_i, \rho, \varepsilon) \\ &= \sum_{x_i=1}^{+\infty} \frac{\lambda_i^{x_i} t^{x_i-1} \exp(-\lambda_i t)}{\Gamma(x_i)} (1 - (1 - G_i \varepsilon) \rho)^{x_i-1} (1 - G_i \varepsilon) \rho \\ &= \lambda_i (1 - G_i \varepsilon) \rho \exp(-\lambda_i (1 - G_i \varepsilon) \rho t) \times \sum_{x_i=1}^{+\infty} \frac{[\lambda_i t (1 - (1 - G_i \varepsilon) \rho)]^{x_i-1} \exp(-\lambda_i t (1 - (1 - G_i \varepsilon) \rho))}{(x_i-1)!} \\ &= [\lambda_i (1 - G_i \varepsilon) \rho] \exp(-\lambda_i (1 - G_i \varepsilon) \rho t), \end{aligned} \quad (1)$$

indicating that $T_i | G_i = 0$ and $T_i | G_i = 1$ are exponentially distributed conditioning on ρ , ε , and λ_i .

To allow heterogeneity between participants in their HIV/STD exposures, λ_i can be treated as a random variable following a gamma distribution $\Gamma(\alpha, \beta)$. The process then becomes a mixed Poisson process (Grandell, 1997). It is easy to show that

$$\begin{aligned} f(t | G_i, \rho, \varepsilon, \alpha, \beta) &= \int_0^{+\infty} f(t | G_i, \lambda_i, \rho, \varepsilon) p(\lambda_i | \alpha, \beta) d\lambda_i \\ &= \int_0^{+\infty} [\lambda_i (1 - G_i \varepsilon) \rho] \exp(-\lambda_i (1 - G_i \varepsilon) \rho t) \frac{\beta^\alpha \lambda_i^{\alpha-1} \exp(-\beta \lambda_i)}{\Gamma(\alpha)} d\lambda_i \\ &= \left[\int_0^{+\infty} \frac{(\beta + (1 - G_i \varepsilon) \rho t)^{\alpha+1} \lambda_i^\alpha \exp[-(\beta + (1 - G_i \varepsilon) \rho t) \lambda_i]}{\Gamma(\alpha+1)} d\lambda_i \right] \times (1 - G_i \varepsilon) \rho \left[\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \right] \left[\frac{\beta^\alpha}{(\beta + (1 - G_i \varepsilon) \rho t)^{\alpha+1}} \right] \\ &= \frac{(1 - G_i \varepsilon) \rho \alpha \beta^\alpha}{(\beta + (1 - G_i \varepsilon) \rho t)^{\alpha+1}}. \end{aligned} \quad (2)$$

Many articles modeling intercourse behavior have shown that it is dependent on various covariates, namely age, parity, and so forth (Zhou et al., 1996; Kim et al., 2010; Sundaram et al., 2012). Our model also incorporates such covariates in the exposure process as well as the self-reported data. An observed rate of intercourse is obtained from the self-reported data for each participant (denoted by ξ_i), and the difference between the observed value and the expected value is adjusted by the baseline covariates Y_i as (3).

$$E(\lambda_i) = \alpha / \beta = \xi_i + Y_i^T \omega. \quad (3)$$

If β is fixed, the shape parameter α is subject-specific as $\alpha_i = (\xi_i + Y_i^T \omega) \beta$. The probability distribution function is now

$$f(t|G_i, \rho, \varepsilon, \alpha_i, \beta) = \frac{(1-G_i\varepsilon)\rho\alpha_i\beta^{\alpha_i}}{(\beta+(1-G_i\varepsilon)\rho t)^{\alpha_i+1}}. \quad (4)$$

The survival function can then be expressed as

$$S(t|G_i, \rho, \varepsilon, \alpha_i, \beta) = \left(\frac{\beta}{\beta+(1-G_i\varepsilon)\rho t} \right)^{\alpha_i}, \quad (5)$$

with the hazard ratio for infection between the intervention and control arms given by

$$r(t|\rho, \varepsilon, \beta) = (1-\varepsilon) \frac{\beta+\rho t}{\beta+(1-\varepsilon)\rho t}. \quad (6)$$

The marginal hazard ratio in (6) is no longer constant over time as with any frailty model.

3. Allowing for Zero Risk

The model presented above assumes that every member of the population is exposed to HIV (so does the Cox model). In fact, many clinical trial participants, although they are sexually active, may not be exposed to HIV throughout the whole study period if their partners are not HIV-infected. These participants will not get infected with HIV within the monitoring time window. To reflect this additional source of heterogeneity, the rate for the exposure process λ_i can be characterized by a zero-inflated gamma distribution such that

$$\lambda_i \sim \begin{cases} 0, & \text{with probability } \phi; \\ \Gamma(\alpha_i, \beta), & \text{with probability } 1-\phi. \end{cases}$$

The parameter ϕ represents the proportion of unexposed subjects in the population, and remains constant over time, reflecting our assumption that exposure to HIV remains steady over time. Accordingly, $T_i = +\infty$ if $\lambda_i = 0$ since participant i is not exposed to HIV throughout the study period. The survival function for the whole population is now given by

$$S_P(t|G_i, \rho, \varepsilon, \phi, \alpha_i, \beta) = \phi + (1-\phi)S(t|G_i, \rho, \varepsilon, \alpha_i, \beta), \quad (7)$$

where $S(t|G_i, \rho, \varepsilon, \alpha_i, \beta)$ is the proper survival function for the non-zero-exposure subjects given by (5). This survival function has a similar expression as some forms of a cure rate model, where ϕ represents the cured fraction.

The population-level hazard ratio of HIV infection at time t under the unexposed rate ϕ is

$$r_P(t|\rho, \varepsilon, \phi, \alpha_i, \beta) = (1-\varepsilon) \frac{\phi(\beta+\rho t)^{\alpha_i+1} + (1-\phi)\beta^{\alpha_i}(\beta+\rho t)}{\phi(\beta+(1-\varepsilon)\rho t)^{\alpha_i+1} + (1-\phi)\beta^{\alpha_i}(\beta+(1-\varepsilon)\rho t)}. \quad (8)$$

This no longer satisfies the proportional hazard assumption (unless ε or ρ is 0). As time of follow-up goes on, the population-level hazard ratio r_P goes to one; hence the population-level effectiveness estimated by the Cox model approaches zero, which deviates from the individual-level effectiveness ε .

4. Bayesian Estimation of the Per-Act Effectiveness

4.1. Likelihood Construction

If we could observe the time of transmission t_i , $i = 1, \dots, N$, directly with a censoring indicator for participants who do not acquire HIV during follow-up ($\delta_i = 1$ means infection was observed, 0 otherwise), the likelihood could be expressed as

$$L(\rho, \varepsilon | \phi, \alpha_i, \beta, (t_i, \delta_i, G_i), i=1, \dots, N) = \prod_{i=1}^N f(t_i | G_i, \rho, \varepsilon, \phi, \alpha_i, \beta)^{\delta_i} S(t_i | G_i, \rho, \varepsilon, \phi, \alpha_i, \beta)^{1-\delta_i}.$$

Instead, we observe the testing times, with the time to infection occurring between the time to last negative test, t_{1i} (equals 0 if there is no negative test) and t_{2i} , the time to the first positive test (equals ∞ if there is no positive test during follow-up). One possible approach is to use the time to the first positive test t_{2i} as the time to infection in the likelihood function above. Alternatively, we could derive the observed likelihood assuming interval censoring, as

$$L_{IC}(\rho, \varepsilon | \phi, \alpha_i, \beta, (t_{1i}, t_{2i}, \delta_i, G_i), i=1, \dots, N) = \prod_{i=1}^N \{F(t_{2i} | G_i, \rho, \varepsilon, \phi, \alpha_i, \beta) - F(t_{1i} | G_i, \rho, \varepsilon, \phi, \alpha_i, \beta)\}.$$

4.2. Prior Distribution Elicitation

Several of the unknown parameters in this model have been studied previously. Powers et al. (2008) undertook a systematic review and meta-analysis of the HIV-1 infectivity per heterosexual sexual act of 27 studies from 15 unique study populations. Boily et al. (2009) extended the above work by adding 43 more publications from 25 different study populations. Powers et al. (2011) also published a modeling study on heterosexual sexual HIV transmission in Lilongwe, Malawi, which includes an estimation of the per-contact transmission probabilities at different stages of infection. All of the works above provide us with some information on ρ .

We define the prior distributions for ρ and ε in a logistic regression model as follows:

$$\log\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 G_i,$$

where P_i is the risk of transmission per exposure for participant i such that $P_i | (G_i = 0) = \rho$

and $P_i | (G_i = 1) = \rho(1 - \varepsilon)$. Equivalently, $\rho = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$ and $\varepsilon = \frac{1 - \exp(\beta_1)}{1 + \exp(\beta_0 + \beta_1)}$. We then set

prior distributions for β_0 and β_1 such that the induced prior distribution for ρ is consistent with the prior information we have, and the prior for ε is non-informative.

The information from the surveillance data or screening data can be borrowed to set the prior distribution for φ . For example, in an HIV prevention trial for women with unknown partner information, the prior for φ can be set as a beta distribution centered at the proportion of HIV-negatives in the population (one minus the HIV prevalence). Note that we assume that each participant can only have sexual intercourse with HIV-positive partners or HIV-negative partners. The partner's HIV status is assumed to be constant within the study period.

A very diffuse prior distribution is used for β and ω in the exposure model (3) as:

$$\begin{cases} \beta \sim \Gamma(0.01, 0.01), \\ \omega \sim N(0, 100). \end{cases} \quad (9)$$

5. Simulation Study

We performed simulation studies to assess the model under various parameter configurations. At each simulation, a sample is generated based upon our model. A sample of size 1500 is generated with 50% to be randomized into the treatment group. Self-reported data and covariates are not considered in the simulation study. For simplicity, we did not include the covariates for α_i and assumed a common α for the whole sample throughout the simulation studies. The exposure process for each subject was generated from a Poisson process with the rate $\lambda_i \sim \Gamma(0.78, 0.01)$. The proportion of non-exposed subjects in each simulated sample, φ , is set to 0.75. All subjects are randomized at time 0 and tested every 90 days until 150 infections have been observed. The data at each simulation include the randomization group indicator, an indicator for whether HIV was detected, the time to the visit when the HIV infection is detected and the time to last visit with negative HIV status. Three values of ρ : 0.2%, 0.7% and 2%, and three values of ε : 0, 0.3, and 0.8 are studied. Under each combination of values for ρ and ε , we replicate the simulation 1000 times.

Table 1 summarizes the simulation results. Generally, the proposed model could obtain an in-line estimate of the per-exposure effectiveness ε . When the intervention does have an effect on preventing the infection ($\varepsilon > 0$) at each exposure, the Cox model actually estimates the overall effectiveness at the population level, rather than the effectiveness at individual exposure ε ; hence a noticeable bias and an insufficient coverage probability are observed in Table 1. As the intervention has more effect with a greater value of ε , or the per-exposure risk of infection without intervention ρ is higher, the more disparity between the Cox estimate and our estimate is observed. As $\varepsilon = 0$ in which the effectiveness at the population level is equal to the effectiveness at the individual exposure level, the Cox model works better. The proposed model loses some efficiency in estimation in terms of a wider credible interval. The inference based on the proposed model is still reasonable, though the point estimates are somewhat biased.

We also investigated the scenarios where the gamma distribution for the rate of exposure process or the value of ϕ are mis-specified. Results are listed in Tables 1 and 2 of Web Appendix A. Mis-specifying the values of α and β in the gamma distribution for the exposure rate does not affect the estimates from the proposed model. Under-specification or over-specification of ϕ does not influence the estimate significantly; ignoring it by setting the value to zero, however, introduces more bias and reduces the coverage probabilities substantially especially when the intervention has a rather strong effect per exposure ($\varepsilon = 0.8$). This series of simulation studies indicates that in the HIV prevention study where the unexposed participants exist, ignoring the unexposed issue will bias the estimate of ε substantially. If the partner information is unknown, a rough approximation for the proportion ϕ could improve the accuracy.

Table 3 of Web Appendix A lists results from an additional simulation study where the fitting model omits the heterogeneity in the per-exposure infectivity ρ in the population. In applications, it is difficult to depict the heterogeneity in ρ . This simulation study ascertains the validity of the results under a common value of ρ .

6. Application: HPTN 035 Data

6.1. Data

We applied the model to HPTN 035 (Abdool Karim et al., 2011), a completed multi-center phase II/IIb clinical trial designed to evaluate the safety and effectiveness of two candidate vaginal microbicides, BufferGel and 0.5% PRO 2000/5 Gel (P), for the prevention of HIV infection in women. To assess the effectiveness of the interventions, study participants were randomly assigned to one of the four arms: BufferGel, 0.5% PRO 2000/5 Gel (P), placebo gel and no gel, and comparisons on the risk of HIV infection were carried out between both candidate products and the placebo gel as well as no gel, respectively. The study enrolled 3099 HIV uninfected women at eight sites, one in the United States and seven in Africa. Abdool Karim et al. (2011) summarized the trial results and concluded that the 0.5% PRO 2000/5 Gel (P) demonstrated a modest 30% reduction in HIV acquisition in women based on the Cox proportional hazard model by site, but the result was not statistically significant.

Participants were seen monthly, however, their HIV status and self-reported sexual behaviors were collected only at quarterly visits. Exploration of the data indicates that the self-reported weekly frequency of sexual acts is consistent with our homogeneous Poisson process assumption in the proposed model. The time unit used here is year, hence we inform ξ_i by multiplying the average of the self-reported weekly frequency by 52. Among the 2137 participants from African sites who were randomized into the arms of 0.5% PRO 2000/5 Gel (P), placebo gel and no gel, the average number of monthly follow-up visits is 20.83 and the average number of quarterly visits is 6.91. 2082 of 2137 participants have reported frequency of vaginal sex at least once. The average weekly frequency is 3.21. The value of ξ for the 55 participants who did not answer this question during follow-up is imputed as the average of the ξ 's of the participants in the site they belong to.

6.2. Analysis

The prior distributions for the main parameters, the per-exposure HIV infectivity without the intervention ρ and the per-exposure effectiveness ε , are configured from the aforementioned logistic regression model. We referred to historic studies to inform the prior distribution for ρ . Powers et al. (2008) reviewed 27 articles from 15 unique study populations and the heterosexual per-act infectivity by the stratified meta-analysis in Africa was 0.091%, and the male-to-female infectivity over all study populations was 0.066%. Boily et al. (2009) extended the meta-analysis by including more study populations and calculated the male-to-female per-act risk of transmission in low-income countries to be 0.3%. A modeling study describing the heterosexual HIV transmission in Lilongwe, Malawi by Powers et al. (2011) estimated the per-contact transmission probability to be around 2–4% in early stage of HIV and 0.07% in asymptotic HIV. Hughes et al. (2012) obtained an estimate of per-act male-to-female risk of transmission as 0.19% from a prospective study of HIV-1-serodiscordant couples. Considering the heterogeneity of the per-act risk of transmission ρ between different populations as discussed by Powers et al. (2008) and Boily et al. (2009), and the heterogeneity between different stages of HIV infection (Powers et al., 2011), we set the mode of the prior distribution for ρ as 0.7% for the HPTN 035 study population in Africa. The prior distribution for our main parameter, the per-exposure treatment effect ε , is non-informative. The prior set (10) embodies such prior information to have the prior mode of ρ as 0.7%.

$$\left\{ \begin{array}{l} \beta_0 \sim N(-4.95, 10), \\ \beta_1 \sim N(0, 100); \\ \rho = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}, \\ \varepsilon = \frac{1 - \exp(\beta_1)}{1 + \exp(\beta_0 + \beta_1)}. \end{array} \right. \quad (10)$$

We set the prior distribution for the unexposed proportion ϕ in each site based on the HIV prevalence data (different from the study population) shown in Table 2. The beta distribution with a mode at the proportion of HIV-negatives is used as the prior distribution for ϕ for each site as suggested in Section 4.2. For example, the model of the prior proportion of ϕ for the Blantyre site is set as $1 - 22.1\% = 77.9\%$. The HIV prevalence for Lusaka, which was not recorded, is imputed as the average of the other sites. We set the 95% percentile around 0.9 to determine the values of both parameters in beta distribution each site. The site-specific priors are listed in Table 2.

6.3. Results

Results were obtained via MCMC using JAGS software (Plummer, 2003) with the rjags package (Plummer, 2011) as the front end in R (R Core Team, 2012).

We compare the effectiveness of 0.5% PRO 2000/5 Gel (P) to both the placebo gel and no gel arms in the African sites. Table 3 lists the number of participants, the person-years of follow-up and the number of events overall and stratified by South African versus non-South African sites. The estimates from the Cox model and the proposed method are listed in Table 4. Overall, the new method produces results that are different from the Cox estimates.

For all the African sites, the 0.5% PRO 2000/5 Gel (P) reduces the risk by 51% compared to no gel, and 26% compared to the placebo gel. The comparison with no gel is now statistically significant at the level of 0.05% since the 95% highest posterior density credible interval excludes zero. The intervention shows less effect on the HIV prevention in the two sites of South Africa (Durban and Hlabisa) with an per-exposure effectiveness as 38%.

The HIV incidence rates are also higher in the two South African sites, especially the Hlabisa site, as shown in Table 2. The posterior estimates of the site-specific unexposed probabilities agree with the pattern from the screening data (Table 4 of Web Appendix B). Excluding the two South African sites, the 0.5% PRO 2000/5 Gel (P) improves the HIV prevention significantly with the risk of infection reduced by 68% compared to the no gel group and 61% compared to the placebo group. The estimates of the per-exposure infectivity ρ in the two controlled arms are around 0.1% among all African sites in the study, which is slightly smaller than the estimate by Hughes et al. (2012).

7. Discussion

In this article, we propose a new model for estimating the effectiveness of interventions in HIV and other STD prevention trials. The proposed model estimates the per-act effectiveness, allowing for heterogeneity in the magnitude of exposure in the study population by incorporating the HIV-exposure path of every participant as an independent mixed Poisson process. The model also allows that a proportion of participants may not be exposed to HIV even with frequent sexual intercourse. The simulation study shows that the estimate of the effectiveness from the proposed model is improved from the traditional Cox model which estimates population-level effectiveness, a quantity that may be impacted by exposure in the population. Not limited to HIV, this model can also be adapted for other STD studies.

Applied to data from HPTN 035, the estimate of the effectiveness of the 0.5% PRO 2000/5 Gel (P) arm versus the no gel arm is statistically significant based on the 95% Bayesian credible interval for all African sites with an effectiveness estimate from the proposed model that is considerably higher than the Cox model estimate. For the two South African sites, where the greatest proportion of participants were exposed to the HIV during the study based on the posterior estimates, the adjusted estimate indicates that the product has a moderately higher effect in preventing HIV infection than the Cox model estimates. For the comparison between the 0.5% PRO 2000/5 Gel (P) and the placebo gel, the difference between the proposed estimates and the Cox estimates is negligible over all African sites and non-South African sites, while the point estimate from the proposed model has an opposite interpretation to the Cox estimate for the two South African sites. This needs further investigation in view of the fact that there might be other factors associated with the effectiveness, such as condom usage, missing follow-up visits, adherence to the assigned intervention, and so forth.

There are two key assumptions in the proposed model: (1) the distribution of sexual frequency is independent of the participant's HIV exposure status; and (2) each participating woman is either exposed or unexposed to HIV throughout follow-up, that is, the ϕ is

presumed to be a constant across all the sexual intercourses. In HIV prevention trials, such as HPTN 035, where high risk women are enrolled independently of their partners and sometimes without their partners' knowledge, it is not possible to know the partners' HIV status. Although a model that allows a woman's exposure status to change over time might be desirable, without the appropriate data, it is not identifiable. However, such a model may be possible in serodiscordant couple studies where dissolution of partnerships is recorded. Violation of these assumptions in our current analysis may bias the estimates of α , β and ϕ , though it should have minimal effect on the estimates of the primary parameters ρ and ε , according to simulation studies (Web Appendix A).

The proposed model could be extended by relaxing the two main assumptions on the exposure process. First, the rate of exposure could be allowed to be time-varying using a heterogeneous Poisson process. Moreover, in practice, ρ may vary over time due to changing characteristics in the woman's vulnerability to HIV and the infectivity of her partner(s). A natural extension, then, would be to allow ρ to vary over time. Additionally, we may want to relate risk to actual adherence to the intervention versus randomization. An instinctive extension is to modify the per-exposure probability of acquiring HIV by $(1 - G_i A_{ij} \varepsilon) \rho$, where A_{ij} is the adherence measure at the j th exposed sexual act for participant i . However, there is no adherence data available for HPTN 035, and even for trials with objective adherence markers, A_{ij} is unknown. Typically, these objective markers are only measured at distinct time points on a strategic subset of participants. Therefore, implementation of this extension requires a model for A_{ij} that can both interpolate adherence between measurement times and impute adherence throughout follow-up for those participants without objective adherence measures. Another extension would allow for a woman's exposure status to change over time based on self-reported partner information. Such extensions on the model need further investigation.

8. Supplementary Materials

Web Appendices referenced in Sections 5 and 6 and the computational code are available with this article at the *Biometrics* website on Wiley Online Library.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIAID grant 5 UM1 AI068615. We are thankful to Dr. James Hughes for the helpful comments. We are also grateful to the Associate editor and the anonymous referee for their insightful comments.

References

- Aalen OO. Heterogeneity in survival analysis. *Statistics in Medicine*. 1988; 7:1121–1137. [PubMed: 3201038]
- Abdool Karim Q, Abdool Karim SS, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, Kharsany ABM, Sibeko S, Mlisana KP, Omar Z, Gengiah TN, Maarschalk S, Arulappan N, Mlotshwa M, Morris L, Taylor D. on behalf of the CAPRISA 004 Trial Group. Effectiveness and safety of

tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science*. 2010; 329:1168–1174. [PubMed: 20643915]

Abdool Karim SS, Richardson BA, Ramjee G, Hoffman IF, Chirenje ZM, Taha T, Kapina M, Maslankowski L, Coletti A, Profy A, Moench TR, Piwowar-Manning E, Mâsse B, Hillier SL, Soto-Torres L. on behalf of the HIV Prevention Trials Network (HPTN) 035 Study Team. Safety and effectiveness of BufferGel and 0.5% PRO2000 gel for the prevention of HIV infection in women. *AIDS*. 2011; 25:957–966. [PubMed: 21330907]

Auvert B, Sitta R, Zarca K, Mahiane S, Pretorius C, Lissouba P. The effect of heterogeneity on HIV prevention trials. *Clinical Trials (London, England)*. 2011; 8:144–154.

Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, Tappero JW, Bukusi EA, Cohen CR, Katabira E, Ronald A, Tumwesigye E, Were E, Fife KH, Kiarie J, Farquhar C, John-Stewart G, Kakia A, Odoyo J, Mucunguzi A, Nakku-Joloba E, Twesigye R, Ngure K, Apaka C, Tamoo H, Gabona F, Mujugira A, Panteleeff D, Thomas KK, Kidoguchi L, Krows M, Revall J, Morrison S, Haugen H, Emmanuel-Ogier M, Ondrejcek L, Coombs RW, Frenkel L, Hendrix C, Bumpus NN, Bangsberg D, Haberer JE, Stevens WS, Lingappa JR, Celum C. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *New England Journal of Medicine*. 2012; 367:399–410. [PubMed: 22784037]

Barrett JC, Marshall J. The risk of conception on different days of the menstrual cycle. *Population Studies*. 1969; 23:455–461. [PubMed: 22073960]

Blower S, Boe C. Sex acts, sex partners, and sex budgets: Implications for risk factor analysis and estimation of HIV transmission probabilities. *Journal of Acquired Immune Deficiency Syndromes*. 1993; 6:1347–52. [PubMed: 8254474]

Boily MC, Baggaley RF, Wang L, Masse B, White RG, Hayes RJ, Alary M. Heterosexual risk of HIV-1 infection per sexual act: Systematic review and meta-analysis of observational studies. *The Lancet Infectious Diseases*. 2009; 9:118–129. [PubMed: 19179227]

CONRAD. ClinicalTrials.gov [Internet]. Bethesda, MD: National Library of Medicine (US); 2011. Safety and effectiveness of Tenofovir gel in the prevention of human immunodeficiency virus (HIV-1) infection in women and the effects of Tenofovir gel on the incidence of herpes simplex virus (HSV-2) infection. Available from: <http://clinicaltrials.gov/show/NCT01386294> NLM Identifier: NCT01386294 [accessed January 1, 2012]

Cox, DR.; Oakes, D. *Analysis of Survival Data*. New York: Chapman and Hall; 1984.

Grandell, J. *Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series*. Chapman & Hall; New York: 1997. *Mixed Poisson Processes*.

Hughes JP, Baeten JM, Lingappa JR, Magaret AS, Wald A, de Bruyn G, Kiarie J, Inambao M, Kilembe W, Farquhar C, Celum C. the Partners in Prevention HSV/HIV Transmission Study Team. Determinants of per-coital-act HIV-1 infectivity among African HIV-1-serodiscordant couples. *Journal of Infectious Diseases*. 2012; 205:358–365. [PubMed: 22241800]

International Partnership for Microbicides, Inc. ClinicalTrials.gov [Internet]. Bethesda, MD: National Library of Medicine (US); 2012. Safety and efficacy trial of a dapivirine vaginal matrix ring in healthy HIV-negative women. Available from: <http://clinicaltrials.gov/show/NCT01539226> NLM Identifier: NCT01539226 [accessed January 2, 2012]

International Partnership for Microbicides, Inc. and National Institutes of Health (NIH). ClinicalTrials.gov [Internet]. Bethesda, MD: National Library of Medicine (US); 2012. Phase 3 safety and effectiveness trial of dapivirine vaginal ring for prevention of hiv-1 in women (ASPIRE). Available from: <http://clinicaltrials.gov/show/NCT01617096> NLM Identifier: NCT01617096 [accessed January 2, 2012]

Joint United Nations Programme on HIV/AIDS (UNAIDS). [accessed January 11, 2013] Global report: UNAIDS report on the global AIDS epidemic: 2012. 2012. Available at: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_en.pdf

Kim S, Sundaram R, Buck Louis GM. Joint modeling of intercourse behavior and human fecundability using structural equation models. *Biostatistics*. 2010; 11:559–571. [PubMed: 20173100]

McCormack S, Ramjee G, Kamali A, Rees H, Crook AM, Gafos M, Jentsch U, Pool R, Chisembele M, Kapiga S, Mutemwa R, Vallely A, Palanee T, Sookrajh Y, Lacey CJ, Darbyshire J, Grosskurth H, Profy A, Nunn A, Hayes R, Weber J. PRO2000 vaginal gel for prevention of HIV-1 infection

- (Microbicides Development Programme 301): A phase 3, randomised, double-blind, parallel-group trial. *The Lancet*. 2010; 376:1329–1337.
- National Institute of Allergy and Infectious Diseases (NIAID). ClinicalTrials.gov [Internet]. Bethesda, MD: National Library of Medicine (US); 2008. Safety and effectiveness of tenofovir 1% gel, tenofovir disoproxil fumarate, and emtricitabine/tenofovir disoproxil fumarate tablets in preventing hiv in women. Available from: <http://clinicaltrials.gov/show/NCT00705679> NLM Identifier: NCT00705679 [accessed January 1, 2012]
- O'Hagan J, Hernán M, Walensky R, Lipsitch M. Apparent declining efficacy in randomized trials: examples of the Thai RV144 HIV vaccine and South African CAPRISA 004 microbicide trials. *AIDS* (London, England). 2012; 26:123–126.
- Plummer, M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria. 2003.
- Plummer, M. R package version 3–5. 2011. rjags: Bayesian Graphical Models Using MCMC.
- Powers KA, Ghani AC, Miller WC, Hoffman IF, Pettifor AE, Kamanga G, Martinson FE, Cohen MS. The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: A modelling study. *The Lancet*. 2011; 378:256–268.
- Powers KA, Poole C, Pettifor AE, Cohen MS. Rethinking the heterosexual infectivity of HIV-1: A systematic review and meta-analysis. *The Lancet Infectious Diseases*. 2008; 8:553–563. [PubMed: 18684670]
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2012.
- Rhodes PH, Halloran ME, Longini J, Ira M. Counting process models for infectious disease data: Distinguishing exposure to infection from susceptibility. *Journal of the Royal Statistical Society, Series B*. 1996; 58:751–762.
- Ross, SM. Wiley. Stochastic Process. New York: Wiley; 1995.
- Satten GA, Mastro TD, Longini IM. Modelling the female-to-male per-act HIV transmission probability in an emerging epidemic in Asia. *Statistics in Medicine*. 1994; 13:2097–2106. [PubMed: 7846413]
- Skoler-Karppoff S, Ramjee G, Ahmed K, Altini L, Plagianos MG, Friedland B, Govender S, Kock AD, Cassim N, Palanee T, Dozier G, Maguire R, Lahteenmaki P. Efficacy of Carraguard for prevention of HIV infection in women in South Africa: A randomised, double-blind, placebo-controlled trial. *The Lancet*. 2008; 372:1977–1987.
- Sundaram R, McLain AC, Buck Louis GM. A survival analysis approach to modeling human fecundity. *Biostatistics*. 2012; 13:4–17. [PubMed: 21697247]
- Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM, Henderson FL, Pathak SR, Soud FA, Chillag KL, Mutanhaurwa R, Chirwa LI, Kasonde M, Abebe D, Buliva E, Gvetadze RJ, Johnson S, Sukalac T, Thomas VT, Hart C, Johnson JA, Malotte CK, Hendrix CW, Brooks JT. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *New England Journal of Medicine*. 2012; 367:423–434. [PubMed: 22784038]
- Van Damme L, Ramjee G, Alary M, Vuylsteke B, Chandeying V, Rees H, Sirivongrangson P, Tshibaka LM, Ettiègne-Traoré V, Uaheowitchai C, Karim SSA, Mâsse B, Perriens J, Laga M. Effectiveness of COL-1492, a nonoxynol-9 vaginal gel, on HIV-1 transmission in female sex workers: A randomised controlled trial. *The Lancet*. 2002; 360:971–977.
- Vittinghoff E, Douglas J, Judon F, McKiman D, MacQueen K, Buchinder SP. Per-contact risk of human immunodeficiency virus transmission between male sexual partners. *American Journal of Epidemiology*. 1999; 150:306–311. [PubMed: 10430236]
- Wilson DP. Interpreting sexually transmissible infection prevention trials by adjusting for the magnitude of exposure. *Clinical Trials*. 2010; 7:36–43. [PubMed: 20007407]
- Zhou H, Weinberg CR, Wilcox AJ, Baird DD. A random-effects model for cycle viability in fertility studies. *Journal of the American Statistical Association*. 1996; 91:1413–1422.

Table 1

Summary of the simulation study based on 1000 replicates. Statistics of the estimated effectiveness ε listed for each estimation model are the mean, median (Q_{50}), mean squared error (MSE) and coverage probability (CP%) of the 95% confidence intervals for the Cox model and 95% highest posterior density credible intervals for the other two models.

ρ	ε	Cox estimate				Proposed model			
		Mean	Q_{50}	MSE	CP%	Mean	Q_{50}	MSE	CP%
0.2%	0	-0.01	0.002	0.03	94.9	-0.06	-0.03	0.07	96.3
	0.3	0.22	0.23	0.02	90.9	0.26	0.29	0.03	96.8
	0.8	0.70	0.70	0.01	39.4	0.80	0.80	0.002	95.6
0.7%	0	-0.01	0.01	0.03	94.4	-0.06	-0.03	0.07	95.2
	0.3	0.22	0.23	0.02	92.1	0.27	0.29	0.03	95.4
	0.8	0.69	0.69	0.02	36.8	0.79	0.80	0.003	94.3
2%	0	-0.01	0.01	0.02	95.5	-0.06	-0.02	0.07	95.8
	0.3	0.20	0.21	0.03	87.5	0.24	0.27	0.04	96.3
	0.8	0.68	0.68	0.02	25.4	0.78	0.79	0.003	96.0

Table 2

HIV prevalence and incidence rates by site in HPTN 035 data

Site	HIV prevalence (%)	HIV incidence (per 100 person-year)	Prior distribution
Blantyre	22.1	3.67	Beta(12, 4)
Lilongwe	18.0	1.42	Beta(28, 7)
Durban	21.6	4.60	Beta(12, 4)
Hlabisa	28.2	9.10	Beta(6, 3)
Lusaka (Kamwala)	Not available	4.10	Beta(12, 4)
Chitungwiza	18.8	2.45	Beta(23, 6)
Harare	21.4	2.49	Beta(12, 4)

Table 3

Summary of the sample size, person-years of follow-up and number of HIV infections for the analyses

	0.5% PRO 2000/5 Gel (P)				Placebo			No gel		
	Sample size	Person-years	No. of HIV		Sample size	Person- years	No. of HIV	Sample size	Person-years	No. of HIV
Africa	721	1228	36		720	1201	49	721	1209.75	53
South Africa	261	474.75	27		262	473	28	262	470.75	30
Non-South Africa	460	753.25	9		458	728	21	459	739	23

Table 4

Summary of the posterior distribution of ρ and ε : posterior mean, posterior standard deviation, and the 95% highest posterior density interval

	Cox estimate	Proposed model	
	ε	ρ	ε
0.5% PRO 2000/5 Gel (P) versus no gel			
African sites	0.33 (0.14) (−0.02, 0.56)	0.0011 (0.0003) (0.0006, 0.0017)	0.51 (0.12) (0.26, 0.73)
South African sites	0.12 (0.23) (−0.49, 0.47)	0.0022 (0.0009) (0.0008, 0.0041)	0.38 (0.22) (−0.06, 0.75)
Non-South African sites	0.61 (0.15) (0.16, 0.82)	0.0009 (0.0003) (0.0004, 0.0015)	0.68 (0.14) (0.40, 0.90)
0.5% PRO 2000/5 Gel (P) versus placebo gel			
African sites	0.27 (0.16) (−0.13, 0.52)	0.0008 (0.0002) (0.0005, 0.0012)	0.26 (0.18) (−0.11, 0.60)
South African sites	0.02 (0.26) (−0.66, 0.42)	0.0010 (0.0003) (0.0004, 0.0017)	−0.13 (0.38) (−0.88, 0.51)
Non-South African sites	0.59 (0.17) (0.09, 0.81)	0.0008 (0.0003) (0.0004, 0.0014)	0.61 (0.17) (0.28, 0.88)