Published in final edited form as:

Stat Med. 2016 July 10; 35(15): 2609–2634. doi:10.1002/sim.6884.

Estimating effectiveness in HIV prevention trials with a Bayesian hierarchical compound Poisson frailty model

Rebecca Yates Coley^{a,*} and Elizabeth R. Browne^{a,b}

^aDepartment of Biostatistics, University of Washington, Seattle, WA, U.S.A

^bFred Hutchinson Cancer Research Center, Seattle, WA, U.S.A

Abstract

Inconsistent results in recent HIV prevention trials of pre-exposure prophylactic interventions may be due to heterogeneity in risk among study participants. Intervention effectiveness is most commonly estimated with the Cox model, which compares event times between populations. When heterogeneity is present, this population-level measure underestimates intervention effectiveness for individuals who are at risk. We propose a likelihood-based Bayesian hierarchical model that estimates the individual-level effectiveness of candidate interventions by accounting for heterogeneity in risk with a compound Poisson-distributed frailty term. This model reflects the mechanisms of HIV risk and allows that some participants are not exposed to HIV and, therefore, have no risk of seroconversion during the study. We assess model performance via simulation and apply the model to data from an HIV prevention trial.

Keywords

Bayesian analysis; data augmentation; heterogeneity; latent variable; survival analysis

1. Introduction

Slowing the spread of HIV remains a major focus of public health research. Of particular concern is the heterosexual transmission of the virus in Sub-Saharan Africa, which accounts for more than half of all new infections worldwide [1]. Attention has recently turned to the possibility that anti-retroviral (ARV) treatments, known to be effective in treating the virus, may also prevent infection in this population. Several trials have shown topical microbicides and oral chemoprophylaxis to be effective in reducing HIV transmission [2, 3, 4]. However, other trials of similar interventions have found no effect [5, 6, 7, 8]. While these inconsistent results are often attributed to low adherence in some studies, heterogeneity in risk of seroconversion among study participants offers another possible explanation. A study participant's risk of seroconversion depends on the number and type of sex acts with infected partners, infected partners' viral load, and other factors that will vary among individuals. Furthermore, in general population studies, some participants will not engage in sexual activity with an HIV-infected person during the trial and, therefore, will have no risk

*Correspondence to: R. Yates Coley, Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, U.S.A.. rycoley@gmail.com.

of seroconversion. Even in serodiscordant partner studies, a participant may have essentially no risk of seroconversion if their partner's viral load is very low or if they always use condoms.

In most HIV prevention trials, intervention effectiveness is estimated with the Cox model, which assumes proportional hazards between treatment and control groups [9]. However, when heterogeneity in risk is present and an intervention acts multiplicatively on an individual's risk, proportionality does not hold on the population level, as explained by Aalen [10] and explored in relation to HIV prevention trials in Auvert *et al.* [11], and O'Hagan *et al.* [12]. While randomization typically guarantees similar risk profiles across trial arms at baseline, the distribution of risk across arms will not remain similar as the study progresses and events accrue if the intervention has a non-null effect [13]. Specifically, if the intervention reduces risk of seroconversion, the intervention arm will contain, on average, participants with higher risk. As the proportion of individuals at higher risk in the control arm is depleted, the intervention appears less effective and, as a result, the Cox model's population-level estimate underestimates the individual-level effectiveness. In this sense, the Cox estimate of effectiveness is a biased estimate of the measure we are most interested in: the reduction in risk for an at-risk individual receiving the intervention. In fact, as examined by Aalen [10], when some portion of the study population is not at risk of an event, an intervention that is beneficial on the individual level may appear ineffective, or even harmful, on the population level. This difference is particularly alarming in the case of HIV prevention trials, where we expect many participants will not be at risk. Moreover, instead of estimating an average treatment effect, we may be more interested in quantifying how effective the intervention is for at-risk individuals.

There are several approaches that accommodate heterogeneity in risk in time-to-event models. The most common is to include risk factors as covariates in a Cox regression model. Unfortunately, participant-reported information on sexual behavior is unreliable, while other predictors, like the HIV status of a participant's partners, are rarely collected. Even if this information were available, it would still be necessary to assume proportional effects of covariates. Furthermore, it would be difficult to constrain the model so as to guarantee that near-zero risk (due to non-exposure) would be predicted for some individuals. Hence, this model would not allow for the possibility that a portion of the study participants may actually have no risk of an event.

Frailty models offer another approach to account for heterogeneity in risk that does not rely on self-reported data. A frailty term is a random effect added to a survival model to account for unobserved heterogeneity [14]. Frailty modeling has been used in the analysis of HIV vaccine efficacy, as in Longini and Halloran [15], but otherwise remains absent from HIV prevention research. Frailty is commonly modeled with a gamma distribution, but this choice is typically for the sake of mathematical convenience, not based on biological rationale [16]. We prefer to select a frailty distribution based on the sources of heterogeneity in HIV risk.

In this paper, we propose a model for risk of seroconversion that is mechanistically motivated and reflects several aspects of HIV transmission. We outline a conceptual framework for the proposed model, which is presented more formally in Section 2: First, an

individual must have sexual contact with an HIV positive partner to have any risk of seroconversion. This contact occurs in the context of that individual's patterns of sexual behavior, which are not fully observed. Second, the risk of seroconversion is not equal for each exposure or sexual partnership, rather, it depends on the number and type of acts, viral load of the infected partner(s), and other factors that may be unobserved. Here, we consider separate behavior patterns to be distinct exposure processes that present constant risk over time. For example, unprotected vaginal sex with an infected partner constitutes one exposure process, while unprotected anal sex with the same partner would be another. We model the number of exposure processes as a Poisson random variable, which allows for a participant to have no risk (i.e., zero exposures) of seroconversion. For individuals with nonzero exposure processes, the risk associated with each exposure process is modeled by a random variable distributed continuously over the positive half-line. Finally, the risks associated with each exposure process are independent competing risks; given exposure, one of the processes eventually leads to seroconversion. It follows that an individual's risk of seroconversion during a specified interval of time is the sum of these competing risks. This risk model aligns with the construction of a compound Poisson random variable. The use of the compound Poisson distribution as a frailty distribution in survival analysis was first suggested by Aalen [10]. The compound Poisson frailty model has been used in many applications but has yet to be applied to HIV prevention research [17, 18, 19, 20, 21, 22, 23, 24].

This paper is also the first to demonstrate Bayesian estimation of a compound Poisson frailty model with univariate survival data. In particular, we propose a novel data augmentation MCMC algorithm for posterior sampling of frailty components. Development of Bayesian frailty methods has more recently focused on multivariate or clustered survival data, arising from the observation of repeated events for a single sampling unit, as in Pennell and Dunson [25], or from the assumption of a shared random effect across multiple units, as is frequently done in spatial modeling for observations within the same geographic region, as in Banerjee, Wall, and Carlin [26] and Zhou *et al.* [27]. Chen, Ibrahim, and Sinha [28] and Yin [29, 30] address the possibility that all units within a cluster have no risk of an event by incorporating a cured fraction into multivariate frailty models. Bayesian approaches to univariate survival data have frequently used joint modeling in order to leverage information available in the trajectory of longitudinal measurements of biomarkers to learn about heterogeneity in risk [31, 32]. In HIV prevention trials, however, neither longitudinal measurements of relevant biomarkers nor reliable self-reported risk data are available, making a frailty model most appropriate.

The paper proceeds as follows. First, we propose a hierarchical definition of the compound Poisson frailty model and outline Bayesian model estimation procedures. Then, model performance is evaluated via simulation. Finally, we illustrate the use of this model to estimate the individual-level effectiveness of an intervention in an HIV prevention trial.

2. A New Bayesian Compound Poisson Frailty model

In this section, we first summarize the notation used in the subsequent model presentation. Next, we present a hierarchical definition of the compound Poisson distribution for frailty

before describing a survival model with frailty and defining the likelihood for the proposed model. We complete Bayesian model specification by discussing prior selection and outlining an appropriate algorithm for posterior sampling.

2.1. Notation

For each study participant, indexed by $i = 1, \dots, n$, we observe (x_i, t_i, δ_i) where x_i is an indicator of randomization to the trial's intervention, t_i is the potentially censored event time, and δ_i is an indicator variable equaling one if seroconversion is observed and zero otherwise. For the survival model, let β be the log-hazard ratio (HR) associated with randomization to the intervention and $\lambda(t)$ denote the baseline hazard function common to all participants. The latent variables defining each individual's unobserved risk are $(N_i, U_{i1}, \dots, U_{iN_i}, Z_i)$ where N_i is the number of exposure processes, $U_{i1}, \dots, U_{iN_i}$ denote the risk associated with each exposure process, and $Z_i = U_{i1} + \dots + U_{iN_i}$ is the total risk or frailty.

2.2. Hierarchical Definition of Compound Poisson Distribution

Here, we provide a hierarchical definition of the compound Poisson distribution for frailty that aligns with the conceptual framework of HIV risk given in Section 1.

First, the level of exposure to HIV for an individual is the result of that individual's patterns of sexual behavior. Separate behavioral patterns, e.g. unprotected vaginal sex with an infected partner or anal sex with the same partner, each establish distinct exposure processes for that individual. To reflect this aspect of HIV risk, we model the number of exposure processes for individual i , N_i , as a Poisson-distributed random variable with mean ρ : $N_i \sim \text{Poisson}(\rho)$.

Next, the risk posed by each exposure process can vary. We define the risk associated with the i th participant's m th exposure process at time t as equal to $\lambda(t)U_{im}$ where $\lambda(t)$ is the baseline hazard function and $U_{im} \sim \text{Gamma}(\eta, \nu)$. (Here, $\text{Gamma}(\eta, \nu)$ is the gamma distribution with shape parameter η and rate parameter ν .)

Multiple exposure processes for an individual pose competing risks. Assuming these competing risks are independent, the risk of seroconversion is additive across exposure

processes such that participant i 's risk of seroconversion at time t is $\lambda(t) \sum_{m=1}^{N_i} U_{im}$. We define an individual's frailty, Z_i , such that:

$$Z_i = \begin{cases} \sum_{m=1}^{N_i} U_{im} \sim \text{Gamma}(N_i \eta, \nu) & \text{if } N_i > 0 \\ 0 & \text{if } N_i = 0. \end{cases} \quad (1)$$

Formally, Z_i follows a compound Poisson distribution.

By convention, the expected value of Z_i is set equal to one so that $\lambda(t)$ is interpreted as the average baseline hazard. Furthermore, assuming a finite mean for the frailty distribution, in addition to the assumption of proportional hazards for covariates x_i , ensures identifiability of frailty model parameters, even in the absence of proper prior specification [33, 34]. We set

the marginal expectation over all Z_i , $E(E(Z_i|N_i)) = \rho\eta/\nu$, equal to one by defining ν in terms of the other frailty distribution parameters, ρ and η : $\nu = \rho\eta$.

We also note that $P(Z_i > 0)$, i.e., the probability that the frailty for a randomly selected individual is greater than zero, is equal to $P(N_i > 0) = 1 - \exp(-\rho)$. In the context of HIV prevention trials, this probability corresponds to the proportion of participants with some positive risk of seroconversion during the study. This identity can be exploited in specifying priors for ρ or, as demonstrated in Section 4.1, the baseline hazard function.

Finally, there are a number of assumptions inherent in this conceptualization of HIV risk. In particular, we assume that each of an individual's exposure processes modifies her baseline hazard by a constant amount during the study period, U_{im} . While we do not expect a participant's frailty to be constant, this assumption is necessary for identifiability of individual-level frailty in univariate survival models (when at most one event is observed per person). This assumption is considered further in Section 5. The proposed formulation of frailty also assumes that the competing risks associated with each exposure process are independent, meaning, first, that each exposure process acts independently to lead to infection— a valid assumption for HIV infection— and, second, that the risks associated with each process are independent and identically distributed conditional on observed covariates. This second assumption would be reasonable, for example, in a serodiscordant partners study where the infected partner's viral load could be included in the proportional hazards function. We discuss other potential approaches for adjusting for covariates in Section 5. We further note that the distributional assumptions on the number of exposure processes, N_i , and the risk associated with each, $U_{i1}, \dots, U_{iN_i}$ may not be appropriate in every context but are more flexible than the typical assumption of gamma-distributed frailty.

2.3. Survival Model with Frailty

We define the i th participant's hazard function as $\lambda_i(t; x_i) = \lambda(t) Z_i \exp(x_i\beta)$ where β is the log-HR associated with randomization to the intervention group, indicated by $x_i = 1$. It follows that the subject-specific hazard density (2) and survival functions (3) for participants at risk are:

$$f(t|x_i, Z_i) = \lambda(t) Z_i \exp(x_i\beta) S(t|x_i, Z_i) \quad (2)$$

$$S(t|x_i, Z_i) = \exp(-Z_i \exp(x_i\beta) \Lambda(t)), \quad (3)$$

where $\Lambda(t) = \int_0^t \lambda(t) dt$ is the cumulative baseline hazard at time t . For an individual with no risk of an event, $Z_i = 0$, the hazard function necessarily equals zero and the survival time distribution is a point mass at infinity.

We make a few remarks regarding parameter interpretation and model specification before proceeding to define the model likelihood. First, while this model allows for individual-level variability in risk, as represented by Z_i , it does not account for any heterogeneity in the

effect of the treatment, that is, β is assumed to be the same for all participants at risk of seroconversion. Furthermore, the interpretation of intervention effectiveness, $1 - \exp(\beta)$, differs between the Cox model and compound Poisson frailty model. While, in the Cox model, the estimate of effectiveness refers to the reduction in risk for a population receiving the intervention versus one that does not, the inclusion of a compound Poisson-distributed frailty term changes this interpretation. Instead, in the proposed model, effectiveness is a measure of risk reduction for an individual at risk of an event. Participants with no risk do not contribute to the estimate.

In contrast, interpretation of the baseline hazard function in the compound Poisson model includes all individuals. Given that the hazard at time t for individual i is $Z_i\lambda(t)$ and that the expectation of Z_i is one, the baseline hazard, $\lambda(t)$, represents the average baseline hazard at time t among all participants, including those with no risk (for whom $Z_i = 0$). It is possible, however, to calculate the baseline hazard among only those individuals with some risk of an event. The average baseline hazard among participants at risk is a function of $\lambda(t)$ and the proportion of the population at risk: $E(Z_i\lambda(t) | Z_i > 0) = \lambda(t)/(1 - \exp(-\rho))$. (We will use this property when specifying priors for our application in Section 4.1.)

The form of the baseline hazard in this model's presentation is intentionally general and would accommodate a variety of specifications, a few of which are discussed here. Specifying a constant baseline hazard function, for example, assumes that each participant has a constant risk of an event throughout the trial with the level of risk for individuals varying according to a compound Poisson distribution. Many previous applications of the compound Poisson frailty distribution have modeled the baseline hazard with a Weibull function [10, 17, 18, 19, 20, 21, 22], which assumes a monotonic trend in hazard for an individual. The Weibull function is a common choice, both due to its mathematical convenience and its clinical plausibility in many medical applications (for e.g., the Armitage-Doll multistage model of cancer [37, 38]). If we do not want to make assumptions about the form of the baseline hazard, less parametric alternatives include a piecewise constant hazard function, spline representation of the baseline hazard, or a Gamma process model [39]. When considering these and other options, the form of the baseline hazard function should be chosen based on scientific rational and supported by examination of the data.

Lastly, in practice, we would consider stratifying the model by site, as is typically done for analysis of data from multi-site clinical trials. Site-stratification would involve estimating the baseline hazard, $\lambda(t)$, as well as the parameters of the frailty distribution, ρ and η , separately for each study site. In addition to following convention, site-stratification would likely improve heterogeneity modeling as we would expect the probability and magnitude of risk to vary across sites in an HIV prevention trial.

2.4. Likelihood

Having defined an appropriate survival model and frailty distribution, we can define the model likelihood given time-to-event data:

$$\begin{aligned}
& L(\beta, \lambda(t), \rho, \eta | (t_i, \delta_i, x_i, N_i, Z_i), i \\
& \quad = 1, \dots, n) \\
& = \prod_{i=1}^n f(N_i | \rho) f(Z_i | N_i, \rho, \eta) f(t_i, \delta_i | x_i, Z_i, \beta, \lambda(t)) \\
& = \prod_{i=1}^n f(N_i | \rho_i) [f(Z_i | N_i, \rho, \eta) f(t_i, \delta_i | x_i, Z_i, \beta, \lambda(t))]^{1_{[N_i > 0]}} \\
& \quad \times [f(Z_i | N_i, \rho, \eta) f(t_i, \delta_i | x_i, Z_i, \beta, \lambda)]^{1_{[N_i = 0]}} \\
& = \prod_{i=1}^n \frac{\rho^{N_i}}{\Gamma(N_i + 1)} \exp(\\
& \quad - \rho) \left[\frac{\rho \eta^{N_i}}{\Gamma(N_i \eta)} Z_i^{N_i \eta - 1} \exp(-Z_i \rho \eta) (\lambda(t) Z_i \exp(x_i \beta))^{\delta_i} \exp(-Z_i \exp(x_i \beta) \Lambda(t_i)) \right]^{1_{[N_i > 0]}} \\
& \quad \times [\Delta_0(Z_i) \Delta_\infty(t_i)]^{1_{[N_i = 0]}},
\end{aligned}
\tag{4}$$

where $\Delta_0(Z_i)$ and $\Delta_\infty(t_i)$ indicate point mass for frailty, Z_i , at zero and time-to-event, t_i , at infinity for participant i with $N_i = 0$, i.e., no risk.

2.5. Prior Elicitation

To complete Bayesian specification of the proposed model, we select prior distributions for model parameters. In a clinical trial setting, a non-informative prior on an intervention's effectiveness is frequently an appropriate choice. Here, we choose a diffuse normal prior for β : $\pi(\beta) = N(M_\beta, \Sigma_\beta^2)$, where $N(M, \Sigma^2)$ is the normal distribution with mean M and variance Σ^2 .

There are several reasonable priors for ρ , the mean parameter for the number of exposure processes. If proper knowledge about the number of relevant exposure processes in the study population was available, that information could be incorporated into a gamma prior for ρ : $\pi(\rho) = \text{Gamma}(O_\rho, T_\rho)$. Alternatively, the prior on ρ could be specified to reflect prior knowledge about the proportion at risk in the study population. Recall that the proportion of participants at risk, which we denote $P(\text{Risk})$, is also a function of ρ : $P(\text{Risk}) = P(N_i > 0) = 1 - \exp(-\rho)$. We specify $P(\text{Risk}) \sim \text{Beta}(A, B)$, where $\text{Beta}(A, B)$ is the beta distribution with shape parameters A and B . The prior distribution for ρ is then:

$$\pi(\rho) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} (1 - \exp(-\rho))^{A-1} \exp(-\rho)^B.$$

Setting $A = 1$ and $B = 1$ is equivalent to a Uniform(0,1) distribution on the proportion at risk, $P(\text{Risk})$, and an exponential prior with mean one on the expected number of exposure processes, ρ . This prior is appropriate when we do not have, or do not want to specify, any

information about the proportion at risk or the number of risk processes for participants. A more informative prior for the proportion at risk could be determined using trial data. For example, in a general population study, screening seroprevalence could be used. We consider this option more closely in Section 4.1

The gamma distribution has its support on the positive half-line and, thus, is a natural choice for the prior distribution on η , which determines the level of risk associated with each exposure process: $\pi(\eta) = \text{Gamma}(O_\eta, T_\eta)$. Data from previous studies could also be used to inform the prior since the number and timing of events observed would be more likely under certain values of η .

Prior selection for the baseline hazard function will depend on the form chosen for $\lambda(t)$. For example, under the assumption of a constant baseline hazard, $\lambda(t) = \lambda$ for all t , the gamma distribution is a conjugate prior, and we set $\pi(\lambda) = \text{Gamma}(O_\lambda, T_\lambda)$. Similarly, when the baseline hazard is assumed to be piecewise constant, we place independent gamma priors on the hazard constant for each of K intervals: $\pi(\lambda_k) = \text{Gamma}(O_{\lambda_k}, T_{\lambda_k})$, $k = 1, \dots, K$. A gamma process prior is another appropriate choice for a piecewise constant baseline hazard. It is also possible to specify a prior on the baseline hazard that is correlated with the prior on the proportion at risk. We present details about a correlated prior selected for our particular application in Section 4.1.

2.6. Posterior Estimation Procedure

After specifying priors for all model parameters, we can define the joint posterior density of the parameters and latent variables as the product of each parameter's prior density and the model likelihood in (4):

$$\mathbf{p}(\beta, \lambda(t), \rho, \eta, \mathbf{N}, \mathbf{Z} | ((t_i, \delta_i, x_i), i=1, \dots, n), \Theta) \propto L(\beta, \lambda(t), \rho, \eta | (t_i, \delta_i, x_i, N_i, Z_i), i=1, \dots, n) \times \pi(\beta, \lambda(t), \rho, \eta | \Theta) \quad (5)$$

where $\mathbf{p}(\cdot | \text{data})$ denotes the joint posterior density, $\pi(\cdot | \Theta)$ denotes the joint prior density for model parameters with hyperparameters Θ , and $\mathbf{N}$ and $\mathbf{Z}$ are vectors of latent variables $[N_1, \dots, N_n]$ and $[Z_1, \dots, Z_n]$.

Posterior sampling of model parameters and latent variables follows a novel data augmentation MCMC algorithm [40] developed for this application, which we describe here. Posterior updating alternates between, first, generating posterior samples of missing data—latent variables (N_i, Z_i) —and, second, generating posterior samples of parameters for the survival model with frailty— $\rho, \eta, \lambda(t)$ and β . For the latter step, slice sampling [41] is used to sample from each parameter's full conditional posterior when conditionally conjugate priors are not available. (Full conditional posteriors are obtained from Equation (5) and explicitly given in the Appendix.)

At each iteration of the proposed data augmentation algorithm, the unobserved number of exposure processes, N_i , and frailty, Z_i , are sampled for all i . Instead of sampling each pair of (N_i, Z_i) simultaneously from the joint full conditional posterior— an approach which would require slower numerical methods for all i at each sampling iteration— we improve efficiency and simplify sampling with a blocking approach in which we first marginalize over Z_i to sample N_i before sampling Z_i from its full conditional [42]. In this way, we are able to sample the number of exposure processes, N_i , without conditioning on total frailty, Z_i , which would preclude sampling zero exposure processes for an individual with a current sample of positive frailty or, conversely, sampling a positive number of exposure processes for an individual with a current sample of zero frailty.

The full conditional posterior for N_i with Z_i integrated out is

$$p(N_i | \rho, \eta, \lambda(t), \beta, t_i, \delta_i, x_i) \propto \frac{\rho^{N_i}}{\Gamma(N_i+1)} \int_0^\infty \left(\frac{\rho\eta^{N_i\eta}}{\Gamma(N_i\eta)} Z_i^{N_i\eta-1} \exp(-Z_i\rho\eta) \right. \\ \left. (\lambda(t)Z_i \exp(x_i\beta))^{\delta_i} \exp(-Z_i \exp(x_i\beta)\Lambda(t_i)) \right) dZ_i \\ \propto \frac{\rho^{N_i}}{\Gamma(N_i+1)} \left(\frac{\rho\eta}{\rho\eta + \exp(x_i\beta)\Lambda(t_i)} \right)^{N_i\eta} \left(\frac{N_i\eta}{\rho\eta + \exp(x_i\beta)\Lambda(t_i)} \right)^{\delta_i}. \quad (6)$$

Equation (6) is used to obtain valid posterior samples for all N_i . Sampling for each participant can be performed most efficiently with separate approaches contingent upon whether seroconversion was observed. When $\delta_i = 0$, that is, seroconversion was not observed for participant i , the conditional posterior density given in Equation (6) is proportional to that of a Poisson random variable with mean $\rho(\rho\eta/(\rho\eta + \exp(x_i\beta)\Lambda(t_i)))^\eta$. We draw from this Poisson distribution to obtain N_i . When $\delta_i = 1$, that is, seroconversion was observed, numerical methods are necessary to sample N_i . We propose a Metropolis-Hastings algorithm to sample N_i from Equation (6), with the constraint that N_i must be positive. (Since seroconversion occurred, the participant must have had at least one exposure process.)

After sampling the number of exposure processes for an individual, we sample frailty. If the sampled number of exposure processes for participant i , N_i , is nonzero, the conditional posterior for frailty, Z_i , is proportional to a gamma distribution with shape = $N_i\eta + \delta_i$ and rate = $\rho\eta + \exp(x_i\beta)\Lambda(t_i)$. Hence, to obtain Z_i conditional on N_i , observed data, and all parameters, we simply draw from this gamma distribution. If the sampled number of exposure processes is zero, then frailty is necessarily zero.

The proposed data augmentation MCMC algorithm is presented in greater detail in the Appendix.

2.7. Model Assessment

Predictive accuracy of Bayesian hierarchical models is commonly assessed with the deviance information criterion (DIC) [43]. However, DIC does not provide correct comparison for high dimension latent variable models [44], that is, models with one or more

latent parameters per observation, like the compound Poisson frailty model proposed here. As explored in Celeux *et al.* [45] and highlighted in a recent review paper by Spiegelhalter *et al.* [46], DIC calculations using the likelihood conditional on posterior samples of latent variables does not lead to correct model comparison. Integrating over latent variables when calculating DIC may improve performance [47], but to do so is often difficult and computationally expensive, especially so in the case of compound Poisson-distributed frailty as the density is not available in closed form. In our simulations, we found that DIC provided incorrect comparison between the frailty model (which was used to generate the data) and the Cox model. Moreover, we also obtained implausible values of the effective number of parameters, a concern also discussed in Gustafson [47]. The Watanabe-Akaike information criteria [48] and conditional predictive ordinant [49], common alternatives to the DIC, rely on the same conditional likelihood and also resulted in incorrect comparisons in simulation studies. Finally, while criteria based on the posterior predictive distribution, such as predictive concordance [50] and the L -measure [39], offer alternative approaches to model comparison, these criteria can not be calculated for survival data with a cured fraction. As a consequence, there is no common measure of predictive accuracy appropriate for our context.

In this paper, we quantify the discrepancy between observed and predicted survival using Hellinger's distance [51]. The Hellinger distance between two discrete distribution functions $P = (p_1, \dots, p_L)$ and $Q = (q_1, \dots, q_L)$ is defined as:

$$HD(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{l=1}^L (\sqrt{p_l} - \sqrt{q_l})^2}. \quad (7)$$

For both model simulations (Section 3) and application (Section 4), we compare posterior predicted survival to observed survival in discrete intervals (e.g., months) via Equation (7). While this metric can be used to identify, among a group of candidate models, which most closely predicts observed survival, it does not take into account over-fitting, as done by DIC and WAIC.

3. Simulations

3.1. Methods

We performed a series of simulations to assess the properties of the proposed model. For each simulation, data were generated from the proposed model for a sample size of $n = 3,000$, divided evenly between intervention and control groups. Compound Poisson frailties were generated using a shared value of ρ , the mean parameter for the number of exposure processes. Separate studies were done for a range of values for ρ , which corresponded to varying proportions at risk of an event, $1 - \exp(-\rho) = 0.15, 0.25, \dots, 0.55$. The shape parameter for the frailty distribution, η , was fixed at 2 for all simulation studies. (We chose to focus on the effects of varying ρ in these simulations instead of η because ρ has a more direct and interesting interpretation.) Event times were generated based on the simulated compound Poisson-distributed frailties, a constant baseline hazard function of 5 events per

100 person-years (PY), $\lambda(t) = 0.05$, and an intervention effectiveness of 50%, $\exp(\beta) = 0.5$. (Values of η and λ were selected in order to observe events at a pace comparable to ARV-based HIV prevention trials.) For each simulated data set, censoring was done after a pre-specified number of events: 90, 180, 270, and 360, that is, event rates of 3%, 6%, 9%, and 12%. Event rates as low as 3% are uncommon and typically only seen when a trial is stopped early [3]. Event rates as high as 12% are also uncommon but possible [2]. An event rate of about 6% was observed in the trial examined in the application (Section 4).

In the initial set of simulations, the prior mean was set equal to the true value for the baseline hazard and frailty model parameters, yet prior distributions were broad enough to allow for a wide range of values for posterior samples. The prior on β was non-informative. We set the following priors on model parameters.

$$\text{Prior: } \begin{cases} \beta \sim N(0, 100) \\ \lambda \sim \text{Gamma}(0.05 \times 50, 50) \\ P(\text{Risk}) = 1 - \exp(-\rho) \sim \text{Beta}(2 \times \frac{P(\text{Risk})}{1-P(\text{Risk})}, 2) \\ \eta \sim \text{Gamma}(1, \frac{1}{2}) \end{cases} \quad (8)$$

At each replication, posterior medians for model parameters and surrounding 95% highest posterior density (HPD) intervals were identified for the proposed compound Poisson frailty model. Calculation of HPD intervals was performed with the `HPDinterval` function in the `coda` package available for R. Cox model estimates of intervention effectiveness and model-based 95% confidence intervals were estimated using the `coxph` function in the `survival` package available in R. Simulations for each scenario were replicated 250 times. Posterior samples were the results of 25,000 replications of the proposed sampling algorithm, of which 10,000 were discarded for burn-in. Convergence was assessed for a few replications of each simulation by examining trace plots and cumulative quantile plots.

Model performance under various prior specifications was also examined. We repeated the simulation study for 25% of participants at risk and an event rate of 6% under three prior specification scenarios. (This event rate and proportion at risk were chosen because they were close to the event rate and screening seroprevalence, respectively, for the application in Section 4.) The scenarios examined were as follows:

- I. Incorrect prior on proportion at risk:** The prior distribution for the proportion at risk had a mean of 50%, $\pi(1 - \exp(-\rho)) = \text{Beta}(2, 2)$, twice the true value of 25%.
- II. Incorrect prior on baseline hazard:** The prior distribution on the constant baseline hazard was centered at 10 events per 100PY, $\pi(\lambda) = \text{Gamma}(0.1 \times 50, 50)$, twice the true value of 5 events per 100PY.
- III. Non-informative priors:** Uniform priors were placed on all parameters, including a flat prior on the proportion at risk, $\pi(1 - \exp(-\rho)) = \text{Beta}(1, 1)$. An upper bound of 20 was placed on the prior for η : $\pi(\eta) = \text{Uniform}(0, 20)$.

For these simulation scenarios, we identified posterior medians for all model parameters as well as coverage of 95% HPD intervals. We also calculated the Hellinger's distance (HD) between observed and predicted monthly survival. For each simulation replication, event times generated at each iteration of the posterior sampling algorithm were used to estimate monthly survival and compared to the observed survival for that replication. 250 replications were performed for each simulation study. Simulation code is available on github (<http://github.com/rycoley/cp-frailty>).

3.2. Results

Figure 1 shows hazard ratio (HR) estimates (upper panel) and coverage of intervals around the estimated HR (lower panel) from the proposed compound Poisson (CP) frailty model using priors centered around true parameter values (solid lines) and the Cox model (dashed lines) at a range of event rates and proportions at risk. We see that the proposed estimation procedure produced unbiased estimates and 95% HPD intervals with nominal coverage for β . As expected, effectiveness estimates from the Cox model underestimated individual-level effectiveness and had less than nominal coverage. Estimate attenuation and inadequate coverage were more pronounced as the proportion of the sample at risk declined. When the proportion at risk was higher and the number of events lower, however, estimates from the Cox and compound Poisson frailty models were similar.

Table 1 gives the average posterior median and coverage of the 95% HPD intervals for the compound Poisson frailty model when data were simulated with a 6% event rate and varying proportions at risk ($P(Risk)$). We see that estimates for the baseline hazard, in addition to those for the HR , were unbiased. 95% HPD intervals for model parameters were generally conservative, with the possible exception of β . We also see that the proposed model slightly overestimates the proportion at risk, $1 - \exp(-p)$ while underestimating η . It is likely that η was underestimated because the prior for that parameter, while centered at the true value, was quite skewed in order to allow for more variability in the posterior. Furthermore, as discussed below, the posterior distribution for η appears to follow the prior closely. Since η was underestimated, it is to be expected that the proportion at risk would be overestimated. Plots showing the prior and posterior distributions for model parameters from a single replication of the simulation with 25% at risk and 6% event rate are given in the Appendix.

Simulation results under non-informative and misspecified priors are given in Table 2. We see that hazard ratio estimation is quite robust to prior misspecification. Effectiveness estimates are unbiased, and coverage of 95% credible intervals is nominal. Furthermore, the difference between observed and predicted survival (as measured by HD) was constant across studies, indicating that posterior predictive accuracy is also robust to prior misspecification. Alternatively, similar HD s suggest this metric may not be particularly useful in identifying optimal prior specification.

Posterior estimates of other model parameters are less robust to the choice of priors. Bias in frailty and hazard parameters is easily explained by the nature of prior misspecification. For example, when the prior distribution on the proportion at risk was twice the true value in scenario (I), the proportion at risk was overestimated. (Results from each scenario are discussed in more detail in the Appendix.) Although asymptotic identifiability of this class

of frailty models is well established [33, 34, 52], our simulation-based exploration of estimate properties suggest that frailty distribution parameters may be non-identifiable in finite samples in the absence of proper prior specification. While identifiability is guaranteed in a Bayesian framework with placement of proper priors, we see that posterior samples of frailty model parameters may be particularly sensitive to prior specification.

4. Application

4.1. Data and Analysis

We applied the model to data from HPTN 035, a completed Phase II/IIb trial to evaluate the safety and effectiveness of BufferGel and 0.5% PRO 2000 microbicide gels in preventing male-to-female transmission of HIV [6]. Beginning in February 2005, 3,099 women were enrolled in the study from seven sites in Sub-Saharan Africa and one site in Philadelphia, USA. Participants were randomized to one of four arms: BufferGel, 0.5 % PRO 2000, placebo gel, and no gel and tested quarterly for seroconversion for twelve to thirty months. The primary analysis used the Cox proportional hazards model to estimate the hazard ratio of seroconversion between each candidate gel and both the placebo gel and no gel. BufferGel did not reduce the risk of seroconversion, nor did placebo gel in comparison to no gel. The 0.5% PRO 2000 gel showed a 27% reduction in HIV acquisition compared to the placebo gel, but this result was not statistically significant at the 0.05 level.

We estimated the individual-level effectiveness of 0.5% PRO 2000 gel in reducing HIV seroconversion risk relative to placebo gel with a site-stratified compound Poisson frailty model, i.e., the baseline hazard and frailty distribution parameters were estimated separately for each study site. Our analysis focused on 0.5% PRO 2000 gel because it showed some effect in the primary analysis. We limited analysis to data from the seven African sites because too few events were observed in Philadelphia for a stratified analysis. Following the primary analysis from the trial, we excluded participants who were infected with HIV at baseline but did not seroconvert until after enrollment. For participants without seroconversion observed, we used the last negative test as the censored event time. For participants with seroconversion observed, we sampled the event time at each iteration of the proposed sampling algorithm using slice sampling techniques and a uniform prior over the interval covering the time between a participant's last negative and first positive serology test. A 95% highest posterior density credible interval around the estimate of effectiveness was also identified.

The estimated effectiveness for an exposed individual was compared to the population-averaged estimate of effectiveness obtained by the site-stratified Cox model. In this analysis, we used the midpoint between last negative and first positive test for event time of seroconverters. (By comparison, the primary analysis used the time of first positive test as the event time in the primary analysis of this data.) A model-based 95% confidence interval was also calculated.

For the stratified compound Poisson frailty model, we used screening data to set the prior on the mean number of exposure processes for each site, p_j for all sites $j = 1, \dots, 7$, as described in Section 2.5. We set the expectation of a beta prior for the proportion at risk in each site

equal to that site's screening prevalence using the following rationale: The screening prevalence in the eligible population may provide a reasonable estimate of the prevalence of HIV among study participants' current and potential partners. Certainly, participants who were seronegative at screening likely differ in their future risk from those who were positive, so screening prevalence may overestimate the risk of exposure during the trial. Alternatively, screening seroprevalence may underestimate the proportion at risk since many of those exposed to HIV never become infected. Ultimately, we chose to set the prior expectation of the proportion at risk within a site equal to the seroprevalence observed at screening but specify a broad distribution that allows for considerable variation from this mean. HIV prevalence among those screened at each site is given in Table 3.

Specification options for the baseline hazard were limited as three of the sites had five or fewer seroconversions observed- too few events to support stratified estimation of a flexible baseline hazard (e.g., piecewise constant). As an alternative, we specified a constant baseline hazard, λ_j for each site j , and allowed that hazard to vary by a proportion λ_{6mo} within the first six months of the study to account for the expected lag time between infection and seroconversion:

$$\lambda(t) = \begin{cases} \lambda_{6mo}\lambda_j & \text{for site } j \text{ and years } t \leq 0.5 \\ \lambda_j & \text{for site } j \text{ and years } t > 0.5. \end{cases}$$

Examination of the data confirmed these assumptions were reasonable.

We also developed a hierarchical prior for site-specific baseline hazards in order to share information across sites and reflect the expected correlation between the baseline hazard and proportion at risk, which we now explain. As defined in Section 2, the baseline hazard among at-risk participants in site j is a function of the average baseline hazard, λ_j , and the proportion at risk, $1 - \exp(-\rho_j)$, and is equal to $\lambda_j/(1 - \exp(-\rho_j))$. In our stratified analysis, we assumed that the probability of exposure to HIV varied across sites due to varying prevalences but that the risk of seroconversion given exposure was roughly similar. To reflect this assumption, we set the expectation of the baseline hazard among at-risk individuals to a single, constant hazard, θ , shared across sites: $\theta = \lambda_j/(1 - \exp(-\rho_j))$ for all sites j . It follows that the prior distribution for the average baseline hazard for all participants in site j , λ_j , had expectation $\theta(1 - \exp(-\rho_j))$. Conjugate gamma priors were specified for λ_j accordingly: $\pi(\lambda_j|\rho_j, \theta, \tau) = \text{Gamma}(\theta(1 - \exp(-\rho_j)) \times \tau, \tau)$, where τ is a rate parameter. A gamma prior was also set on θ , and an inverse gamma prior was set on τ .

The relationship between the proportion at risk and baseline hazard imposed by this prior is illustrated in Figure 2, which shows the prior distribution of average baseline incidence, $\lambda_j \times 100$ PYs, for four hypothetical sites. Each site has a different proportion of its population at risk, but the average incidence among those at risk is the same ($\theta = 0.4$). We see that the prior distribution for the average baseline hazard among all participants in a site is higher when the proportion of the site's population at risk is larger. The observed relationship makes sense intuitively; when more of the population is exposed but the probability of

seroconversion given exposure remains the same, we expect to observe a greater seroconversion.

Finally, the prior for our parameter of primary interest, β , was non-informative. We set the following priors on model parameters:

$$\text{Prior: } \begin{cases} \beta \sim N(0, 100) \\ 1 - \exp(-\rho_j) \sim \text{Beta}(2 \times \frac{Prev_j}{1 - Prev_j}, 2) & \text{for each site } j \\ \eta_j \sim \text{Gamma}(0.5, 2) & \text{for each site } j \\ \lambda_{6mo} \sim \text{Gamma}(3, 4) \\ \lambda_j | \rho_j, \theta, \tau \sim \text{Gamma}(\theta(1 - \exp(-\rho_j)) \times \tau, \tau) & \text{for each site } j \\ \theta \sim \text{Gamma}(10, 28) \\ \tau \sim \text{InverseGamma}(2, 20) \end{cases}$$

where $Prev_j$ is the prevalence for site j , as given in Table 3.

Prior distributions were specified in such a way that the expected distribution of incidence in the placebo arm would be consistent with that previously observed in similar HIV prevention trials. Figure 3 shows the prior distribution of seroconversion in the placebo arm for the stratified compound Poisson frailty model, produced by 2,500 replications of a simulation generating event times under the model for 1,000 participants in each of seven study sites. Rates of seroconversion observed in the placebo arms of similar previously published trials are indicated by plotting symbols. (We note that the incidence observed in the CAPRISA 004 trial [2] is nearly twice as high as that observed in the other studies. This outlying observation can partly be explained by differences in enrollment criteria.) Overall, we see that the distribution of expected seroconversion in the placebo arm under the prior specified is consistent with those previously observed. Also, while the distribution of seroconversion under the prior appears quite informative, we note that prior distributions for individual parameters were specified to allow for considerable variability. As part of model assessment, we compared prior and posterior distributions for all model parameters. We also performed a sensitivity analysis, allowing priors for frailty distribution and baseline hazard parameters to vary from those specified above.

To obtain posterior samples from the stratified compound Poisson frailty models, we repeated the posterior sampling algorithm described in Section 2.6 for 300,000 iterations. The first 100,000 samples were discarded for burn-in and the remaining samples from parameters' posterior distributions were thinned, keeping every tenth sample. Parallel chains were run from different starting values to confirm that models converged to the same estimates. Sampling was performed in R version 3.0.3 with independently written code available on github (<http://github.com/rycoley/cp-frailty>). Highest posterior density (HPD) intervals were obtained via the `HPDInterval` command in the `coda` package. Other model diagnostics performed included examining cumulative distribution and trace plots. Finally, we generated event times from the posterior to evaluate how well the proposed model captured the underlying data generating mechanism. Kaplan Meier curves were generated from posterior predicted survival and visually compared to the Kaplan Meier curve for the

observed data. HD between posterior predicted and observed survival was also calculated at each saved iteration of the posterior sampler. A 95% bootstrapped interval for the HD was obtained from 100 site- and event-stratified bootstrap samples.

4.2. Results

Table 4 reports estimates of the population- and individual-level effectiveness of 0.5% PRO 2000 gel in reducing risk of seroconversion from the Cox model and the proposed compound Poisson (CP) frailty model, respectively. Under the site-stratified compound Poisson frailty model, 0.5% PRO 2000 gel was estimated to be 37% effective in reducing the risk of seroconversion for an exposed individual (95% HPDI: $-0.049, 0.63$). In comparison, the stratified Cox model estimate of population-level effectiveness was 27% (95% CI: $-0.12, 0.52$).

Results of the sensitivity analysis indicate that the estimate of effectiveness in this analysis was robust to prior specification. Figure 4 shows results from a sensitivity analysis varying site-specific priors on ρ and η . The left-hand column shows four site-specific prior densities for the proportion at risk, $1 - \exp(-\rho)$: the first three involve varying the Beta distribution parameter, B , used to scale the observed baseline odds of HIV infection (calculated from screening seroprevalence) while the fourth assumes the same prior distribution on the proportion at risk for all sites. The middle column shows the prior seroincidence corresponding to the prior for ρ in that row and nine different priors on η —indicated by color and line type, as defined in the legend at the bottom of the figure. The solid, black line shows expected seroincidence under the priors used in the primary analysis, as given in (9). The right-hand column shows the posterior distribution of the HR for each model. We see here that the posterior HR distribution is robust to prior specification; posterior distributions under all prior specifications are similar to the posterior for the primary analysis (black, solid line), and the estimated effectiveness for all scenarios aligns with that reported in Table 4 (shown with the vertical dashed line). Results for all combinations of nine different priors on both θ and τ were similar and are given in the Appendix.

Figure 5 shows site-specific estimates from the stratified compound Poisson frailty model of the baseline seroincidence, $\lambda_j \times 100PY$, and proportion at risk, $1 - \exp(-\rho_j)$, plotted against the seroincidence observed in the placebo arm. Unsurprisingly, estimates of the proportion at risk and the baseline incidence for each site were strongly correlated with the observed seroincidence. Moreover, the HPD intervals around the baseline hazard estimates for Durban and Hlabisa actually exclude the interval for Lilongwe. The variability in site-specific estimates of ρ_j and λ_j suggest the stratified model extends the frailty model's ability to account for heterogeneity in risk.

Kaplan Meier curves in Figure 6 show the observed (bold line) and posterior predicted (gray lines) survival for the stratified compound Poisson frailty model. The midpoint between last negative and first positive test was used as the event time for the observed data. 95% confidence intervals are also given for the observed survival (dashed lines). We see here that the proposed model appears to provide an appropriate model for the data. The HD , given in Table 4, measures the discrepancy between posterior predicted survival and that which was observed while the bootstrapped 95% quantifies the variability.

Plots comparing prior and posterior distributions for all stratified compound Poisson frailty model parameters are given in the Appendix.

5. Discussion

Inconsistent results in recent HIV prevention trials of pre-exposure prophylactic interventions may be due to heterogeneity in risk among study participants [12]. When unobserved heterogeneity is present, population-level effectiveness estimates from the Cox model underestimate the reduction in risk for an individual receiving the intervention [10]. Although estimating a population-specific average causal effect is necessary to evaluate an intervention's likely impact in a population of interest, the individual-level effect is a more relevant target parameter for identifying effective interventions.

We have presented a Bayesian hierarchical model that adjusts for heterogeneity in risk when estimating the individual-level effectiveness of candidate interventions for HIV prevention. Using a frailty term to model heterogeneity does not rely on self-reported measures of risk nor on *post hoc* identification of risk-related covariates [?]. Furthermore, unlike distributions more commonly chosen for modeling frailty, the compound Poisson distribution reflects the mechanisms of HIV risk and allows some participants to have no exposure to HIV and, therefore, no risk of seroconversion. The hierarchical definition of the compound Poisson distribution presented here is intuitive, and its parameters are interpretable. As such, the proposed Bayesian model is preferable to frequentist analogs. Frequentist formulations of the compound Poisson frailty model have relied on the Laplace transformation of the compound Poisson density, the parameters of which have no meaningful interpretations. Furthermore, a computational strength of our model is that it exploits hierarchical modeling and Bayesian estimation methods to avoid approximating an infinite series. The density of the continuous part of a compound Poisson random variable is not available in closed form, which burdens frequentist model estimation. We propose a novel data augmentation algorithm to sample the frailty term conditional on the underlying Poisson random variable, and this distribution is available in closed form.

While our model assumes a shared frailty distribution, we suspect that heterogeneity in risk, even within site, may be better modeled by a mixture of compound Poisson distributions. Participants likely have different individual probabilities of exposure and different levels of risk given exposure. If so, a single, site-specific estimate of the frailty distribution parameters may be an inadequate model for the risk of all participants. The compound Poisson frailty model has previously been extended in applications with familial dependence to allow parameter values of a frailty distribution to vary across clusters of observations [19, 53]. While HIV prevention trials contain no inherent grouping of participants, we could instead allow the frailty distribution parameters to vary depending on individual covariate data, including demographic information, STI status, reported sexual behaviors, and other risk factors. Rather than using information within pre-defined groups to estimate frailty distribution parameters, this approach could share information across all participants via regression models relating parameters to risk factors. Furthermore, this approach would allow an individual's hazard to vary stochastically conditional on risk-related covariates,

instead of directly modeling the effect of covariates on an individual's hazard through a proportional hazards function, as in the Cox regression model.

The approach proposed also assumes that a participant's frailty is constant over time. In the case of HPTN 035, this assumption may be reasonable, since the majority of participants remained in the same, sole partnership throughout the study. In general, allowing participants' frailties to change over time in order to reflect a change in partnerships or behavior would likely improve heterogeneity modeling. To do so would present challenges, however, as current time-varying frailty methods are inappropriate for univariate survival data.

Despite these limitations, we have demonstrated that the inclusion of a compound Poisson frailty term in a survival model is able to capture some of the heterogeneity in HIV risk. When applied to data from HPTN 035, the site-stratified compound Poisson frailty model estimated that 0.5% PRO 2000 gel is 37% effective in reducing risk of seroconversion for exposed individuals in comparison to placebo gel, suggesting a greater potential impact of this intervention on risk of seroconversion than the Cox model's population-averaged estimate of 27% effectiveness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contract/grant sponsor: This work was supported by NIAID grant 5 UM1 AI068615.

The authors thank the associate editor, two referees, Jared Baeten, Jim Hughes, Lurdes Inoue, and Megan Othus for their helpful comments.

References

1. Joint United Nations Programme on HIV/AIDS. Global report: UNAIDS report on the global AIDS epidemic: 2012. 2013 [accessed January 1, 2016] online, Available at http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/20121120_UNAIDS_Global_Report_2012_with_annexes_en.pdf.
2. Karim QA, Karim SSA, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, Kharsany AB, Sibeko S, Mlisana KP, Omar Z, et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science*. 2010; 329(5996):1168–1174. [PubMed: 20643915]
3. Baeten JM, Donnell D, Ndase P, Mugo NR, Campbell JD, Wangisi J, Tappero JW, Bukusi EA, Cohen CR, Katabira E, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *New England Journal of Medicine*. 2012; 367(5):399–410. [PubMed: 22784037]
4. Thigpen MC, Kebaabetswe PM, Paxton LA, Smith DK, Rose CE, Segolodi TM, Henderson FL, Pathak SR, Soud FA, Chillag KL, et al. Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *New England Journal of Medicine*. 2012; 367(5):423–434. [PubMed: 22784038]
5. McCormack S, Ramjee G, Kamali A, Rees H, Crook AM, Gafos M, Jentsch U, Pool R, Chisembele M, Kapiga S, et al. PRO2000 vaginal gel for prevention of HIV-1 infection (Microbicides Development Programme 301): a phase 3, randomised, double-blind, parallel-group trial. *Lancet*. 2010; 376(9749):1329–1337. [PubMed: 20851460]

6. Karim SSA, Richardson BA, Ramjee G, Hoffman IF, Chirenje ZM, Taha T, Kapina M, Maslankowski L, Coletti MA, Profy A, et al. Safety and effectiveness of BufferGel and 0.5% PRO2000 gel for the prevention of HIV infection in women. *AIDS*. 2011; 25(7):957–966. [PubMed: 21330907]
7. Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S, Malahleha M, Owino F, Manongi R, Onyango J, et al. Preexposure prophylaxis for HIV infection among African women. *New England Journal of Medicine*. 2012; 367(5):411–422. [PubMed: 22784040]
8. Marrazzo JM, Ramjee G, Richardson BA, Gomez K, Mgodini N, Nair G, Palanee T, Nakabiito C, van der Straten A, Noguchi L, et al. Tenofvir-based preexposure prophylaxis for HIV infection among African women. *New England Journal of Medicine*. 2015; 372(6):509–518. [PubMed: 25651245]
9. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972; 34(2):187–220.
10. Aalen OO. Heterogeneity in survival analysis. *Statistics in Medicine*. 1988; 7(11):1121–1137. [PubMed: 3201038]
11. Auvert B, Sitta R, Zarca K, Mahiane SG, Pretorius C, Lissouba P. The effect of heterogeneity on HIV prevention trials. *Clinical Trials*. 2011; 8(2):144–154. [PubMed: 21478327]
12. O'Hagan JJ, Hernán MA, Walensky RP, Lipsitch M. Apparent declining efficacy in randomized trials: Examples of the RV144 HIV vaccine and CAPRISA 004 microbicide trials. *AIDS*. 2012; 26(2):123–126. [PubMed: 22045345]
13. Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010; 21(1):13–15. [PubMed: 20010207]
14. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979; 16(3):439–454. [PubMed: 510638]
15. Longini IM, Halloran ME. A frailty mixture model for estimating vaccine efficacy. *Applied Statistics*. 1996; 45(2):165–173.
16. Hougaard P. Frailty models for survival data. *Lifetime Data Analysis*. 1995; 1(3):255–273. [PubMed: 9385105]
17. Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. *The Annals of Applied Probability*. 1992; 2(4):951–972.
18. Aalen OO, Tretli S. Analyzing incidence of testis cancer by means of a frailty model. *Cancer Causes & Control*. 1999; 10(4):285–292. [PubMed: 10482487]
19. Moger TA, Aalen OO, Heimdal K, Gjessing HK. Analysis of testicular cancer data using a frailty model with familial dependence. *Statistics in Medicine*. 2004; 23(4):617–632. [PubMed: 14755393]
20. Moger TA, Aalen OO, Halvorsen TO, Storm HH, Tretli S. Frailty modelling of testicular cancer incidence using Scandinavian data. *Biostatistics*. 2004; 5(1):1–14. [PubMed: 14744824]
21. Moger TA, Aalen OO. Regression models for infant mortality data in Norwegian siblings, using a compound Poisson frailty distribution with random scale. *Biostatistics*. 2008; 9(3):577–591. [PubMed: 18304994]
22. Price DL, Manatunga AK. Modelling survival data with a cured fraction using frailty models. *Statistics in Medicine*. 2001; 20(9–10):1515–1527. [PubMed: 11343371]
23. Rodrigues J, Cancho VG, de Castro M, Balakrishnan N. A Bayesian destructive weighted Poisson cure rate model and an application to a cutaneous melanoma data. *Statistical Methods in Medical Research*. 2012; 21(6):585–597. [PubMed: 21131689]
24. Tsodikov A. Estimation of survival based on proportional hazards when cure is a possibility. *Mathematical and Computer Modelling*. 2001; 33(12):1227–1236.
25. Pennell ML, Dunson DB. Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics*. 2006; 62(4):1044–1052. [PubMed: 17156278]
26. Banerjee S, Wall MM, Carlin BP. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. *Biostatistics*. 2003; 4(1):123–142. [PubMed: 12925334]
27. Zhou H, Lawson AB, Hebert JR, Slate EH, Hill EG. A Bayesian hierarchical modeling approach for studying the factors affecting the stage at diagnosis of prostate cancer. *Statistics in Medicine*. 2008; 27(9):1468–1489. [PubMed: 17708511]

28. Chen MH, Ibrahim JG, Sinha D. Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis*. 2002; 80(1):101–126.
29. Yin G. Bayesian cure rate frailty models with application to a root canal therapy study. *Biometrics*. 2005; 61(2):552–558. [PubMed: 16011704]
30. Yin G. Bayesian transformation cure frailty models with multivariate failure time data. *Statistics in Medicine*. 2008; 27(28):5929–5940. [PubMed: 18618427]
31. Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*. 2003; 59(2):221–228. [PubMed: 12926706]
32. Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*. 2011; 30(12):1366–1380. [PubMed: 21337596]
33. Elbers C, Ridder G. True and spurious duration dependence: The identifiability of the proportional hazard model. *The Review of Economic Studies*. 1982; 49(3):403–409.
34. Hougaard, P. Frailty models. In: Klein, JP.; Van Houwelingen, HC.; Ibrahim, JG.; Scheike, TH., editors. *Handbook of Survival Analysis*. CRC Press; 2013. p. 457–473.
35. Aalen OO, Cook RJ, Røysland K. Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis*. 2015; 21(4):579–593. [PubMed: 26100005]
36. Hernán, MA.; Robins, JM. *Causal Inference*. Boca Raton: Chapman and Hall/CRC; Forthcoming
37. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*. 1954; 8(1):1. [PubMed: 13172380]
38. Day NE, Brown CC. Multistage models and primary prevention of cancer. *Journal of the National Cancer Institute*. 1980; 64(4):977–989. [PubMed: 6929006]
39. Ibrahim, JG.; Chen, MH.; Debajyoti, S. *Bayesian Survival Analysis*. New York: Springer; 2010.
40. Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*. 1987; 82(398):528–540.
41. Neal RM. Slice sampling. *Annals of Statistics*. 2003; 31(3):705–741.
42. Liu JS, Wong WH, Kong A. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*. 1994; 81(1):27–40.
43. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2002; 64(4): 583–639.
44. Millar RB. Comparison of hierarchical Bayesian models for overdispersed count data using DIC and Bayes' factors. *Biometrics*. 2009; 65(3):962–969. [PubMed: 19173704]
45. Celeux G, Forbes F, Robert CP, Titterton DM, et al. Deviance information criteria for missing data models. *Bayesian Analysis*. 2006; 1(4):651–673.
46. Spiegelhalter DJ, Best NG, Carlin BP, Linde A. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2014; 76(3):485–493.
47. Gustafson, P. Bayesian analysis of frailty models. In: Klein, JP.; Van Houwelingen, HC.; Ibrahim, JG.; Scheike, TH., editors. *Handbook of Survival Analysis*. CRC Press; 2013. p. 475–487.
48. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*. 2010; 11:3571–3594.
49. Geisser, S. *Predictive Inference*. Vol. 55. CRC Press; 1993.
50. Gelfand, A. Model determination using sampling based methods. In: Gilks, W.; Richardson, S.; Spiegelhalter, D., editors. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall; 1996. p. 145–161.
51. Yang, GL.; Le Cam, LM. *Asymptotics in Statistics: Some Basic Concepts*. Berlin: Springer; 2000.
52. Vaupel, JW.; Yashin, AI. *The deviant dynamics of death in heterogeneous populations*. Laxenburg, Austria: International Institute for Applied Systems Analysis; 1983.
53. Moger TA, Aalen OO. A distribution for multivariate frailty based on the compound Poisson distribution with random scale. *Lifetime Data Analysis*. 2005; 11(1):41–59. [PubMed: 15747589]

Figure 1. Estimated effectiveness (upper panel) and coverage of 95% intervals (lower panel) from simulation study, priors from Equation (8).

Figure 2.

Prior distribution of average baseline incidence when proportion at risk varies and average seroincidence among those at risk is equal across all sites.

Figure 3.
Density of seroincidence in placebo arm under specified priors.

Figure 4.

Results of sensitivity analysis for priors on frailty distribution parameters. Plots in left-hand column give the site-specific prior densities for the proportion at risk. Plots in the middle column give prior seroincidence for all combinations of priors (see legend at the bottom of the figure for hyperparameters for $\pi(\eta)$) in comparison to that used in the primary analysis (solid, black line). Posterior distributions for the *HR* associated with 0.5% PRO 2000 gel are given (on the log-scale) in the right-hand column for each prior. The black, solid line gives

the posterior HR from the primary analysis (priors given in (9)), and the vertical dashed line indicates the estimated effectiveness (given in Table 4).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5.

Site-specific estimates of proportion of participants at risk of seroconversion, $1 - \exp(-\rho_j)$, (left panel), and baseline seroincidence, $\lambda_j \times 100$, (right panel) versus observed seroincidence in placebo arm, represented by plotting symbols. Solid lines depict 95% HPD intervals around site-specific estimates.

Figure 6.

Kaplan Meier curves for survival data simulated with stratified compound Poisson frailty model estimates (gray lines) and observed survival data (bold, solid line) with a 95% confidence interval around the observed survival (bold, dashed lines).

Table 1

Average posterior median for compound Poisson frailty model parameters and coverage of 95% HPD intervals from 250 replications at 6% event rate and varying proportion at risk.

	HR			P(Risk)			η			λ		
	P(Risk)	Est	Cov	Est	Cov	Est	Est	Cov	Est	Est	Cov	Cov
0.15		0.49	95%	0.17	100%	1.8	100%	1.8	100%	5.1	95%	
0.25		0.51	96%	0.28	98%	1.8	100%	1.8	100%	5.1	97%	
0.35		0.50	96%	0.38	96%	1.8	100%	1.8	100%	5.0	97%	
0.45		0.50	96%	0.48	96%	1.8	100%	1.8	100%	5.0	96%	
0.55		0.50	96%	0.60	90%	1.8	100%	1.8	100%	4.9	98%	
Truth	0.5					2				5		100PY

Average posterior median for compound Poisson frailty model parameters and coverage of 95% HPD intervals under various prior specification scenarios from 250 simulation replications at 6% event rate. The mean Hellinger Distance (HD) and quantile-based 95% intervals of the posterior mean across all replications are also given.

Table 2

Scenario	HR		P(Risk)		η		λ		HD	
	Est	Cov	Est	Cov	Est	Cov	Est	Cov	Est	95% Interval
(0): Correct priors	0.49	96%	0.29	100%	1.8	100%	5.1	97%	0.047	(0.039, 0.054)
(I): Incorrect $\boldsymbol{\pi}$ ($P(Risk)$)	0.51	96%	0.50	80%	1.4	98%	4.8	93%	0.047	(0.039, 0.054)
(II): Incorrect $\boldsymbol{\pi}$ (λ)	0.47	96%	0.28	100%	1.7	100%	5.4	96%	0.047	(0.039, 0.054)
(III): Non-informative priors	0.50	94%	0.49	66%	9.5	62%	4.8	85%	0.047	(0.039, 0.055)
Truth	0.5		0.25		2		5 / 100PY			

Table 3

HIV prevalence at screening and observed seroincidence per 100PY.

Site	Prevalence (%)	Seroincidence (100PY)
Hlabisa, South Africa	28%	7.0
Durban, South Africa	22%	5.2
Kamwala, Zambia *	22%	4.1
Blantyre, Malawi	22%	3.1
Harare, Zimbabwe	21%	1.2
Chitungwiza, Zimbabwe	19%	1.6
Lilongwe, Malawi	18%	0.89

Note: (*) Screening data not available; prevalence estimated based on seroincidence.

Table 4

Estimated effectiveness and 95% interval of 0.5% PRO 2000 gel relative to placebo gel in reducing risk of seroconversion for the Cox and compound Poisson (CP) frailty model. *HD* quantifies distance between observed and predicted survival and the 95% interval is based on 100 bootstrap samples.

<u>Cox Model</u>		<u>CP Frailty Model</u>	
Est.	(Interval)	Est.	(Interval)
0.27	(−0.12, 0.52)	0.37	(−0.049, 0.63)
		0.067	(0.066, 0.098)