Published in final edited form as:

Biometrics. 2015 June ; 71(2): 386–396. doi:10.1111/biom.12277.

Incorporating Founder Virus Information in Vaccine Field Trials

Dean Follmann^{1,*} and Chiung-Yu Huang²

¹Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, 6700B Rockledge Drive MSC 7609, Bethesda, Maryland 20892, U.S.A

²Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, 550 N. Broadway, Baltimore, Maryland 21205, U.S.A

Summary

Vaccine clinical trials with active surveillance for infection often use the time to infection as the primary endpoint. A common method of analysis for such trials is to compare the times to infection between the vaccine and placebo groups using a Cox regression model. With new technology, we can sometimes additionally record the precise number of virions that cause infection rather than just the indicator that infection occurred. In this article, we develop a unified approach for vaccine trials that couples the time to infection with the number of infecting or founder viruses. We assume that the instantaneous risk of a potentially infectious exposure for individuals in the placebo and vaccine groups follows the same proportional intensity model. Following exposure, the number of founder viruses X^* is assumed to be generated from some distribution on $0, 1, \dots$, which is allowed to be different for the two groups. Exposures that result in $X^* = 0$ are unobservable. We denote the placebo and vaccine means of X^* by μ and μ_v so that $1 - \mu_v/\mu$ measures the proportion reduction in the mean number of infecting virions due to vaccination per exposure. We develop different semi-parametric methods of estimating μ_v/μ . We allow the distribution of X^* to be Poisson or unspecified, and discuss how to incorporate covariates that impact the time to exposure and/or X^* . Interestingly μ_v/μ , which is a ratio of untruncated means, can be reliably estimated using truncated data ($X^* > 0$), even if the placebo and vaccine distributions of X^* are completely unspecified. Simulations of vaccine clinical trials show that the method can reliably recover μ_v/μ in realistic settings. We apply our methods to an HIV vaccine trial conducted in injecting drug users.

Keywords

Burden of illness; Competing risks; Cox regression; Empirical process; Infectious disease; Marked process

*dfollmann@niaid.nih.gov

8. Supplementary Materials

Web Appendices, R-code for the WEE method, and data referenced in Sections 3.2, 4.2, and 6 are available with this paper at the *Biometrics* website on Wiley Online Library.

1. Introduction

Some randomized vaccine trials use the time to clinically relevant infection as the primary endpoint with Cox regression the method of analysis. This approach is appealing as the baseline hazard function is left unspecified and the familiar Cox regression parameter is easily transformed into the vaccine efficacy or proportion reduction in the risk of infection for the vaccine arm compared to the placebo arm. With new technology we can sometimes additionally record the number of founder viruses that actually infected human cells following a potentially infectious exposure (see Keele et al. (2008) for HIV and Li et al. (2012) for hepatitis C). Intuitively, fewer founder viruses in the vaccine group should provide additional information that a vaccine is efficacious. While a comparison of the vaccine and placebo distributions of founder viruses can be easily conducted, such an analysis effectively conditions on infection status, which provides separate evidence of how a vaccine works. It is unclear how to blend founder virus information with time to infection data in a principled way.

Different measures of vaccine efficacy have been proposed (see, e.g., Halloran, Longini, and Struchiner, 2010); they typically can be interpreted as the proportion reduction in some infectious event in the vaccine arm relative to placebo and focus on preventing human disease or infection. These include vaccine efficacy for individuals, communities, and for the prevention of infection of others. This article uses a different metric for analysis of vaccine trials where the number of founder viruses, X^* , is recorded along with the times of infection. Our metric is one minus the ratio of the mean of X^* on vaccine divided by the mean of X^* on placebo. This can be interpreted as the vaccine:placebo proportion reduction in an individual virion's probability of infecting a cell, given an exposure, under weak assumptions. We view this as a measure of vaccine efficacy from the perspective of a virion. Surprisingly, by incorporating information from the times of infection, we can recover the ratio of untruncated means ($E(X^* | Z = 1)/E(X^* | Z = 0)$, where Z is the vaccine indicator), even though only truncated $X^* > 0$ is observed.

We begin by discussing a simple model where the conditional distribution of X^* given Z follows a Poisson distribution with mean μ^Z . We couple this Poisson model with an unspecified intensity function $\omega(t)$ for the risk of a potentially infectious exposure, which we assume is the same in the two groups. We discuss maximum profile likelihood estimates of μ , which are obtained after profiling out the baseline hazard.

In field trials of vaccines, the amount of presented virus may vary substantially from exposure to exposure. We thus allow arbitrary distributions of X^* with mean μ (μ^Z) in the placebo (vaccine) group. We show how the product of the Cox estimate of the vaccine:placebo hazard ratio for infection multiplied by the vaccine:placebo ratio of truncated means of X^* is consistent for μ . Finally, we develop a novel weighted estimating equation estimator based on individual level stochastic processes that jump from 0 to X^* when infection occurs. With only a treatment indicator covariate, the weighted estimating equation estimator is virtually identical to the product estimator, but it allows for proper incorporation of covariates that impact the risk of exposure and covariates that impact the

distribution of X^* . Simulations of vaccine clinical trials show that can be reliably recovered. We apply our methods to an HIV vaccine trial and end with some commentary.

2. Preliminaries

Vaccine efficacy clinical trials with active surveillance for infection often use the time to infection as the primary endpoint. A common method of analysis for such trials is to compare the times to infection between the vaccine and placebo groups using a Cox regression model where the time to infection is governed by a proportional hazards model with hazard function given by

$$h(t) = h_0(t) \exp(Z\beta),$$

where Z is the vaccine indicator, and $h_0(t)$ the baseline hazard for the placebo group. Vaccine efficacy for the human is measured by the proportion reduction in the hazard of infection on vaccine compared to placebo:

$$VE_H = 1 - \exp(\beta).$$

This is an example of vaccine efficacy for susceptibility for infection as defined in Halloran et al. (2010).

A different way to think about this model is to recognize that two things are required for infection (see Rhodes, Halloran, and Longini, 1996; Halloran et al., 2010). First, a person needs to be exposed to the pathogen, for example, sexual encounter, tainted water, mosquito bite, spores in the soil, etc. and second, the exposure needs to result in the pathogens actually infecting the host. For viruses, an infecting virion or founder virus is one that initiates a branching process of virions infecting cells that does not become extinct.

Let $\omega(t)$ be the instantaneous risk of exposure or intensity function and let $P(X^* > 0 | Z)$ be the transmission probability following an exposure in group Z . We assume a blinded trial for which it is plausible to assume that $\omega(t)$ is the same in the placebo and vaccine groups. Thus, we can write the instantaneous risk of infection as the risk of exposure times the probability of infection or $\omega(t)P(X^* > 0 | Z)$. We can write this in the familiar proportional hazards form

$$\begin{aligned} h(t) &= \omega(t)P(X^* > 0 | Z=0) \exp \left(\log \left\{ \frac{P(X^* > 0 | Z=1)}{P(X^* > 0 | Z=0)} \right\} Z \right) \\ &= h_0(t) \exp(Z\beta). \end{aligned}$$

Thus, we can re-interpret VE^H from Cox regression as

$$VE_H = 1 - \frac{P(X^* > 0 | Z=1)}{P(X^* > 0 | Z=0)},$$

or the vaccine:placebo proportion reduction in the human infection rate *conditional* on exposure.

Intuitively, smaller X^* s among the vaccine infected individuals compared to the placebo infected individuals should provide additional information that the vaccine is working beyond any imbalance in infections. To blend this information in a principled way, we consider a metric for efficacy that takes the virion's perspective

$$VE_V = 1 - \frac{E(X^*|Z=1)}{E(X^*|Z=0)} = 1 - \Delta,$$

or the vaccine:placebo proportion reduction in the mean number of founder viruses conditional on exposure. The above metric makes no assumptions about the distribution of X^* , but has an additional interpretation under weak assumptions. Suppose that an exposure releases N virions, each with a small probability p of infecting a cell. Then $VE_V = 1 -$ gives the per-virion reduction in the probability of infection conditional on exposure. This interpretation obtains even if N varies from exposure to exposure. This approach was used in monkey challenge studies where the outcome from each exposure is known (Nason, 2014; Roederer et al., 2014).

Some data is given in Figure 1 for illustration. The observable data from this trial only reveals that volunteers 1, 5, and 6 are censored at 3 years, while volunteers 2, 3, and 4 are infected in the middle of followup with 2, 2, and 1 founder viruses, respectively. Exposures that result in $X^* = 0$ founder viruses are unknown. It may be helpful to recognize that this process is stochastically equivalent to one where individuals go to a casino at some rate $\alpha(t)$ and then spin a roulette wheel with pie slices, that is, “payouts” numbered 0, 1, . . . , with the size of the pie slice labeled x proportional to the probability $P(X^* = x)$. The casino has separate wheels for vaccines and placebo volunteers. Winners spin zero and can visit the casino again. Losers spin $X^* > 0$ and cannot return. Our goal is to understand how the vaccine and placebo wheels differ by using the ratio of average payouts 0, 1, 2, Covariates, which we will introduce later, can be thought of as either modulating the arrival time to the casino (i.e., exposure) and/or modifying the size of the pie slices on the roulette wheel (i.e., X^*).

The observable data structure in this article is similar to two different bodies of work in the survival setting; competing risks and marked processes (see Prentice et al., 1978; Gilbert, McKeague, and Sun, 2008a; Sun, Gilbert, and McKeague, 2009, respectively). Just as time to death by stroke, heart attack or other causes compete with each other, so does infection by 1, 2, 3, . . . founder viruses. We can also view the time to infection as having an associated marking variable X . Our approach will develop an explicit connection between either the competing risks or the time to event and marking variables.

3. Proportional Hazards-Truncated Poisson

In this section, we formulate a specific model to estimate μ . We will call this the Proportional Hazards-Truncated Poisson (PH-TP) approach. Suppose that n volunteers are randomized. Define $Y = \min(T, C)$ as the time to infection (T) or censoring (C) and let $\eta = I(T < C)$ be the infection indicator. Assume that given Z , X^* follows a Poisson distribution with mean μ^Z . As a result, the number of founder viruses after infection, denoted by X , follows a truncated Poisson distribution. That is, $X \sim X^* | (Z, X^* > 0)$. We use X instead of X^* to emphasize it is a truncated random variable. We assume C is independent of T, X^* and that Y and X are conditionally independent given Z . This assumption could be checked by comparing the distributions of X for small versus large Y .

Our observable data consists of $(Y_i, Z_i, \eta_i, X_i, \Delta_i)$, $i = 1, \dots, n$. Under our assumptions, we can write down a likelihood

$$\begin{aligned} \mathcal{L} &= \prod_{i=1}^n h(Y_i | Z_i)^{\eta_i} S(Y_i | Z_i) f(X_i | Z_i; \mu, \Delta)^{\eta_i} \\ &= \prod_{i=1}^n \left\{ h_0(t) e^{\beta Z_i} \right\}^{\eta_i} e^{-H_0(t) e^{\beta Z_i}} f(X_i | Z_i; \mu, \Delta)^{\eta_i}, \end{aligned}$$

where $S(\cdot)$ is the survivor function for T , $\beta = \log\{1 - e^{-\mu}\} / (1 - e^{-\mu})$, $H_0(t) = \int_0^t h_0(s) ds$ and

$$f(X | Z; \mu, \Delta) = \frac{e^{-\mu \Delta^Z} \mu^X \Delta^{XZ}}{X! \{1 - e^{-\mu}\}^{1-Z} \{1 - e^{-\mu \Delta}\}^Z},$$

is the probability mass function for a truncated Poisson random variable. Because $H_0(\cdot)$ is a nuisance parameter, we apply the profile likelihood method to obtain estimates of μ . For fixed μ and Δ , the likelihood ℓ is maximized as a function of the baseline cumulative hazard function $H_0(t)$ by the Breslow estimator

$$\hat{H}_0(t) = \sum_{i=1}^n \int_0^t \frac{dN_i(t)}{\sum_{j=1}^n I(Y_j \geq u) e^{\beta Z_j}},$$

where $N_i(t) = \eta_i I(Y_i \leq t)$ is 0 until volunteer i is infected and 1 afterwards. After profiling out H_0 , the profile likelihood is proportional to

$$\mathcal{L}_P(\mu, \Delta) = \prod_{i=1}^n \left\{ \frac{e^{\beta Z_i}}{\sum_{j=1}^n I(Y_j \geq Y_i) e^{\beta Z_j}} \right\}^{\eta_i} f(X_i | Z_i; \mu, \Delta)^{\eta_i},$$

or the usual Cox partial likelihood times a truncated Poisson likelihood. Note that β is a function of μ . We can maximize $\mathcal{L}_P$ over μ and Δ to obtain the maximum (profile)

likelihood estimates. Asymptotic properties of the maximum likelihood estimator follow from arguments similar to the profiling arguments of Murphy and van der Vaart (2000).

With the estimated $\hat{\mu}$ we can directly form $\widehat{VE}_V = 1 - \hat{\Delta}$. Interestingly, we can additionally obtain a new estimate of the effect of the vaccine on human infection:

$$\widehat{VE}_H = 1 - \frac{1 - \exp(-\hat{\mu}\hat{\Delta})}{1 - \exp(-\hat{\mu})}.$$

Intuitively, this should be more efficient than the estimate of $VE_H = 1 - \exp(\hat{\beta})$ from the usual Cox model if the Poisson assumption is approximately correct. Standard errors for $\hat{\mu}$, $\hat{\Delta}$ (or functions of them) can be based on the second derivative of the profile log-likelihood (coupled with the delta-method) or by use of bootstrap methods.

3.1. Product Method

While simple, the Poisson model for X^* may be unrealistic in settings where μ varies appreciably from exposure to exposure. Additionally, we may have baseline covariates that affect the time to exposure, say $\mathbf{W}^E$ (which cannot include Z), so that the intensity of exposure is given by the proportional intensity model $\omega(t) \exp(\boldsymbol{\theta}'\mathbf{W}^E)$. For any distribution of X^* , we can write the hazard for infection as

$$\begin{aligned} h(t) &= \omega(t) \exp(\boldsymbol{\theta}'\mathbf{W}^E) P(X^* > 0 | Z=0) \times \exp \left[\log \left\{ \frac{P(X^* > 0 | Z=1)}{P(X^* > 0 | Z=0)} \right\} Z \right] \\ &= h_0(t) \exp(\beta Z + \boldsymbol{\theta}'\mathbf{W}^E), \end{aligned}$$

so $\beta = \log\{P(X^* > 0 | Z=1)/P(X^* > 0 | Z=0)\}$ and $h_0(t) = \omega(t)P(X^* > 0 | Z=0)$. Next, note that the mean of untruncated X^* for group Z is proportional to the mean of truncated X ,

$$\begin{aligned} E(X^* | Z) &= \sum_{x=0}^{\infty} x P(X^*=x | Z) = \sum_{x=1}^{\infty} x P(X^*=x | Z) \\ &= \sum_{x=1}^{\infty} x P(X^*=x | Z) \frac{P(X^*>0 | Z)}{P(X^*>0 | Z)} \\ &= E(X^* | X^*>0, Z) P(X^*>0 | Z) \\ &= E(X | Z) P(X^*>0 | Z). \end{aligned}$$

Thus, $E(X | Z) = E(X^* | Z)/P(X^* > 0 | Z)$. Let $\bar{X}_Z$ be the sample mean of X in group Z . Because a sample mean converges to its expected value, we can multiply the Cox estimate by the ratio of sample means to obtain a consistent estimator

$$\begin{aligned} \hat{\Delta} &= e^{\hat{\beta}} \left(\frac{\bar{X}_1}{\bar{X}_0} \right) \rightarrow \frac{P(X^*>0 | Z=1)}{P(X^*>0 | Z=0)} \left\{ \frac{\frac{E(X^* | Z=1)}{P(X^*>0 | Z=1)}}{\frac{E(X^* | Z=0)}{P(X^*>0 | Z=0)}} \right\} \\ &= \frac{E(X^* | Z=1)}{E(X^* | Z=0)} = \Delta. \end{aligned} \quad (1)$$

We call this the product estimate. Remarkably, it estimates the vaccine:placebo ratio of the untruncated means of X^* under essentially no assumptions about the placebo and vaccine distributions of X^* ; and with no direct knowledge of how many noninfectious exposures occur in either group. Furthermore, exposure itself need not be crisply defined.

Asymptotics for the log of the product estimator are straightforward as

$$\hat{\phi} = \log(\hat{\Delta}) = \log \left\{ e^{\hat{\beta}} \left(\frac{\bar{X}_1}{\bar{X}_0} \right) \right\} = \hat{\beta} + \log(\bar{X}_1) - \log(\bar{X}_0).$$

The first term is the usual Cox estimate of the log-hazard ratio, which is asymptotically normal. The X_Z are asymptotically normal from a standard central limit argument for the sample mean, and the asymptotic normality of $\log(X_Z)$ follows from the delta-method. Note that X_0 and X_1 are independent and we assume independence of Y and X . Thus, asymptotically $\hat{\phi}$ is the sum of independent normals and thus normal.

Let I_Z and S_Z^2 be the number infected and the sample variance of the X s in group Z . We can test $\phi = \log(\Delta) = 0$ by a Wald test statistic

$$\frac{\hat{\phi}}{\sqrt{\hat{V}}},$$

where

$$\hat{V} = \widehat{\text{var}}(\hat{\beta}) + \frac{S_1^2}{I_1 \bar{X}_1^2} + \frac{S_0^2}{I_0 \bar{X}_0^2}.$$

An asymptotic confidence interval for $\text{VE}_V = 1 - \Delta$ can be formed as

$$1 - \exp(\hat{\phi} \pm 1.96 \sqrt{\hat{V}}).$$

Another tack would be to use the bootstrap to estimate V for the Wald test and to form a percentile-bootstrap confidence interval for $1 - \Delta$.

The product estimator has a simple and interesting form under an additional assumption of an exponential time to infection with parameter $\omega_Z = \omega P(X^* > 0 | Z)$ for group Z . Under this assumption, we can estimate the hazard ratio for infection by the ratio of exponential maximum likelihood estimates, $\hat{\omega}_Z = I_Z/T_Z$, where $T_Z = \sum_{i=1}^n Y_i I(Z_i = Z)$ is the total time of observation for group Z . Under the exponential assumption, the product estimate reduces to

$$\frac{I_1/T_1}{I_0/T_0} \frac{\overline{X_1}}{\overline{X_0}} = \frac{X_{1+}/T_1}{I_{0+}/T_0},$$

where X_{Z+} is the total number of infecting virions in group Z .

3.2. Weighted Estimating Equation Approach

While the product method can handle covariates that impact exposure, it does not naturally handle covariates that impact X^* , other than Z . We are thus led to consider a different modeling approach. Denote the covariates that impact X^* by $\mathbf{W}^X$. A similar distinction between covariates that impact exposure to pathogens versus susceptibility to infection was introduced by Rhodes et al. (1996). We assume that given $Z, \mathbf{W}^X, \mathbf{W}^E$, the risk of exposure depends only on $\mathbf{W}^E$ and the distribution of X^* only depends on Z and $\mathbf{W}^X$. We require that the elements of $\mathbf{W}^X$ and $\mathbf{W}^E$ be distinct and that neither includes Z . Specifically, we assume that the intensity of exposure given $Z, \mathbf{W}^X, \mathbf{W}^E$ is given by the proportional intensity model

$$\omega(t|\mathbf{W}^E, \mathbf{W}^X) = \omega(t) \exp(\boldsymbol{\theta}' \mathbf{W}^E),$$

and assume that the mean of X^* given $Z, \mathbf{W}^E, \mathbf{W}^X$ is given by the proportional mean model

$$E(X^*|Z, \mathbf{W}^X, \mathbf{W}^E) = E(X^*|Z, \mathbf{W}^X) = \mu \exp(\varphi Z + \boldsymbol{\psi}' \mathbf{W}^X),$$

where $\varphi = \log(\quad)$. Here, we define $\quad = E(X^* | Z = 1, \mathbf{W}^X) / E(X^* | Z = 0, \mathbf{W}^X)$. The hazard function for the time to infection is given by

$$\omega(t) \exp(\boldsymbol{\theta}' \mathbf{W}^E) P(X^* > 0 | Z, \mathbf{W}^X).$$

We assume that the censoring time C is independent of (X^*, T) given $Z, \mathbf{W}^X, \mathbf{W}^E$. With these assumptions, we have

$$\begin{aligned} E(X_i dN_i(t) | Z, \mathbf{W}^X, \mathbf{W}^E) &= \frac{E(X^* | Z, \mathbf{W}^X)}{P(X^* > 0 | Z, \mathbf{W}^X)} P(X^* > 0 | Z, \mathbf{W}^X) \times \omega(t | \mathbf{W}^E) E\{I(T_i \geq t, C_i \geq t) | Z, \mathbf{W}^X, \mathbf{W}^E\} \\ &= \mu \exp(\varphi Z + \boldsymbol{\psi}' \mathbf{W}^X) \omega(t) \exp(\boldsymbol{\theta}' \mathbf{W}^E) E\{I(Y_i \geq t) | Z, \mathbf{W}^X, \mathbf{W}^E\} \\ &= \lambda(t) \exp(\varphi Z + \boldsymbol{\psi}' \mathbf{W}^X + \boldsymbol{\theta}' \mathbf{W}^E) E\{I(Y_i \geq t) | Z, \mathbf{W}^X, \mathbf{W}^E\}, \end{aligned}$$

where $\lambda(t) = \mu \omega(t)$.

Denote $\boldsymbol{\alpha} = (\varphi, \boldsymbol{\psi}, \boldsymbol{\theta})$ and $\mathbf{W}_i = (Z_i, \mathbf{W}_i^X, \mathbf{W}_i^E)$. Define the empirical process

$$M_i(t) = \int_0^t X_i dN_i(u) - \int_0^t \lambda(u) \exp(\boldsymbol{\alpha}' \mathbf{W}_i) I(Y_i \geq u) du.$$

It is easy to see that

$$E \left\{ \sum_{i=1}^n \int_0^t dM_i(u) \right\} = 0 \text{ for all } t \in [0, \tau], \quad (2)$$

$$E \left\{ \sum_{i=1}^n \int_0^\tau \mathbf{W}_i dM_i(u) \right\} = 0, \quad (3)$$

where τ is a constant satisfying $\tau < \sup\{t : \text{pr}(Y_1 > t) > 0\}$. Define $\Lambda(t) = \int_0^t \lambda(s) ds$. Solving the equation $\sum_{i=1}^n \int_0^t dM_i(u) = 0$ for each t yields

$$\hat{\Lambda}(t) = \sum_{i=1}^n \int_0^t \frac{X_i dN_i(u)}{\sum_{k=1}^n \exp(\boldsymbol{\alpha}' \mathbf{W}_k) I(Y_k \geq u)}. \quad (4)$$

Replacing $\Lambda(t)$ with (4) in the estimating equations $\sum_{i=1}^n \int_0^\tau \mathbf{W}_i dM_i(u) = 0$ yields the estimating equations

$$\sum_{i=1}^n \int_0^\tau X_i \left\{ \mathbf{W}_i - \frac{\sum_{j=1}^n \mathbf{W}_j \exp(\boldsymbol{\alpha}' \mathbf{W}_j) I(Y_j \geq u)}{\sum_{k=1}^n \exp(\boldsymbol{\alpha}' \mathbf{W}_k) I(Y_k \geq u)} \right\} dN_i(u) = 0. \quad (5)$$

Asymptotics for $\hat{\boldsymbol{\alpha}}$, the solution to (5), are given in Web Appendix A (van der Vaart 2000).

Let U be a generic covariate that impacts X^* . If $\mathbf{W} = (Z, U)$ then $E(X^*) = \mu \exp(\alpha_1 Z + \alpha_2 U)$ and $\text{VE}_V = 1 - \exp(\alpha_1) = 1 - \mu$ is the overall virion vaccine efficacy. A more interesting case is where $\mathbf{W} = (Z, U, ZU)$, so that $E(X^*) = \mu \exp(\alpha_1 Z + \alpha_2 U + \alpha_3 ZU)$ and thus

$$\text{VE}_V(U) = 1 - \exp(\alpha_1 + \alpha_3 U) = 1 - \frac{E(X^* | Z=1, U)}{E(X^* | Z=0, U)}$$

is the virion vaccine efficacy evaluated at U . An important example of such a U is the actual (in the vaccine group) or counterfactual (in the placebo group) immune response to the vaccine measured shortly after vaccination (see Follmann, 2006; Gilbert, Qin, and Self, 2008b). This approach may be more efficient at identifying important immune responses compared to methods that only record infection.

Note that if a covariate U truly impacts both X^* and exposure, say, where $E(X^*) = \exp(Z\phi + U\phi_X)$, and the intensity function is $\omega(t) \exp(U\theta_E)$, our observables will be governed by $\exp\{U(\theta_E + \phi_X)\}$. Using the weighted estimating equations, we will need to specify U as

part of $\mathbf{W}^E$, say, but we will only be able to estimate $\alpha = \theta_E + \phi_X$ as the single coefficient for such a dual effect U . Thus, any conclusion that such an α estimates the effect on, for example, exposure must rely on judgment unsupported by data (η, Y, X).

Interestingly, the estimating equations (5) can be viewed as the score function derived from an “X-weighted” partial log-likelihood where a person who has X infecting virions has his likelihood contribution repeated X times:

$$\prod_{i=1}^n \left\{ \frac{\exp(\boldsymbol{\alpha}' \mathbf{W}_i)}{\sum_{j=1}^n \exp(\boldsymbol{\alpha}' \mathbf{W}_j) I(Y_j \geq Y_i)} \right\}^{\eta_i X_i}.$$

4. Simulations

4.1. No Covariates Other Than Z

In this section, we evaluate different estimators for $\beta = \log\{P(X^* > 0 | Z = 1)/P(X^* > 0 | Z = 0)\}$ and $\phi = \log(\quad)$ when there are no covariates (other than Z). We consider two estimators of β :

1. Cox: Cox regression estimate of β based on maximizing the usual partial likelihood based on infection data alone.
2. PH-TP: We maximize $\ell_P(\mu, \quad)$ obtaining $\mu, \hat{\quad}$ and form $\hat{\beta} = \log\{1 - \exp(-\hat{\mu})/1 - \exp(-\mu)\}$.

We evaluate four estimates of ϕ :

1. PH-TP: $\hat{\phi}_{PH-TP} = \log(\hat{\quad})$ uses the maximum likelihood estimator for $\quad$ obtained by maximizing $\ell_P(\mu, \quad)$.
2. Product: $\hat{\phi}_P = \hat{\beta}_{\text{Cox}} X_1 / X_0$, where $\hat{\beta}_{\text{Cox}}$ is the usual Cox regression estimate based on maximizing the partial likelihood.
3. WEE: $\hat{\phi}_M = \alpha_1$ is based on solving the weighted estimating equation (5) for α_1 when $W = Z$.
4. Omniscient: Let R_Z be the *total* number of exposures for group Z , whether they result in infection or not, let X_k^* be the untruncated number of founder viruses from the k th sequential exposure, and let Z_k^* be the associated vaccine indicator. With this omniscience we form the vaccine:placebo ratio of sample averages of X^* :

$$\hat{\Delta}_O = \frac{\sum_{k=1}^R X_k^* Z_k^* / R_1}{\sum_{k=1}^R X_k^* (1 - Z_k^*) / R_0},$$

where $R = R_1 + R_0$. The Omniscient method provides a benchmark for performance.

For each simulated trial, $n = 3000$ volunteers are randomized and repeated exposures are generated under a constant intensity ω . Following exposure, X^* is generated under a Poisson

model with $E(X^* | Z) = \mu^Z$, or a mixture of Poissons where μ follows a gamma mixing distribution (i.e., negative binomial). If $X^* > 0$ that person is infected and removed from the risk set. The trial continues until a total of 200 infections have accumulated. We set $\mu = 0.5$ so $\phi = -0.69$ while $\beta = \log\{P(X^* > 0 | Z = 1)/P(X^* > 0 | Z = 0)\}$ differs for each scenario. For each scenario, 1000 trials were simulated. Table 1 provides details of the scenarios, while Table 2 provides the simulation results.

Under the Poisson model of scenarios 1 and 2, all estimators are unbiased. For β , the PH-TP likelihood estimate has similar efficiency to the Cox model for scenario 1 but appreciably greater efficiency under scenario 2. This is not surprising as scenario 1 has most of the X s equal to 1 while scenario 2 has most X s > 1 . For ϕ , the PH-TP likelihood estimate has similar or better performance to competitors. The product and weighted estimating equation methods are competitive with the PH-TP likelihood estimate for scenario 1, but less so for scenario 2. None of this is particularly surprising as the data are generated under a model consistent with the PH-TP model.

Under the Gamma mixture model of scenarios 3–4, the PHTP likelihood model is misspecified, which leads to noticeable bias for β and modest bias for ϕ for scenario 3. The other estimators are unbiased. The product and weighted estimating equation methods are competitive or better than the PH-TP likelihood method for ϕ for scenario 3, while for scenario 4, the PH-TP likelihood appears best.

For these simulations, the PH-TP likelihood appears to be a good choice for estimation of ϕ in scenarios 1, 2, and 4, but less so for scenario 3. Use of the PH-TP likelihood estimate of β may be risky unless X is “reasonably” close to a truncated Poisson. Interestingly, the product and weighted estimating equation estimates were virtually identical for each simulated dataset. The product and weighted estimating equation methods are conceptually quite robust and are a safer bet in uncertain settings.

4.2. Regression Simulations

In this subsection, we perform limited simulations to evaluate the performance of the product and weighted estimating equation methods when there is an additional binary covariate U distributed Bernoulli(0.5), which impacts either the time to exposure or the value of X^* given an exposure. We consider three scenarios

1. Exposure. We generate data where time to exposure is exponential with parameter $\omega \exp(U\theta)$ and $X^* | Z$ is Poisson μ^Z . We set $\mu = 0.5, 0.5$ and let $\theta = 1$ or 3.
2. $E(X^*)$. For the effect of U on X^* , we generate time to exposure as exponential with parameter ω and $X^* | Z, U$ as Poisson $\mu^Z \exp(U\psi)$. We set $\mu = 0.5, 0.5$ and let $\psi = 1$ or 3.
3. Both. We generate data where time to exposure is exponential with parameter $\omega \exp(U\theta)$ and $X^* | Z$ and $X^* | Z, U$ as Poisson $\mu^Z \exp(U\psi)$. We set $\mu = 0.5, 0.5$ and let (ψ, θ) sum to 1, that is, $(2/3, 1/3)$, $(1/2, 1/2)$, $(1/3, 2/3)$ or sum to 3, that is, $(2, 1)$, $(1.5, 1.5)$, $(1, 2)$.

For each scenario, 1000 trials were simulated. Each trial randomized 3000 participants and accrued 200 infections before stopping. For each simulated trial, we considered three different estimates of $\phi = \log(\)$.

- Product 1. A product estimate based on a simple Cox model with a single covariate

$$\omega(t)\exp(Z\beta),$$

so $\widehat{\varphi}_{\text{prod}}^{(1)} = \log\{\exp(\hat{\beta})\overline{X}_1/\overline{X}_0\}$. This ignores U so its performance will evaluate the impact of unmeasured subject-level heterogeneity on the estimator.

- Product 2. A product estimate based on a Cox model with two covariates

$$\omega(t)\exp(Z\beta + U\theta),$$

so $\widehat{\varphi}_{\text{prod}}^{(2)} = \log\{\exp(\hat{\beta})\overline{X}_1/\overline{X}_0\}$. Note that this is a proper specification when U impacts exposure (scenario 1) but not when U impacts X^* (scenario 2).

- WEE. A weighted estimating equation estimator based on $E\{X_i dN_i(t)\} = \lambda_0(t)\exp(\phi Z + \psi U)$. Thus, we solve (5) with $W = (Z, U)$. This is a properly specified model whether U impacts exposure or X^* or both.

For the product 2 and weighted estimating equation methods, we also report the performance of estimators of the coefficient of U , respectively, θ and ψ . Table 3 presents the results.

We see that all methods are extremely similar in terms of estimating $\phi = \log(\)$ for all scenarios. When U impacts time to exposure (rows 1–4), we see that both the product 2 and weighted estimating equation approaches are unbiased for the coefficient of the covariate U , but the product 2 method is more efficient. It could be that with the weighted estimating equations, the weight X is random and irrelevant for the U coefficient and weighting by X provides for a slightly more variable estimate (WEE variance of 0.156 compared to product variance of 0.136). When U impacts X^* (row 5–8) the product 2 estimate of the coefficient for U is biased. As expected, the weighted estimating equation method is unbiased for the coefficient for U , when U impacts X^* . The final 12 rows provide results when U impacts both exposure and $E(X^*)$. We see that the product 2 method is biased for $\phi + \theta$ while the weighted estimating equation approach is not. The sampling variance of $\hat{\psi}$ for the WEE approach increases with θ when $\theta + \psi = 3$.

Interestingly, product method 1 based on a single covariate Z is virtually identical to the other methods, which incorporate U . This is nice as the performance of product method 1 in these scenarios can be viewed as demonstrating its robustness to unmeasured heterogeneity between volunteers. In HIV, unmeasured heterogeneity is a potential concern as some volunteers may have repeated encounters with a partner with lots of virus resulting in an X^* with a relatively large mean, while others may have encounters with a partner with virtually no virus. Limited simulations in Web Appendix B demonstrate that the virtually identical

product and weighted estimating equation methods using only Z are robust to substantial subject level heterogeneity.

To sum up, for inference about $\phi = \log(\phi)$, all three methods provide quite similar performance for the scenarios of this simulation. The weighted estimating equation approach is preferred if covariates that impact X^* are of interest. If covariates only impact exposure and they are of interest, the product method is best.

5. Simulated HIV Vaccine Trial

In this section, we evaluate the performance of the different estimators under a realistic simulated HIV vaccine clinical trial. A key design feature is the distribution of founder viruses given a risky encounter. We use two different sources of information. An article by Wawer et al. (2005) estimates the probability of infection per coital act in serodiscordant couples as being between 0.01 and 0.001. An article by Keele et al. (2008) evaluated founder viruses in 102 humans: 78 had 1 founder virus and 24 had 2–5. Analyses suggested that truncated Poisson and truncated negative binomial data were questionable for these data and thus we fit a two component mixture of Poisson models truncated at zero to these data

$$P(X=x) = \frac{\pi((\mu_1^x e^{-\mu_1})/x!) + (1-\pi)((\mu_2^x e^{-\mu_2})/x!)}{\pi(1-e^{-\mu_1}) + (1-\pi)(1-e^{-\mu_2})}. \quad (6)$$

This model is difficult to fit to the sparse data as we have a flat log-likelihood as $\mu_1 \rightarrow 0$. We thus use the profile log-likelihood $\ell\{\pi(\hat{\mu}_1), \mu_2(\hat{\mu}_1)\}$ but set $\mu_1 = 0.0039$. This results in profile maximum likelihood estimates of $\pi(0.0039) = 0.998$ and $\mu_2(0.0039) = 2.24$. These estimates are appealing as the maximized profile likelihood is quite similar to the maximized overall likelihood, and the overall infection probability is estimated at 0.006, which is in the middle of the range reported by Wawer et al. (2005). This estimated mixture model is used to generate X^* in the placebo group. For the vaccine group, we replace μ_j with μ_j for $j = 1, 2$. This estimated model is meant primarily to be a device to generate X^* s that reasonably match observed data and is not an endorsement that there are only two types of extremely different exposures in field studies.

We consider two types of trials. The first is a phase 2b type trial that will accrue 75 infections. It has about 80% power to detect a 0.50 VE_H using the Cox model for analysis. The second is a phase 3 trial that will accrue 206 infections. It has about 80% power to detect a 0.33 VE_H using the Cox model. Based on the specified VE_H and (6) we can deduce β and ϕ which are given in Table 4. Because the product and WEE estimates are virtually identical with Z the only covariate, we just report the product estimates for brevity.

Table 4 displays the results. We see that the Cox model is unbiased for β while the PH-TP likelihood based estimate of β is about 8% (15%) biased for the phase 2b (phase 3) trials. For ϕ , the PH-TP likelihood estimate shows negligible bias. The product method is unbiased. The product estimate has a modestly smaller variance than the PH-TP likelihood estimate and noticeably better power than the PH-TP method for testing $\phi = 0$.

The Cox and WEE approaches are the preferred estimators of β and ϕ , respectively. The PH-TP approach, under this extremely non-Poisson model, is biased for β and has poorer power for ϕ . The power for testing $\beta = 0$ under Cox regression is about 0.80 while the power for testing $\phi = 0$ under the product method is about 0.88. To compare the Cox and product-empirical process methods in terms of relative efficiency for testing the null hypothesis of no vaccine effect, note that the sample size formula for a generic method is proportional to τ^2/δ^2 where τ^2 is the variance of the estimator and δ the true value of the estimator. For the phase 2b trial, we can estimate this ratio as $\{0.0620/(-0.705)^2\}/\{0.0811/(-0.889)^2\} = 1.22$. Thus, if testing for a null vaccine effect is the sole purpose, the new methods are about 22% more efficient for this scenario than Cox regression.

6. Example

To illustrate our method, we use data from a randomized trial of the VAX003 HIV vaccine conducted in Thailand (Pitisuttithum et al., 2006). A total of 2546 volunteers were recruited from 17 Bangkok drug treatment clinics and randomized 1:1 between vaccine and placebo. Inclusion criteria included injection drug use in the year prior to enrollment, being HIV negative at baseline, and age between 20 and 60 years. A total of 211 infections were accrued. Using the Cox regression model with Z the single covariate, the overall estimate of the log hazard ratio was -0.00245 with an estimated standard error of 0.138. Thus, the vaccine efficacy from the perspective of a human is virtually nil $1 - \exp(-0.00245) = 0.002$. An asymptotic 95% confidence interval for VE_H is $(-0.31, 0.24)$.

Founder viruses were determined in a subset of 39 volunteers who had high baseline risk of infection and whose infection was estimated to have occurred no more than 100 days before detection (Sterrett et al., 2014). Determination of the number of founder viruses for samples beyond the 100-day window was felt to be unreliable. We can view this subset as being representative of the distribution of founder viruses of the 211 infections if

1. The distribution of X is unaffected by proximity to a visit where infection (and number of founder viruses) is measured.
2. The distribution of X is unaffected by whether volunteers have a history of high or low risk behavior for injection drug use at baseline.

Assumption 1 seems quite plausible. Assumption 2 may be less plausible, though one might expect high risk individuals to have an increased risk of exposure rather than a larger inoculum size. In any event, we will use these 39 volunteers to illustrate our calculation, understanding its potential limitations. About 30% of these infections had $X > 1$, and the sample means (standard deviations) were 1.33, 1.67, (0.62, 1.15), respectively, for the vaccine and placebo groups. Under the assumption that these 39 volunteers are representative of the 211 infected individuals, we can use the product method to calculate an estimate of the vaccine efficacy from the perspective of the virion

$$VE_V = 1 - \exp(-0.00245) \frac{1.33}{1.67} = 0.21,$$

which is directionally in favor of the vaccine group. An asymptotic 95% confidence interval using the delta method is $(-0.33, 0.52)$, which shows little evidence of benefit. If we recalculate the confidence interval pretending all 211 patients had founder virus data recorded but the means and standard deviations of X were unchanged, we obtain a confidence interval of $(-0.09, 0.42)$, which is noticeably narrower, but does not change the conclusion of a null VE_V .

7. Discussion

This manuscript has developed a new approach to incorporate founder virus information into vaccine trials. Our metric for vaccine evaluation, VE_V , allows us to blend the infection and founder virus information in a principled way. By assuming the same arbitrary intensity of exposure for both groups, VE_V can be recovered even though only truncated data are available and the times of noninfectious exposures unknown. We introduced three different estimators of VE_V , a Proportional Hazards-Truncated Poisson estimator, a product estimator, and an weighted estimating equation estimator. If the Poisson assumption is approximately correct, the PH-TP estimate can be quite efficient for VE_V and be more efficient than Cox regression for VE_H . The Poisson assumption seems implausible for HIV vaccine trials, but may be a useful approximation in other settings. The product method is quite simple, but cannot handle covariates that impact the distribution of X^* . The weighted estimating equation method is quite general and can handle covariates that impact X^* as well as exposure.

The test of $VE_V = 0$ based on the methods of this article can be more powerful than the test of VE_H using traditional methods that ignore founder virus information. Nonetheless, for definitive evaluation of a vaccine, prevention of consequential human infection is our goal and VE_H the natural metric. We feel that our methods can provide additional useful information to help understand how an effective vaccine works, but should not supplant traditional methods of analysis for licensure decisions. The use of VE_V reflects the impact of a vaccine on a virion's ability to productively infect a cell and thus complements VE_H . It may be a more efficient metric for screening phase II vaccine candidates than VE_H .

This article has implicitly assumed that X is reliably measured. However, in practice the precision of the determination of X is not well understood, though it is known to be less reliable for samples collected far from infection in HIV. Refinements of the methods of this article could be entertained such as allowing for "measurement error" in X . Classical robustness is another potential concern. For example if a couple of X s in the placebo group are quite large, they may be quite influential in the estimate of VE_V . While this may not be a problem in HIV it could be for other settings and may deserve development of more robust approaches.

There is a growing literature in the vaccine world on combining infection information with post-infection outcomes (see, e.g., Chang, Guess, and Heyse, 1994; Lachenbruch, 2001; Mehrotra, Li, and Gilbert, 2006; Shepherd, Gilbert, and Lumley, 2007; Gilbert, Shepherd, and Hudgens, 2013). Comparing groups solely in terms of the post-infection outcome can be dangerous as the vaccine and placebo infected individuals are selected by infection and thus

not a randomized comparison. Methods that combine both infection and post-infection information may be more attractive. Examples of post-infection outcomes X include the steady state viral load measured 6 months after HIV infection, or the severity of diarrheal illness following an enteric infection. Beyond infectious disease, consider a trial in patients with heart failure where T might be the time to hospitalization with X the length of the hospital stay, or T might be time to stroke and X stroke severity. While the analog to exposure is not always clear, the methods of this article can be directly applied to such settings and may be more powerful than existing methods. Effectively the methods of this article provide a kind of weighting of the two sources of information. Nonetheless the interpretation of λ as the ratio of untruncated means per “exposure” may be lost.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Martha Nason and Michael Proschan for helpful discussion, Peter Gilbert and Paul Edlefsen for help arranging use of VAX003 data, and Katharine Bar and Global Solutions for Infectious Diseases for providing the VAX003 data. We also thank the volunteers in the VAX003 clinical trial and the Bangkok Vaccine Evaluation Group.

References

- Chang MN, Guess HA, Heyse JF. Reduction in burden of illness: A new efficacy measure for prevention trials. *Statistics in Medicine*. 1994; 13:1807–1814. [PubMed: 7997714]
- Follmann D. Augmented designs to assess immune response in vaccine trials. *Biometrics*. 2006; 62:1161–1169. [PubMed: 17156291]
- Gilbert PB, McKeague IW, Sun Y. The two-sample problem for failure rates depending on a continuous mark: An application to vaccine efficacy. *Biostatistics*. 2008; 9:263–276. [PubMed: 17704528]
- Gilbert PB, Qin L, Self SG. Evaluating a surrogate endpoint at three levels, with application to vaccine development. *Statistics in Medicine*. 2008; 27:4758–4778. [PubMed: 17979212]
- Gilbert PB, Shepherd BE, Hudgens MG. Sensitivity analysis of per-protocol time-to-event treatment efficacy in randomized clinical trials. *Journal of the American Statistical Association*. 2013; 108:789–800.
- Halloran, ME.; Longini, IM.; Struchiner, CJ. *Design and Analysis of Vaccine Studies*. Springer-Verlag; New York: 2010.
- Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterisation of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:7552–7557. [PubMed: 18490657]
- Lachenbruch PA. Comparisons of two-part models with competitors. *Statistics in Medicine*. 2001; 20:1215–1234. [PubMed: 11304737]
- Li H, Stoddard MB, Wang SY, Blair LM, Giorgi EE, Parrish EH, et al. Elucidation of hepatitis C virus transmission and early diversification by single genome sequencing. *PLOS Pathogens*. 2012; 8:e1002880. [PubMed: 22927816]
- Mehrotra DV, Li X, Gilbert PB. A comparison of eight methods for the dual-endpoint evaluation of efficacy in a proof-of-concept HIV vaccine trial. *Biometrics*. 2006; 62:893–900. [PubMed: 16984333]
- Murphy SA, van der Vaart AW. On profile likelihood. *Journal of the American Statistical Association*. 2000; 95:449–465.

- Nason M. 2014 The use of founder virus information in repeated low dose challenge models. Working Manuscript
- Pitisuttithum P, Gilbert P, Gurwith M, Heyward W, Martin M, van Griensven F, et al. Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *Journal of Infectious Diseases*. 2006; 194:1661–1771. [PubMed: 17109337]
- Prentice RL, Kalbfleisch JD, Peterson AV, Fluornoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics*. 1978; 34:541–554. [PubMed: 373811]
- Rhodes PH, Halloran ME, Longini IM. 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.
- Roederer M, Keele BF, Schmidt SD, Mason RD, Welles HC, Fischer W, et al. Immunological and virological mechanisms of vaccine-mediated protection against SIV and HIV. *Nature*. 2014; 505:502–508. [PubMed: 24352234]
- Sun YQ, Gilbert PB, McKeague IW. Proportional hazards models with continuous marks. *The Annals of Statistics*. 2009; 37:394–426. [PubMed: 19838313]
- Shepherd BE, Gilbert PB, Lumley T. Sensitivity analyses comparing time-to-event outcomes existing only in a subset selected postrandomization. *Journal of the American Statistical Association*. 2007; 102:573–582. [PubMed: 19122791]
- Sterrett S, Learn G, Edlefsen P, Haynes BF, Hahn BH, Shaw GM, et al. Low multiplicity of HIV-1 infection and no vaccine enhance in VAX003 injection drug users. 2014 Working Manuscript.
- van der Vaart, AW. *Asymptotic Statistics*. Cambridge, United Kingdom: Cambridge University Press; 2000.
- Wawer MJ, Gray RH, Sewankambo NK, Serwadda D, Li X, Laeyendecker O, et al. Rates of HIV-1 transmission per coital act, by stage of HIV-1 infection, in Rakai, Uganda. *Journal of Infectious Diseases*. 2005; 191:1403–1409. [PubMed: 15809897]

Figure 1.

Illustrative data from a vaccine trial where the number of founder viruses following exposure is Poisson with parameter $\mu = -\log(0.5)$. The number of founder viruses at the time of each risky encounter are displayed; $X^* = 0$ s are unobserved, $X^* > 0$ s are observed and result in infection.

Table 1

Features of the distribution of the number of founder viruses in the placebo group. Note that $E(\mu)$ is the mean number of founder viruses following a risky encounter and $P(X^* > 0 \mid Z = 0)$ is the probability of infection given a risky encounter. The columns on the right provide a tabulation of the $X_s > 0$ from the placebo group for a randomly selected clinical trial with 200 total infections.

Tally of the X_s from the placebo group from a randomly selected trial													
Scenario	$E(\mu)$	$V(\mu)$	$P(X^* > 0 \mid Z = 0)$	1	2	3	4	5	6	7	8	9	10
Poisson 1	0.105	0	0.10	142	6								
Poisson 2	2.30	0	0.90	29	36	25	16	10	3	1			
Neg-Bi 3	0.105	.2	0.06	61	29	12	3	2	2				
Neg-Bi 4	2.30	.2	0.89	25	37	16	13	7	7	2	0	1	

Table 2

Simulation results for different estimators of β and ϕ . Each row is based on 1000 simulated trials. Infectious exposures follow an exponential distribution with X^* 's following either a Poisson or negative binomial. Each block of three rows provides, respectively, the sample mean, sample variance and sample mean divided by the square root of the sample variance.

Scenario	β			ϕ			True	
	Cox	PH-TP	Product	WEE	PH-TP	Omniscient	β	ϕ
Poisson 1	-0.666	-0.668	-0.694	-0.694	-0.694	-0.696	-0.667	-0.693
	0.0232	0.0224	0.0239	0.0239	0.0239	0.0217		
	-4.37	-4.46	-4.49	-4.49	-4.49	-4.72		
Poisson 2	-0.279	-0.281	-0.698	-0.698	-0.700	-0.692	-0.275	-0.693
	0.0193	0.0039	0.0254	0.0253	0.0160	0.0104		
	-2.01	-4.50	-4.38	-4.38	-5.53	-6.79		
Neg-Bi 3	-0.459	-0.374	-0.696	-0.696	-0.644	-0.695	-0.456	-0.693
	0.0224	0.0137	0.0311	0.0311	0.0335	0.0301		
	-3.07	-3.20	-3.95	-3.95	-3.52	-4.01		
Neg-Bi 4	-0.282	-0.274	-0.702	-0.702	-0.697	-0.698	-0.275	-0.693
	0.0189	0.0038	0.0253	0.0253	0.0166	0.0105		
	-2.05	-4.43	-4.42	-4.41	-5.41	-6.81		

Table 3

Performance of different estimators of $\phi = \log(\cdot)$ and the coefficient for U , which is distributed Bernoulli(0.5) and either impacts the time to exposure, $E(X^*)$, or both. Successive pairs of rows provide the sample mean and variance of the estimates.

Case	Impact of U	Product 1 $h(t) = h_0(t) \exp(Z\beta)$		Product 2 $h(t) = h_0(t) \exp(Z\beta + U\theta)$		WEE $E\{X\Delta N(t)\} = \lambda_0(t) \exp(Z\phi + U\psi)$		True	
		ϕ	θ	ϕ	θ	ϕ	ψ	ϕ	ψ
1	Exposure	-0.705	1.016	-0.698	1.016	-0.698	1.020	-0.693	1.0
		0.026	0.026	0.025	0.026	0.025	0.031		
1	Exposure	-0.691	3.057	-0.691	3.057	-0.696	3.069	-0.693	3.0
		0.025	0.136	0.024	0.136	0.025	0.156		
2	$E(X^*)$	-0.712	0.701	-0.703	0.701	-0.704	1.005	-0.693	0.1
		0.028	0.023	0.027	0.023	0.027	0.030		
2	$E(X^*)$	-0.706	1.180	-0.702	1.180	-0.702	3.005	-0.693	3.0
		0.030	0.026	0.029	0.026	0.029	0.032		
3	Both exposure and $E(X^*)$	-0.704	0.925	-0.696	0.925	-0.697	1.002	-0.693	0.667
		0.026	0.026	0.025	0.026	0.025	0.030		
		-0.689	0.884	-0.681	0.884	-0.682	1.007	-0.693	0.500
		0.025	0.024	0.024	0.024	0.024	0.030		
		-0.710	0.827	-0.702	0.827	-0.703	1.000	-0.693	0.333
		0.027	0.024	0.027	0.024	0.027	0.030		
3	Both exposure and $E(X^*)$	-0.695	2.744	-0.692	2.744	-0.693	3.048	-0.693	2.0
		0.026	0.090	0.025	0.090	0.026	0.108		
		-0.696	2.474	-0.694	2.474	-0.694	3.040	-0.693	1.50
		0.027	0.070	0.025	0.070	0.026	0.082		
		-0.698	2.109	-0.697	2.109	-0.696	3.030	-0.693	1.0
		0.028	0.051	0.027	0.051	0.029	0.062		

Table 4

Simulation results for a phase 2b trial with 75 infections (top block), and a phase 3 trial with 206 infections (bottom block). Each block of four rows provides, respectively, the sample mean, sample variance, sample mean divided by the square root of the sample variance, and the power for testing.

β		ϕ			True	
Cox	PH-TP	Product	PH-TP	Omniscient	β	ϕ
-0.705	-0.651	-0.889	-0.868	-0.881	-0.693	-0.879
0.0620	0.054	0.0811	0.0859	0.0162		
-2.83	-2.79	-3.12	-2.96	-6.92		
80	79	88	83	100		
-0.401	-0.345	-0.518	-0.491	-0.510	-0.405	-0.511
0.0208	0.016	0.0282	0.0305	0.0140		
-2.79	-2.72	-3.08	-2.81	-4.31		
80	78	87	81	98		