

DESIGN OF VACCINE TRIALS DURING OUTBREAKS WITH AND WITHOUT A DELAYED VACCINATION COMPARATOR¹

BY NATALIE E. DEAN^{*}, M. ELIZABETH HALLORAN^{†,‡} AND
IRA M. LONGINI^{*}

University of Florida^{}, Fred Hutchinson Cancer Research Center[†] and
University of Washington[‡]*

Conducting vaccine efficacy trials during outbreaks of emerging pathogens poses particular challenges. The “Ebola ça suffit” trial in Guinea used a novel ring vaccination cluster randomized design to target populations at highest risk of infection. Another key feature of the trial was the use of a delayed vaccination arm as a comparator, in which clusters were randomized to immediate vaccination or vaccination 21 days later. This approach, chosen to improve ethical acceptability of the trial, complicates the statistical analysis as participants in the comparison arm are eventually protected by vaccine. Furthermore, for infectious diseases, we observe time of illness onset and not time of infection, and we may not know the time required for the vaccinee to develop a protective immune response. As a result, including events observed shortly after vaccination may bias the per protocol estimate of vaccine efficacy. We provide a framework for approximating the bias and power of any given analysis period as functions of the background infection hazard rate, disease incubation period, and vaccine immune response. We use this framework to provide recommendations for designing standard vaccine efficacy trials and trials with a delayed vaccination comparator. Briefly, narrower analysis periods within the correct window can minimize or eliminate bias but may suffer from reduced power. Designs should be reasonably robust to misspecification of the incubation period and time to develop a vaccine immune response.

1. Background. Evaluating the efficacy of a vaccine candidate during a public health emergency brings unique challenges. Trials in resource-limited settings may face severe logistical constraints; transmission can be highly localized and hard to predict; furthermore, the ethics of a trial in the face of an emergency are complex. Innovative designs can address some of these challenges. During the 2013–2016 West African Ebola epidemic, the “Ebola ça suffit” ring vaccination trial in Guinea randomized clusters of contacts and contacts of contacts of confirmed Ebola virus disease cases to immediate or delayed (after 21 days) vaccination [Ebola ça Suffit Ring Vaccination Trial Consortium (2015)]. The trial used the delayed vaccination arm as a comparator to estimate vaccine efficacy (VE).

Received September 2016; revised August 2017.

¹Supported by National Institutes of Health Grants U54-GM111274 and R37-AI032042.

Key words and phrases. Vaccine trial, infectious diseases, per protocol.

When analyzing data from a vaccine trial with or without a delayed vaccination arm, infection times are unknown and only illness (symptom) onset times are observable. The time between infection and illness onset, known as the incubation period, is an unobserved random variable. Furthermore, there is a delay between vaccination and the development of a robust immune response which we refer to as the vaccine ramp-up period. As a result, individuals with illness onsets occurring shortly after vaccination were likely infected prior to vaccination (incubating cases) and/or were infected when the immune response had not developed. Including these cases in the per protocol analysis can bias the estimated VE toward the null [Horne, Lachenbruch and Goldenthal (2001)]; we refer to the biased expectation of VE as apparent VE. In response, the analysis period for vaccine trials often starts after some delay greater than the maximum incubation period plus the maximum vaccine ramp-up period. For example, the per protocol analysis of the RTS,S malaria vaccine was restricted to cases occurring 14 or more days after the third dose [RTSS Clinical Trials Partnership (2015)]. Unfortunately, this approach can negatively impact power. In the RV144 HIV vaccine efficacy trial, the per protocol analysis (seronegative at 26-week visit and followed protocol) did not achieve statistical significance because it excluded about 25% of study participants and 31% of infections, whereas the modified intent to treat analysis (seronegative at baseline visit) was significant at the 0.05 level [Gilbert et al. (2011)]. No formal guidance exists on how to identify the optimal analysis period for vaccine trials.

A second key design issue is that the comparator arm may receive delayed vaccination for ethical reasons, as was used in the “Ebola ça suffit” trial [Henao-Restrepo et al. (2015)]. This has a few benefits, including that all trial participants receive the vaccine while they are still at risk. Participants may also be more likely to consent if they are offered delayed vaccination as compared to placebo or vaccination for a different disease. Finally, trial partners may be more likely to approve a protocol with this component. The use of delayed vaccination necessarily decreases study power as it restricts the period when the comparator arm is unprotected by vaccine, and thus VE is estimable. This reduction in power as compared to a standard parallel design is well recognized for stepped wedge trials, in which all clusters start in the control arm and then receive the intervention in a randomized order [Hussey and Hughes (2007), Bellan et al. (2015)]. Delayed vaccination is related to this approach, and it can be viewed as a one-way crossover trial. As the use of a delayed vaccination comparator arm outside of stepped wedge trials is novel, we provide guidance on how to select this vaccination delay. Furthermore, we make recommendations on how to conduct the analysis in the presence of a delayed vaccination arm, including when to start including cases in the primary analysis and, importantly, when to stop including cases.

In this paper, we describe a model for the hazard rates of illness onset times in a vaccine trial with and without delayed vaccination. We present closed-form

approximations for apparent VE and power to detect a significant vaccine effect for any given analysis period. We describe a framework for selecting the optimal analysis period start in terms of maximizing power and minimizing bias in trials with a placebo or unrelated vaccine control arm. Next, we consider trials using delayed vaccination and provide recommendations on selecting this delay and conducting the analysis. We describe the principles in the context of the Guinea Ebola vaccine trial.

2. Description of the model. We describe a model for the observable illness onset hazard rate as a function of the unobservable infection hazard rate, time-dependent vaccine protection, and disease incubation period. The background infection hazard rate, defined as the rate at which a susceptible unvaccinated person becomes infected, is a time-dependent function $\lambda_W(w)$ of infection times $W = w$. In the absence of vaccination, the actual infection hazard rate is equal to the background infection hazard rate. After vaccination, as the immune response develops, we assume that the background infection hazard rate is multiplicatively reduced due to the protective effect of the vaccine; this multiplicative reduction, referred to as leaky or partial vaccine protection, is in contrast to “all-or-none” protection in which a proportion of vaccinees are assumed to be completely protected and the rest are susceptible [Halloran et al. (1991)]. True vaccine efficacy VE_0 for a leaky vaccine is defined as one minus the infection hazard ratio for a maximally protected, vaccinated individual as compared to a susceptible, unvaccinated individual; this quantity is referred to the vaccine efficacy for susceptibility, and it measures protection against infection [Halloran, Longini and Struchiner (2010)]. VE_0 is the estimand of interest in vaccine efficacy trials. In vaccine trials, unless participants are tested for asymptomatic infection, we observe vaccine efficacy for preventing symptomatic disease rather than infection; VE_0 is then defined as one minus the illness onset hazard ratio for a maximally protected individual versus a susceptible unvaccinated individual. Vaccine efficacy for susceptibility and vaccine efficacy for preventing disease have slightly different definitions, but for simplicity we use them interchangeably in this paper; if vaccine does not alter the probability that an infection will yield symptomatic disease, the two quantities are equal.

A vaccinated individual is maximally protected after a “vaccine ramp-up” with duration R days. During the ramp-up period, the infection hazard rate in vaccinees is assumed to be linearly decreasing, though this could be readily modified. In equation (1), $\beta(w|s_j, VE_0, R)$ describes the multiplicative reduction in infection hazard rate at infection time w in Arm j vaccinated on day s_j ; it equals 1 before the individual is vaccinated, indicating no change in infection hazard rate, and equals $1 - VE_0$ after the ramp-up period R , indicating that the vaccine is maximally

protective:

$$(1) \quad \beta(w|s_j, VE_0, R) = \begin{cases} 1, & w < s_j, \\ 1 - \frac{VE_0}{R}(w - s_j), & s_j \leq w < s_j + R, \\ 1 - VE_0, & w \geq s_j + R. \end{cases}$$

In a standard vaccine trial, the intervention arm (Arm 1) is vaccinated on day $s_1 = 0$, and the comparator arm (Arm 0) is never vaccinated, $s_0 = \infty$. In a trial with a delayed vaccination arm, the immediate arm (Arm 1) is vaccinated on day $s_1 = 0$, and the delayed arm (Arm 0) is vaccinated on day $s_0 = b$ where b is the vaccination delay. Where trial participants are vaccinated on different days (e.g., stepped rollout as in ring vaccination), the time scale might be standardized so that day 0 is the date of randomization. In a ring vaccination trial, for example, the background infection hazard rate can be interpreted as the infection hazard rate in a population with a recent case that triggered the definition of a ring, assuming that the process of contact tracing and enrollment into the trial will have a similar impact on the infection hazard rate across rings.

Given infection at time $W = w$, the incubation period for an infected individual is a nonnegative random variable $U = u$ that follows some continuous distribution with probability distribution function $f_U(\cdot)$, cumulative distribution function $F_U(\cdot)$, median $u_{0.50}$, 90th percentile $u_{0.90}$, and 99.9th percentile $u_{0.999}$. Examples of distributions are $U \sim \text{Unif}(0, 10)$ or $U \sim \text{Gamma}(\text{shape} = 6, \text{scale} = 1)$. Illness onset time is thus a random variable $T = t$, where an individual develops illness at time $t = w + u$. We assume that t is observable for any individual, though u and w are not. There could be exceptions, such as a known point exposure, but we do not consider these here.

We let \mathbf{X} be the set of parameters comprised of VE_0 , R , $F_U(\cdot)$, and $\lambda_W(\cdot)$ that characterize the vaccine and the disease. Integrating over all possible combinations of infection times and incubation periods as given by \mathbf{X} , the illness onset hazard rate $\lambda_T(t|s_j, \mathbf{X})$ at a particular time t for an individual vaccinated on day s_j has the following form:

$$(2) \quad \lambda_T(t|s_j, \mathbf{X}) = \int_{u=0}^{\infty} \lambda_W(t - u)\beta(t - u|s_j, VE_0, R) f_U(u) du.$$

The illness onset hazard rate simplifies nicely in a few settings. When the background infection hazard rate is constant [$\lambda_W(w) = \lambda_W \forall w$] and there is no ramp-up period ($R = 0$), the illness onset hazard rate at time t in Arm 1 is as follows [derivation in Dean, Halloran and Longini (2018)]:

$$(3) \quad \lambda_T(t|s_1, \mathbf{X}) = \lambda_W[1 - VE_0 F_U(t - s_1)].$$

In the general setting, the apparent VE at time t , denoted $VE_T(t)$, is equal to one minus the expected illness onset hazard ratio, written below:

$$(4) \quad VE_T(t|s_1, s_0, \mathbf{X}) = 1 - \frac{\lambda_T(t|s_1, \mathbf{X})}{\lambda_T(t|s_0, \mathbf{X})}.$$

Apparent VE reflects the observable multiplicative difference in illness onset hazard rate at a given time t , noting that at many times this will not equal our estimand of interest VE_0 . For example, immediately after vaccination, the illness onset hazard rate in Arm 1 will not yet be reduced relative to Arm 0 because vaccine protection has not developed. We define bias as the difference between the apparent VE at any time t and the true vaccine efficacy VE_0 .

3. Illness onset hazard rates and apparent VE. We use the equations in Section 2 to model hypothetical vaccine trials. For all figures in this manuscript, unless otherwise noted, we assume a scenario modeled after Ebola virus disease where the disease incubation period is gamma distributed with a mean of 6 days and scale parameter of 1. The vaccine is fast-acting with a linear four-day vaccine ramp-up period ($R = 4$) before maximal VE of $VE_0 = 90\%$ is achieved. We assume a low event rate, and consider scenarios with a decreasing background infection hazard rate as disease containment procedures are implemented in the population.

The solid lines in Figure 1 depict the illness onset hazard rates in each arm, $\lambda_T(t|s_1, \mathbf{X})$ and $\lambda_T(t|s_0, \mathbf{X})$, and apparent VE, $VE_T(t)$, as determined by equations (2) and (4), respectively, and the dashed lines depict the comparable infection hazard rates. Arm 0 is never vaccinated, and the background infection hazard rate is constant and low ($\lambda_W = 0.001$, resulting in an Arm 0 attack rate $\approx 3\%$ over 30 days). In Panel A, we see that the infection hazard rate in Arm 1 drops sharply after vaccination on day $s_1 = 0$ until minimizing on day R ; the drop is then reflected in the illness onset hazard rate with a delay attributable to the incubation period. In

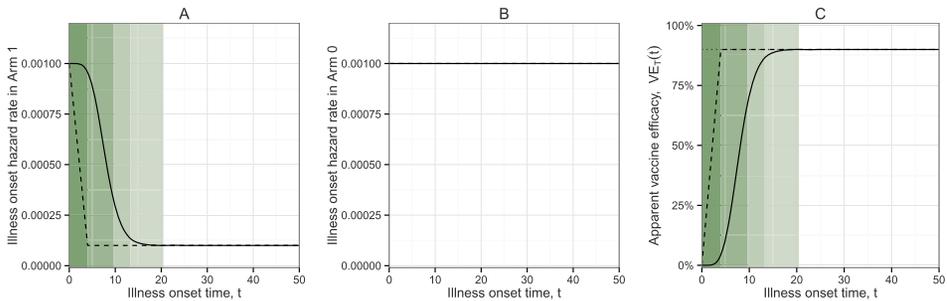


FIG. 1. Arm 1 vaccinated on day $s_1 = 0$; Arm 0 never vaccinated ($b = \infty$). Constant background infection hazard rate $\lambda_W = 0.001$. For this and all following figures, left shaded regions identify key time intervals in Arm 1, from darkest to lightest, (i) s_1 to $s_1 + R$, (ii) to $s_1 + R + u_{0.50}$, (iii) to $s_1 + R + u_{0.90}$, and (iv) to $s_1 + R + u_{0.999}$. (A) Solid line indicates the illness onset hazard rate in Arm 1 as a function of time, $\lambda_T(t|s_1, \mathbf{X})$; dashed line indicates the infection onset hazard rate in Arm 1. (B) Solid line indicates the illness onset hazard rate in Arm 0 as a function of time, $\lambda_T(t|s_0, \mathbf{X})$; dashed line indicates the infection onset hazard rate in Arm 0. (C) Apparent VE ($1 -$ illness onset hazard ratio) as a function of time; dashed line ($1 -$ infection hazard ratio); horizontal dotted line indicates $VE_0 = 90\%$.

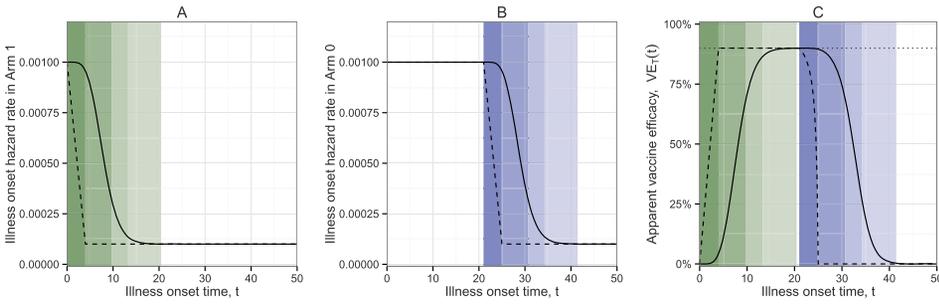


FIG. 2. Arm 1 vaccinated on day $s_1 = 0$; Arm 0 vaccinated on day $s_1 = 21$ ($b = 21$). Constant background infection hazard rate $\lambda_W = 0.001$. For this and all following figures, right shaded regions identify key time intervals in Arm 0, from darkest to lightest, (i) s_0 to $s_0 + R$, (ii) to $s_0 + R + u_{0.50}$, (iii) to $s_0 + R + u_{0.90}$, and (iv) to $s_0 + R + u_{0.999}$. (A) Solid line indicates the illness onset hazard rate in Arm 1 as a function of time, $\lambda_T(t|s_1, \mathbf{X})$; dashed line indicates the infection onset hazard rate in Arm 1. (B) Solid line indicates the illness onset hazard rate in Arm 0 as a function of time, $\lambda_T(t|s_0, \mathbf{X})$; dashed line indicates the infection onset hazard rate in Arm 1. (C) Apparent VE ($1 -$ illness onset hazard ratio) as a function of time; dashed line ($1 -$ infection hazard ratio); horizontal dotted line indicates $VE_0 = 90\%$.

Panel C, the dashed line indicating one minus the infection hazard ratio increases from 0 to $VE_0 = 90\%$ by day R , but apparent VE increases are slower to stabilize. Here, stabilization begins shortly before $R + u_{0.999}$.

Figure 2 shares the same structure as Figure 1 but the comparator arm is vaccinated after 21 days ($b = 21$). The illness onset hazard rate in Arm 0 resembles that of Arm 1 shifted by b days. Apparent VE briefly maximizes at $VE_0 = 90\%$ but returns to 0 shortly after delayed vaccination. Panel C suggests that there is a limited time window in which VE_0 is estimable without bias; at illness onset times before and after this window, apparent VE can be highly attenuated compared to VE_0 .

4. Trial analysis framework. As demonstrated in Section 3, apparent VE can very different from VE_0 at particular illness onset times, especially when delayed vaccination is used. A key design choice is then which illness onset times should contribute to the primary analysis. A commonly used approach is to apply hard cutpoints to create an analysis period in which illness onsets occurring before the starting cutpoint are excluded, and observations occurring after the ending cutpoint are censored. Though statistically more refined approaches are available, this simple approach may be preferred by regulators. We use d to denote the analysis period start, which is the earliest illness onset day, relative to the date of randomization, included in the primary analysis, and we use c to denote the length of the analysis period in days. The analysis period is identical in both trial arms to maintain comparability; a different analysis period for each arm could induce additional bias [Camacho et al. (2015)].

Cox proportional hazards or a piecewise exponential model fit with a log-linear approach can be used to estimate vaccine efficacy if the data are expressed in time-to-event format with times shifted to subtract the analysis period start d to prevent immortal time before d . These models allow for flexibility in the background infection hazard rate and can accommodate additional covariates (e.g., risk factors for infection). Clustering, if a cluster randomized trial design is used, can be modeled with a shared frailty term.

4.1. *Bias approximation.* In the Cox and piecewise exponential models, illness onset hazard rates $\lambda_T(t|s_1, \mathbf{X})$ and $\lambda_T(t|s_0, \mathbf{X})$ can be time-dependent as long as the hazards are proportional (constant hazard ratio). From Section 3, we see that apparent VE is actually time-varying (“time varying effect”). When we assume proportional hazards but proportional hazards are violated, we estimate an average regression effect [Xu and O’Quigley (2000)]. As no closed-form formula is available for the Cox model, we approximate the apparent VE for a given analysis period, $VE_D(d, c|s_1, s_0, \mathbf{X})$, from the ratio of the average illness onset hazard rates observed in $[d, d + c)$ [see equation (5)]:

$$(5) \quad VE_D(d, c|s_1, s_0, \mathbf{X}) = 1 - \frac{\frac{1}{c} \int_{t=d}^{d+c} \lambda_T(t|s_1, \mathbf{X}) dt}{\frac{1}{c} \int_{t=d}^{d+c} \lambda_T(t|s_0, \mathbf{X}) dt}.$$

4.2. *Power approximation.* The power for comparing survival curves under Cox proportional hazards can be approximated by the power of the log rank test statistic to test $H_0 : VE_0 \neq 0\%$ [Rosner (2010), Chapter 14, page 784]. To estimate power for a given analysis period, $B(d, c|s_1, s_0, \mathbf{X})$, in equation (6) we adapt the standard formula, replacing the true VE, VE_0 , with the apparent VE, $VE_D(d, c)$, and calculating the expected number of events (illness onsets) using the illness onset hazard rates:

$$(6) \quad B(d, c|s_1, s_0, \mathbf{X}) = \Phi \left[\frac{\sqrt{m(d, c|s_1, s_0, \mathbf{X})|VE_D(d, c|s_1, s_0, \mathbf{X})|}}{2 - VE_D(d, c|s_1, s_0, \mathbf{X})} - 1.96 \right],$$

where $m(d, c|s_1, s_0, \mathbf{X})$ is the expected total number of illness onsets during the analysis period, calculated as $np_1 + np_0$ where n is the sample size per arm and p_j [in equation (7)] is the probability that a participant does not have an illness onset occurring before time d but has an illness onset during the analysis period $[d, d + c)$ in arm j . We refer to the estimated power calculated using apparent VE as apparent power. Though a time-dependent background infection hazard rate could be used, in practice, one would likely assume a constant rate:

$$(7) \quad p_j = \exp \left\{ - \int_{t=0}^d \lambda_T(t|s_j, \mathbf{X}) dt \right\} \left[1 - \exp \left\{ - \int_{t=d}^{d+c} \lambda_T(t|s_j, \mathbf{X}) dt \right\} \right].$$

Equation (6) can be adapted to test $VE_0 > 30\%$ or some other pre-specified lower bound. Results are presented for individually randomized trials, but this can

be modified for cluster randomized designs by solving equation (6) for the effective sample size n and then multiplying the effective sample size by the trial design effect. The design effect is frequently approximated as $1 + (m - 1)\rho$ where ρ is the intracluster correlation coefficient and m is the per-cluster sample size [Ridout, Demétrio and Firth (1999)].

4.3. *Simulations.* We evaluate the performance of these two approximations using a simulation approach in R [R Core Team (2015)] using the survival package [Therneau (2015)]. Starting 40 days before Arm 1 is vaccinated, a random sample of $n = 1000$ infection times per arm are independently drawn from a piecewise exponential distribution, with hazard rate equal to the background infection hazard rate ($\lambda_W = 0.001$ or $\lambda_W = 0.01$) or multiplicatively reduced by VE_0 (0%, 50%, or 90%) once an individual is vaccinated ($R = 0$). Incubation periods are then drawn from a gamma distribution and added to infection times to obtain illness onset times. Illness onset times prior to d are excluded, and times after $d + c$ are censored. A Cox proportional hazards model is fit to the event times, with a single binary predictor for trial arm and rescaled so that d is time 0. We retain the p-value and VE as 1 minus the estimated hazard ratio, and repeat this process 25,000 times per scenario. Average simulated VE and power for an $\alpha = 0.05$ test are compared to the values predicted by our closed-form approximations.

Results are summarized in Table 1, and we see that approximated VE generally agrees with the simulated average VE within a few percent, though our approximation slightly overestimates apparent VE. The bias approximation performs best when the background infection hazard rate $\lambda_W(\cdot)$ is highest and/or the sample size per arm n is largest. The approximation does not consider censoring due to losses to follow-up during the trial period; if censoring rates are high, there may be a larger discrepancy between approximated and simulated VE [Xu and O'Quigley (2000)].

For the power approximation, generally there is good agreement ($\sim 2\%$ absolute difference) between simulated and approximated power, though the approximation tends to slightly overestimate power, consistent with overestimating apparent VE. The approximation performs poorly when the expected number of events in Arm 1, intervention arm, is small, for example, less than 5, and can perform poorly when the sample size is small, for example, $n = 500$, as can be noted in further simulation results provided in the Supplementary Materials, Dean, Halloran and Longini (2018).

5. Results.

5.1. *Trials with an unvaccinated control arm.* After establishing methods for approximating the apparent VE and power for a given trial design, we investigate how to select the analysis period start d and length c in the simple setting of a vaccine trial in which Arm 0 is never vaccinated with the candidate vaccine. We

TABLE 1

Assessment of bias and power approximations. Sample size $n = 1000$ per arm. Two constant infection hazard rates λ_W are considered: (1) $\lambda_W = 0.001$, yielding a 3% attack rate in Arm 0, and (2) $\lambda_W = 0.01$, yielding a 30% attack rate in Arm 0. Incubation period has mean of 6 days (scale parameter = 1). No ramp-up period is included ($R = 0$). No delayed vaccination in Arm 0 ($b = \infty$). Analysis window of $c = 30$ days starting at either $d = 0, 6$ or 12 . 25,000 simulations run for each scenario. *Power approximation function returns 0.025 assuming other rejection region negligible; we multiply by two. #Expected number of events in Arm 1 is less than 5

	Infection hazard rate	d	Apparent VE		Apparent power	
			Simulated	Approximated	Simulated	Approximated
$VE_0 = 0\%$	$\lambda_W = 0.001$	0	-0.038	0	0.047	0.050*
		6	-0.039	0	0.048	0.050*
		12	-0.039	0	0.048	0.050*
	$\lambda_W = 0.01$	0	-0.006	0	0.049	0.050*
		6	-0.006	0	0.049	0.050*
		12	-0.007	0	0.049	0.050*
$VE_0 = 50\%$	$\lambda_W = 0.001$	0	0.376	0.400	0.380	0.399
		6	0.464	0.484	0.545	0.562
		12	0.480	0.499	0.578	0.593
	$\lambda_W = 0.01$	0	0.387	0.400	0.990	0.997
		6	0.479	0.484	1.000	1.000
		12	0.496	0.499	1.000	1.000
$VE_0 = 90\%$	$\lambda_W = 0.001$	0	0.708	0.720	0.937	0.928
		6	0.866	0.871	0.974#	0.993
		12	0.895	0.899	0.943#	0.996
	$\lambda_W = 0.01$	0	0.707	0.720	1.000	1.000
		6	0.869	0.871	1.000	1.000
		12	0.898	0.899	1.000	1.000

consider varying the analysis period start d while fixing the analysis period end at 50 days, which could occur if the budget only supports follow-up for a fixed time from randomization, and there is incentive to maximize the use of this follow-up time. In Panel A of Figure 3, recalling that $R = 4$ days, we see that d below $R + u_{0.90}$ is associated with bias because it includes a period when the vaccine has not yet reached maximal protection and apparent VE is biased (see Figure 1). Interestingly, in Panel B, the power maximizes at an analysis period start d for which there is some bias in the apparent VE. This is an example of a bias-variance tradeoff, with an earlier start d capturing more events during the period of partial efficacy, thereby increasing power while slightly biasing apparent VE. This tradeoff is especially notable when power is limited because of an early analysis period end or a background infection hazard that is decreasing over time.

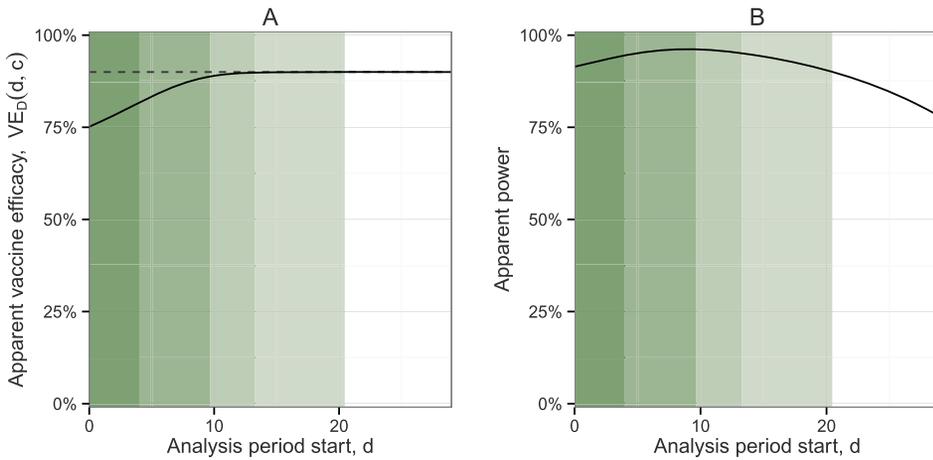


FIG. 3. Arm 1 vaccinated on day $s_1 = 0$; Arm 0 never vaccinated ($b = \infty$). Constant background infection hazard rate $\lambda_W = 0.001$. Sample size $n = 500$ per arm. (Setting as in Figure 1.) Analysis period $[d, 50]$ with length $c = 50 - d$; a range of d values are considered. (A) Apparent VE to assess bias, where the horizontal dotted line indicates $VE_0 = 90\%$. (B) Apparent power.

We assess the impact of other key factors, including the incubation period distribution, the ramp-up period, VE_0 , and the background infection hazard rate. We summarize the results here and in Table 2. Figures for a wide range of scenarios are provided in the Supplementary Materials, Dean, Halloran and Longini (2018).

Longer incubation periods and ramp-up periods shift the optimal d but the same general relationships persist. Illness onsets following long incubation periods are most likely to contribute to bias, but the probability of these types of incubation periods is reflected in the distribution’s quantiles, $u_{0.90}$ and $u_{0.999}$. In this scenario, to minimize bias, an appropriate starting point for the delay could be $R + u_{0.90}$ or $R + u_{0.999}$. For longer ramp-up periods, the power tends to maximize at earlier values of d when the ramp-up period is longer because there is a longer period of partial efficacy. If there is little protective effect until the end of the ramp-up period (e.g., ramp-up is not linear), we will need to select a later d because an earlier d might be prone to more bias (vaccinated group looks similar to unvaccinated group for longer). Conversely, fast-acting vaccines require an earlier d . In an extreme situation, a vaccine with post-exposure prophylactic effects will require the earliest d because vaccinees may be protected from disease even if already infected.

We also consider varying VE_0 , observing a change in power as expected, but only a minimal impact on bias. As VE_0 increases, the optimal d for minimizing bias increases very slightly. This can be explained because when VE_0 is highest, the apparent VE experiences the greatest change from 0; thus, including a period of partial efficacy when the vaccine is highly effective can induce the largest absolute bias. The effect of varying the background infection hazard rate $\lambda_W(w)$ is also minimal. In the constant hazard setting, there is no change in bias, though the

mean squared error decreases as the event rate increases. The effect of event rate on power is as expected, with greater power at the higher event rate, but the location of the optimal value of d is fairly constant.

Given the above results, it is clear that choosing a later analysis period start, d , is preferable for minimizing bias between the apparent VE and true VE_0 . A safe and commonly used option is to select d equal to $R + u_{0.999}$, the maximum ramp-up period plus the maximum incubation period, though using the 90th or 95th percentile of the incubation period distribution can also give good results. Nonetheless, our results suggest that if the goal is to maximize power, there are many settings where we may prefer an earlier value of d . Capturing cases during this period of partial vaccine efficacy may provide critical additional events, especially important if the background infection hazard rate is low and/or decreasing. Similarly, an earlier d may allow for a wider analysis period c , further increasing the expected number of events.

5.2. Trials with a delayed vaccination arm. Next, we consider delayed vaccination of the comparator arm, Arm 0. Using the framework described above, we investigate how to select an appropriate trial design, defined as the vaccination delay b , the analysis start period d , and the analysis period length c . The choice of d follows many of the same principles described in Section 5.1 and should be based on the expected ramp-up and incubation periods and background infection hazard rate, but the trial results are more sensitive to the choice of c when delayed vaccination is used. Issues of power are especially critical because cases must occur during the narrow window before the delayed arm is protected.

A natural starting point is to consider analysis periods that have length equal to the vaccination delay ($c = b$). This approach has a nice symmetry because the cutoff (d) applied to the immediate arm parallels the cutoff ($b + d$) applied to the delayed arm. In practice, if $c = b$, there is no value of d such that apparent VE_D is unbiased because any analysis period captures either unprotected immediate vaccinees or protected delayed vaccinees. For example, in Figure 2, there is no consecutive 21 day period during which apparent VE_T is unbiased, so an average over this period will also be biased, as shown in Panel A of Figure 4. Alternative approaches using $c < b$ could minimize bias, but the $c = b$ approach tends to have higher power; nonetheless, power can be highly sensitive to the choice of d , with a narrow maximum and steep decline on either side, leaving little tolerance for error in pre-specification of the analysis period. Power can be increased by increasing the sample size [see Figure S15 in Supplementary Materials, Dean, Halloran and Longini (2018)]. These results suggest that, if the primary goal is to reduce bias, we should decouple the vaccination delay b and the analysis period length c .

A natural way to decrease bias without negatively impacting power is to increase the length of the vaccination delay b . In Figure 5, we increase the vaccination delay to $b = 35$ days while maintaining an analysis period of length $c = 21$. By increasing the vaccination delay, there is a longer period of time in which Arm 1

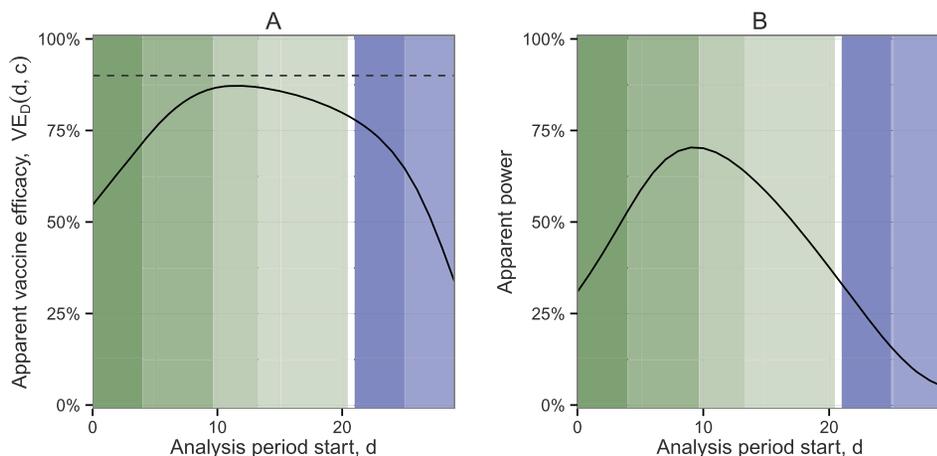


FIG. 4. Arm 1 vaccinated on day $s_1 = 0$; Arm 0 vaccinated on day $s_1 = 21$ ($b = 21$). Constant background infection hazard rate $\lambda_W = 0.001$. Sample size $n = 500$ per arm. (Setting as in Figure 2.) Analysis period $[d, d + 21)$ with length $c = 21$; a range of d values are considered. (A) Apparent VE to assess bias, where the horizontal dotted line indicates $VE_0 = 90\%$. (B) Apparent power.

is protected and Arm 0 is not protected. Thus, there is more observation time in which the apparent $VE_T(t)$ is equal to the true VE_0 (Panel C). As a result, we note that it is possible to identify a 21-day analysis period in which there is little to no bias (Panel D), and bias and power are less sensitive to the choice of d (Panels D and E). This stable region occurs roughly for d values between $R + u_{0,90}$ to $R + u_{0,999}$. This has the advantage that you have more tolerance for error when pre-specifying the delay. As described above, a more powerful approach would be to then similarly increase the length of the analysis period c ; this induces bias, though less bias than when $b = 21$ because $VE_T(t) = VE_0$ for a larger proportion of the analysis period.

We assess the impact of other key factors, including the incubation period distribution, the ramp-up period, VE_0 , and the background infection hazard rate. Figures for a wide range of scenarios are provided in the Supplementary Materials, [Dean, Halloran and Longini \(2018\)](#). In Figure 6, we consider the impact of the underlying vaccine efficacy VE_0 . The optimal value d is similar across all levels of VE_0 , though it is slightly later when VE_0 is highest. This occurs because high VE_0 requires the largest change in illness onset hazard ratio, meaning that the period of partial efficacy can induce the most bias in apparent VE. It is also noteworthy that when $VE_0 = 100\%$, larger values of d do not induce bias in apparent VE as no additional cases are observed in the immediately vaccinated arm.

In general, if bias is the primary concern, we suggest selecting a large b and $c < b$. Furthermore, we do not recommend counting events past $b + R + u_{0,50}$, or, more stringently, $b + u_{0,50}$. In some settings, regulators may question counting any events in Arm 0 following delayed vaccination; thus the analysis period end can

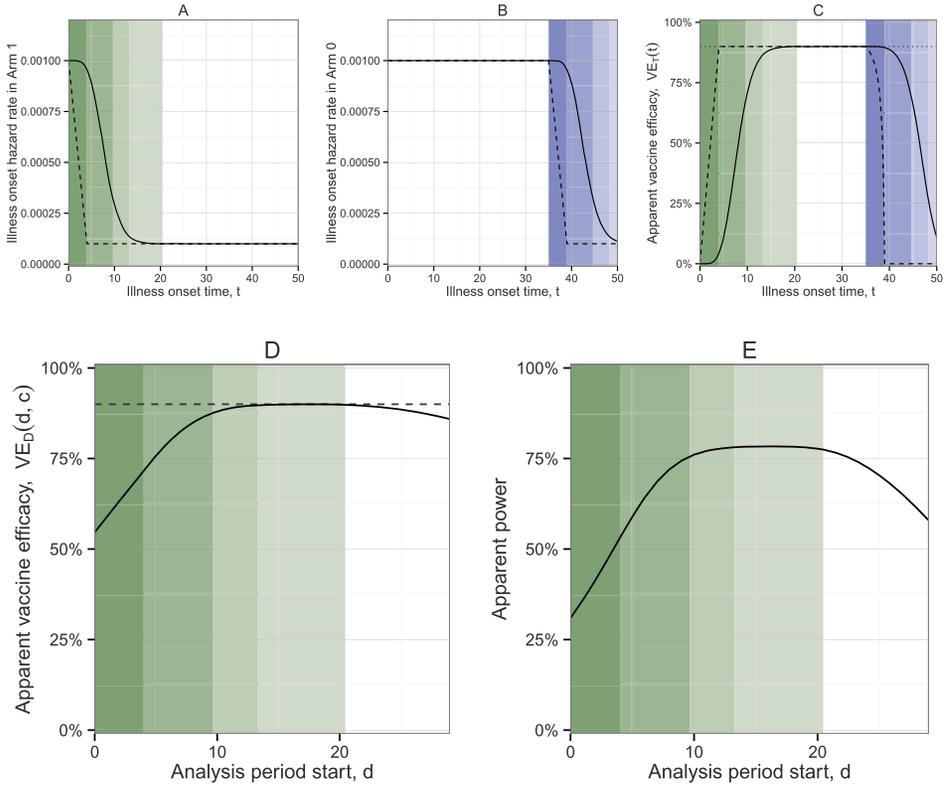


FIG. 5. Arm 1 vaccinated on day $s_1 = 0$; Arm 0 vaccinated on day $s_1 = 35$ ($b = 35$). Constant background infection hazard rate $\lambda_W = 0.001$. Sample size $n = 500$ per arm. (A) Solid line indicates the illness onset hazard rate in Arm 1 as a function of time, $\lambda_T(t|s_1, \mathbf{X})$; dashed line indicates the illness onset hazard rate in Arm 0 as a function of time, $\lambda_T(t|s_0, \mathbf{X})$; dashed line indicates the infection onset hazard rate in Arm 1. (C) Apparent VE ($1 -$ illness onset hazard ratio) as a function of time; dashed line ($1 -$ infection hazard ratio); horizontal dotted line indicates $VE_0 = 90\%$. (D) Apparent VE to assess bias with analysis period $[d, d + 21)$ with length $c = 21$; a range of d values are considered. (E) Apparent power.

be set at b . Selecting $d = R + u_{0.90}$ should reasonably minimize bias, though $d = R + u_{0.999}$ could be selected as a more stringent option. Note that power may be severely impacted if the most stringent options are applied. If maximizing power is the primary concern, selecting $c = b$ is recommended to capture the greatest number of events, and $d = R + u_{0.50}$ is proposed as a starting value.

5.3. *General recommendations.* We provide qualitative trends, but we recommend that investigators prepare their own setting-specific plots. Particularly in the context of emerging infectious diseases, the disease and/or vaccine may not be well characterized when the trial protocol is developed. We want the success of the trial

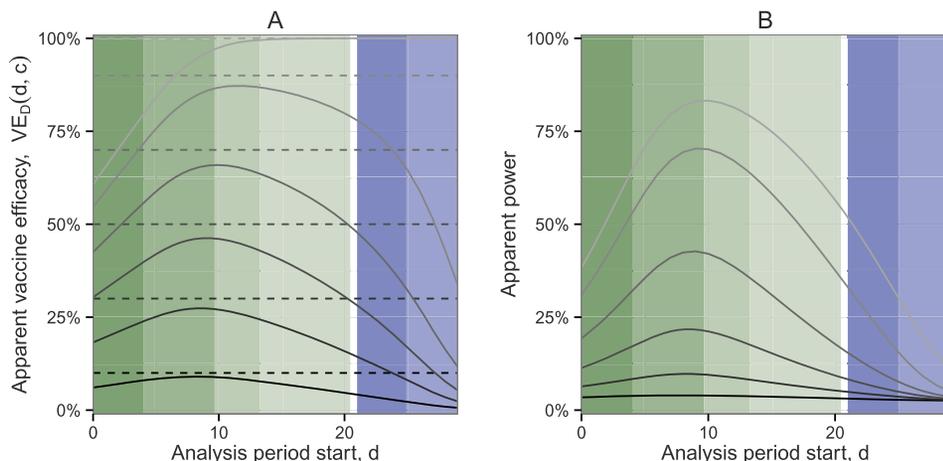


FIG. 6. Arm 1 vaccinated on day $s_1 = 0$; Arm 0 never vaccinated ($b = \infty$). Constant background infection hazard rate $\lambda_W = 0.001$. Sample size $n = 500$ per arm. From darkest to lightest, includes $VE_0 = 10\%$, 30% , 50% , 70% , 90% , and 100% , as indicated by horizontal dotted lines. Analysis period $[d, d + 21]$ with length $c = 21$; a range of d values are considered. (A) Apparent VE to assess bias. (B) Apparent power.

to be reasonably robust to the selected analysis period. They should first define ranges of reasonable values for (1) background infection hazard rate, (2) disease incubation period distribution, (3) time for full vaccine immune response to develop, and (4) predicted VE_0 . Starting with the most likely values for each of the above and an arbitrary sample size, the investigator can construct bias and power plots. From these plots, they should first identify potential trial designs, $[d, d + c]$, with minimal bias. Investigators may then modify the sample size until acceptable power is achieved. Investigators should avoid trial designs for which the bias and/or power have narrow peaks at certain values of d , and recall that sometimes these are not the same values for bias and for power. Finally, the initial assumptions should be modified to cover the range of reasonable values above, ensuring that the trial design is reasonable for all combinations within these ranges; the trial design and/or sample size may need to be modified such that it is robust to assumption misspecification.

Trials with a vaccination delay will need to further consider the choice of delay b . Qualitative recommendations are provided in Table 2. In general, smaller values of b may be more ethically advantageous, but these values will lead to higher bias and lower power.

6. Ebola vaccine trial example. The interim analysis of the “Ebola ça suffit” ring vaccination trial in Guinea provides a real application of the principles described above [Henao-Restrepo et al. (2015)]. In this trial, a delayed vaccination arm was used with vaccination occurring 21 days after randomization as compared

TABLE 2

Recommended trial design modifications, where d is the start of the analysis period, c is the length of the analysis period, b is the vaccination delay, and n is the sample size per arm

Scenario	Suggested approach(es)
Long vaccine ramp-up period	$\uparrow d, \uparrow b$
Long disease incubation period	$\uparrow d, \uparrow b$
Higher variability in incubation period	$\uparrow b$
Low background infection hazard rate	$\uparrow c, \uparrow b, \uparrow n$
Decreasing background infection hazard rate	$\downarrow d$
Minimize bias	$\uparrow b, c < b$
Maximize power	$\uparrow b, c = b, \uparrow n$

to an arm vaccinated immediately after randomization. The analysis period start was set at $d = 10$ days, excluding cases with illness onset between 0 and 9 days after randomization from the per protocol analysis. As no cases were observed in vaccinees in either arm more than 6 days after vaccination, the estimated VE was 100% (95% CI: 75.1, 100%). In fact, the estimated VE is largely insensitive to the choice of d , remaining 100% for any d after the last case in an immediately vaccinated individual, as indicated in Figure 6. Similarly, the estimated VE is insensitive to the analysis period end, which occurs only when the vaccine is so efficacious that no further cases accumulate in the immediately vaccinated arm; thus, the estimated VE is 100% regardless of when the analysis is stopped. A final analysis of the trial was conducted including additional participants, and the same key results were observed [Henao-Restrepo et al. (2017)]. As this paper demonstrates, a risk of delayed vaccination designs is bias toward the null; with VE of 100%, this type of bias was not observed in the trial. The power, on the other hand, is sensitive to the choice of d , with earlier values of d yielding greater power because more events in the delayed arm are retained.

7. Discussion. In this paper, we present a framework for selecting the analysis period for vaccine efficacy trials. This framework accounts for the facts that only illness (symptom) onset times are observed, and illness onsets shortly after vaccination may reflect infections before vaccination or before the immune system developed protection. A sensitive baseline test, if available, can help identify infections occurring prior to vaccination, but it cannot identify infections occurring before the immune response has developed. The per protocol analysis seeks to exclude these early cases, along with participants who do not follow trial protocol, to achieve an unbiased estimate of vaccine efficacy. An unbiased estimate of vaccine efficacy is critical for deciding whether the vaccine product should be licensed. Delayed vaccination trials may return an attenuated vaccine efficacy estimate. We provide closed-form approximations for predicting the bias and power

for a given analysis period, and we describe the bias/variance tradeoff associated with counting these early cases.

Delayed vaccination can be used as a comparator arm in settings where placebo or vaccination for an unrelated disease is not considered ethically acceptable. The design also gives more participants access to vaccine, potentially averting additional cases and enhancing disease control efforts if the vaccine is efficacious. This approach may be preferred for diseases with high case fatality rates, such as Ebola virus disease. Using a delayed vaccination comparator necessarily decreases study power as it limits the time when the two arms can be meaningfully contrasted for estimation of vaccine efficacy. Once delayed vaccination participants are protected by vaccine, the analysis must stop because no appropriate control remains; this can complicate how cases occurring later in the trial are interpreted. Other challenges of the approach include that blinding may be difficult to achieve. If availability of vaccine is a limiting factor, it is noteworthy that delayed vaccination trials require more doses as participants in both arms are vaccinated. Finally, if the vaccine candidate requires multiple doses or takes a long time to develop efficacy, the delay would need to be very long to achieve the desired power, thereby reducing the ethical advantage of the approach.

The framework we provide is intended to support those designing vaccine efficacy trials in calculating sample size, power, and pre-specifying the appropriate analysis period. As for any sample size/power calculation, this approach requires a number of assumptions for which limited information may be available. The closed-form approximations had small discrepancies from simulation results, especially when the sample size was small (<500 per arm). The underlying model assumes independence and does not reflect the complicated nature of infectious disease dynamics, including indirect vaccine effectiveness. In individually randomized trials, indirect vaccine effects should similarly affect both arms and would not bias the results, but indirect effects would likely be differential across arms in cluster randomized trials. The impact of heavy censoring is also not captured, which may induce further bias. We suggest using the closed-form approximations to narrow down the space of trial designs while planning and using more realistic simulations as a confirmatory step. Other limitations include that the ramp-up period is assumed fixed and everyone is assumed to be vaccinated on the same day; these could be readily modified by converting these constants to random variables.

Immunological data from early phase trials can be used to support the design of a vaccine efficacy trial. If these data are rapidly accumulating, as in an emergency setting, and are not available at the time the protocol is written, the statistical analysis plan could include a clause allowing flexibility in the specification of d pending external data. Alternatively, a different value of d could be pre-specified as a secondary analysis. More sophisticated approaches could be used than applying hard cutpoints for the analysis period, though typically simpler approaches are preferred for clinical trial primary endpoints. The incubation and immune ramp-up periods could be explicitly included in the likelihood; if this lag is known, it can

be used to construct optimal weights [Zucker and Lakatos (1990)]. Furthermore, we always recommend reporting survival (or cumulative incidence) curves so that any evidence of an early harmful effect is not missed [Hernán (2013)].

Moving forward, the international community has recognized the need for specific guidance on how to design vaccine trials in emergency settings [Kieny and Salama (2017)]. For emerging pathogens with high case fatality rates, delayed vaccination may be a desirable strategy when evaluating vaccine candidates. This guidance is intended to support those implementing this approach.

Acknowledgments. The authors are grateful to the World Health Organization, Ana Maria Henao-Restrepo and the “Ebola ça Suffit” trial consortium.

SUPPLEMENTARY MATERIAL

Supplement to “Design of vaccine trials during outbreaks with and without a delayed vaccination comparator” (DOI: [10.1214/17-AOAS1095SUPP](https://doi.org/10.1214/17-AOAS1095SUPP); .zip). The supplementary materials contains additional plots and tables under a range of scenarios. Supporting R code is also provided to produce estimates and plots.

REFERENCES

- BELLAN, S. E., PULLIAM, J. R. C., PEARSON, C. A. B., CHAMPREDON, D., FOX, S. J., SKRIP, L., GALVANI, A. P., GAMBHIR, M., LOPMAN, B. A., PORCO, T. C., MEYERS, L. A. and DUSHOFF, J. (2015). Statistical power and validity of Ebola vaccine trials in Sierra Leone: A simulation study of trial design and analysis. *Lancet, Infect. Dis.* **15** 703–710.
- EBOLA ÇA SUFFIT RING VACCINATION TRIAL CONSORTIUM (2015). The ring vaccination trial: A novel cluster randomised controlled trial design to evaluate vaccine efficacy and effectiveness during outbreaks, with special reference to Ebola. *Br. Med. J.* **351** h3740.
- CAMACHO, A., EGGO, R. M., FUNK, S., WATSON, C. H., KUCHARSKI, A. J. and EDMUNDS, W. J. (2015). Estimating the probability of demonstrating vaccine efficacy in the declining Ebola epidemic: A Bayesian modelling approach. *BMJ Open* **5** e009346.
- DEAN, N., HALLORAN, M. E. and LONGINI, I. M. (2018). Supplement to “Design of vaccine trials during outbreaks with and without a delayed vaccination comparator.” DOI:10.1214/17-AOAS1095SUPP.
- GILBERT, P. B., BERGER, J. O., STABLEIN, D., BECKER, S., ESSEX, M., HAMMER, S. M., KIM, J. H. and DEGRUTTOLA, V. G. (2011). Statistical interpretation of the RV144 HIV vaccine efficacy trial in Thailand: A case study for statistical issues in efficacy trials. *J. Infect. Dis.* **203** 969–975.
- HALLORAN, M. E., LONGINI, I. M. JR. and STRUCHINER, C. J. (2010). *Design and Analysis of Vaccine Studies*. Springer, New York.
- HALLORAN, M. E., HABER, M., LONGINI, I. M. and STRUCHINER, C. J. (1991). Direct and indirect effects in vaccine efficacy and effectiveness. *Am. J. Epidemiol.* **133** 323–331.
- HENAO-RESTREPO, A. M., LONGINI, I. M., EGGER, M., DEAN, N. E., EDMUNDS, W. J., CAMACHO, A., CARROLL, M. W., DOUMBIA, M., DRAGUEZ, B., DURAFFOUR, S., ENWERE, G., GRAIS, R., GUNTHER, S., HOSSMANN, S., KONDE, M. K., KONE, S., KUISMA, E., LEVINE, M. M., MANDAL, S., NORHEIM, G., RIVEROS, X., SOUMAH, A., TRELLE, S., VICARI, A. S., WATSON, C. H., KEITA, S., KIENY, M.-P. P. and RÖTTINGEN, J. A. (2015). Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: Interim results from the Guinea ring vaccination cluster-randomised trial. *Lancet* **386** 857–866.

HENAO-RESTREPO, A. M., CAMACHO, A., LONGINI, I. M., WATSON, C. H., EDMUNDS, W. J., EGGER, M., CARROLL, M. W., DEAN, N. E., DATTA, I., DOUMBIA, M., DRAGUEZ, B., DURAFFOUR, S., ENWERE, G., GRAIS, R., GUNTHER, S., GSELL, P.-S., HOSSMANN, S., WATTLE, S. V., KONDE, M. K., KEITA, S., KONE, S., KUISMA, E., LEVINE, M. M., MANDAL, S., MAUGET, T., NORHEIM, G., RIVEROS, X., SOUMAH, A., TRELLE, S., VICARI, A. S., RØTTINGEN, J. A. and KIENY, M.-P. (2017). Efficacy and effectiveness of an rVSV-vectored vaccine preventing Ebola virus disease: Final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola ça suffit!). *Lancet* **389** 505–518.

HERNÁN, M. A. (2013). The hazards of hazard ratios. *Epidemiology* **21** 13–15.

HORNE, A. D., LACHENBRUCH, P. A. and GOLDENTHAL, K. L. (2001). Intent-to-treat analysis and preventive vaccine efficacy. *Vaccine* **19** 319–326.

HUSSEY, M. A. and HUGHES, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemp. Clin. Trials* **28** 182–191.

KIENY, M. P. and SALAMA, P. (2017). WHO R&D blueprint: A global coordination mechanism for R&D preparedness. *Lancet* **389** 2469–2470.

RTSS CLINICAL TRIALS PARTNERSHIP (2015). Efficacy and safety of RTS, S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3, individually randomised, controlled trial. *Lancet* **386** 31–45.

RIDOUT, M. S., DEMÉTRIO, C. G. and FIRTH, D. (1999). Estimating intraclass correlation for binary data. *Biometrics* **55** 137–148.

ROSNER, B. (2010). *Fundamentals of Biostatistics*, 7th ed. Cengage Learning, Boston.

R CORE TEAM (2015). *A Language and Environment for Statistical Computing*. Available at <http://www.r-project.org/>.

THERNEAU, T. M. (2015). Package ‘survival’: Survival analysis. R package. Available at <http://cran.r-project.org/web/packages/survival/index.html>.

XU, R. and O’QUIGLEY, J. (2000). Estimating average regression effect under non-proportional hazards. *Biostatistics* **1** 423–439.

ZUCKER, D. M. and LAKATOS, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika* **77** 853–864. MR1086695

N. E. DEAN
 I. M. LONGINI
 DEPARTMENT OF BIOSTATISTICS
 UNIVERSITY OF FLORIDA
 GAINESVILLE, FLORIDA 32611
 USA
 E-MAIL: nataliedean@ufl.edu

M. E. HALLORAN
 VACCINE AND INFECTIOUS DISEASE DIVISION
 FRED HUTCHINSON CANCER
 RESEARCH CENTER
 USA
 AND
 DEPARTMENT OF BIOSTATISTICS
 UNIVERSITY OF WASHINGTON
 SEATTLE, WASHINGTON 98109
 USA