NONPARAMETRIC INFERENCE FOR IMMUNE RESPONSE THRESHOLDS OF RISK IN VACCINE STUDIES[1]

BY KEVIN M. DONOVAN[*], MICHAEL G. HUDGENS[*] AND
PETER B. GILBERT[†]

*University of North Carolina at Chapel Hill[*] and Fred Hutchinson Cancer
Research Center[†]*

An important objective in vaccine studies entails identifying an immune response which is predictive of disease risk. Nonparametric methods are developed for inference on immune response thresholds that are associated with specified levels of disease risk, including where the risk level is zero. This threshold is defined as the minimum immune response value above which disease risk is less than or equal to the desired level. The proposed nonparametric methods are compared to previously developed parametric methods in simulation studies. The methods are extended for use in studies that only measure the immune response in a subset of participants, such as case-cohort or case-control studies, and with right censored time to disease outcomes. Finally, these methods are used to estimate neutralizing antibody thresholds for virologically confirmed dengue risk using data from two recent dengue vaccine trials.

1. Introduction. Infectious disease is one of the most prevalent causes of death globally. In 2010 approximately 15 million deaths worldwide were caused directly by infectious disease (Dye (2014)). Historically, the use of vaccines to reduce the burden of infectious diseases has been extremely effective. Previously devastating illnesses such as smallpox and measles have been dramatically reduced, and sometimes eradicated completely, due to vaccination (Centers for Disease Control and Prevention (2014)). However, infectious diseases such as HIV/AIDS, malaria and tuberculosis continue to cause millions of deaths worldwide as their treatment and prevention have proved difficult (Centers for Disease Control and Prevention (2014)). As a result, it is imperative that the effectiveness of vaccines continues to be studied.

One goal of vaccine research entails finding an immune response which is predictive of the clinical outcome of interest. Here we consider a binary outcome which is generically referred to as *disease*. The inferential goal is to estimate the minimum immune response value above which the risk of disease is less than or

equal to some specified value (e.g., 0.01%). This value of immune response is referred to as a *risk threshold* and the specified risk upper bound is referred to as a *risk level*. Inference about these thresholds can be useful to regulatory agencies, policymakers and public health officials in assessing disease risk in a vaccinated population. In particular, for the sake of cost and efficiency, studies of new formulations of existing vaccines and studies of existing vaccines in new populations often collect data only on immune responses and do not also assess disease outcomes. In these studies, estimated risk thresholds can be used to predict disease risk for the new vaccine formulation or in the new population.

There are two common approaches employed in vaccine efficacy trials to identify immune response thresholds of risk. The first approach entails fitting parametric or semiparametric models to quantify an immune response's association with disease risk (e.g., Chan et al. (2002), Dunning (2006), Jokinen and Åhman (2004), Storsaeter et al. (1998), White et al. (1992), Haynes et al. (2012)) and these methods can be adapted for inference on risk thresholds. However, such models are subject to misspecification, especially when drawing inference about immune response thresholds associated with zero disease risk. The second approach considers specifically the zero risk threshold. This approach (e.g., Andrews, Borrow and Miller (2003), Borrow, Balmer and Miller (2005), Jódar et al. (2003), Black et al. (2011)) assumes the existence of an immune response threshold which perfectly discriminates whether an individual is protected from disease. In addition, disease risk is assumed to be independent of vaccination status given the immune response, that is, Prentice's (Prentice (1989)) full mediation condition for a valid surrogate endpoint is assumed. In contrast, this paper introduces methods for inference about risk thresholds corresponding to specified disease risk levels which are nonparametric and do not require Prentice's full mediation condition to hold.

The remainder of this paper's structure is as follows. Section 2 details inference about a risk threshold in a standard vaccine study design. Section 3 describes simulation studies evaluating the inferential methods described in Section 2. Section 4 extends the methods to studies where the immune response is only measured in a subset of participants, such as case-cohort or case-control studies, and where some individuals' disease outcomes are censored. Section 5 presents an application of these methods to two recent dengue vaccine trials. Section 6 concludes the paper and offers further discussion.

2. Methods.

2.1. *Risk threshold.*   Suppose a vaccine is evaluated in a randomized trial, possibly placebo-controlled, in which all participants are initially disease-free and n individuals receive vaccine. (Unless indicated otherwise, the methods described below only utilize data observed from the n vaccinated individuals. Thus the methods can also be applied in observational cohort studies where a random sample of

n vaccinated individuals is followed prospectively.) Let Z denote treatment assignment with $Z = 1$ denoting vaccine and $Z = 0$ denoting control, where *control* refers to a placebo or absence of treatment. Suppose there is an immune response associated with the likelihood of developing disease. Let S denote an individual's immune response, with S measured after the individual receives their assigned treatment (vaccine or control). Let Y denote an individual's outcome, with $Y = 1$ denoting disease and $Y = 0$ otherwise; assume Y is measured some time after S. While for simplicity we assume that all n participants have S measured before Y occurs, under a random censoring assumption the methods also apply for inference on participants at-risk for Y when S is measured.

For risk level $c \in [0, 1]$, define the risk threshold when vaccinated to be

$$v_c = \inf\{v : \Pr(Y = 1 | Z = 1, S \geq v) \leq c\}.$$

In other words, the risk of disease is no greater than c in the stratum of individuals with immune response at least v_c. Such thresholds are useful for understanding the risk of disease in a vaccinated population. For example, suppose $c = 0.01$ and $\Pr(S \geq v_c | Z = 1) = 1$; then the risk of disease in the vaccinated population is at most 1%. As described in the Introduction, these thresholds have utility in assessing disease risk at the population level in settings where only immune response data are available.

An alternative target parameter might be $v_{\text{alt}} = \inf\{v : \Pr(Y = 1 | Z = 1, S = v) \leq c\}$. This alternative parameter has a straightforward interpretation at the individual level, providing an upper bound on disease risk for an individual with immune response v_{alt}. However, this parameter does not have a straightforward population level interpretation, and thus is of less utility to regulatory agencies or policymakers regarding the population level disease risk when vaccinated. An exception is in settings where it is plausible to assume there is a monotonic relationship between the immune response and disease risk, that is, $\Pr[Y = 1 | S = v, Z = 1]$ is a nonincreasing function of v. If monotonicity is assumed, then disease risk would be at most c in the stratum of individuals with immune response at least v_{alt}. Below nonparametric methods are proposed which do not require such a monotonicity assumption, and thus v_c is the target parameter of interest.

Below, methods are considered for estimation and confidence intervals (CIs) of v_c. Assume n independent copies of (S, Y) are observed in the vaccinated individuals. The estimators of v_c considered below are motivated by the equalities

$$(1) \quad \Pr(Y = 1 | S \geq v) = \frac{\Pr(Y = 1, S \geq v)}{\Pr(S \geq v)} = \frac{\int_v^\infty \Pr(Y = 1 | S = s) \, dF_S(s)}{\int_v^\infty dF_S(s)},$$

where $F_S(s)$ is the cumulative distribution function (CDF) of S. Here and elsewhere distributions are conditional on $Z = 1$ unless stated otherwise. The denominator of (1) can be estimated consistently by $\sum_{i=1}^n I(S_i \geq v)/n$. Nonparametric and parametric estimators for the numerator of (1) are considered below, which in

turn lead to estimators of v_c. For now assume $c > 0$; the special case where $c = 0$ is considered in Section 2.2.

Note $\Pr(Y = 1, S \geq v)$ can be estimated consistently by $\sum_{i=1}^n I(Y_i = 1, S_i \geq v)/n$, which motivates the following proposed nonparametric estimator of v_c:

$$(2) \qquad \hat{v}_c = \min\left\{ S_j : \frac{\sum_{i=1}^n I(Y_i = 1, S_i \geq S_j)}{\sum_{i=1}^n I(S_i \geq S_j)} \leq c, j \in \{1, \ldots, n\} \right\}.$$

In words, $\hat{v}_c$ is the minimum v of the set of observed immune responses such that in the stratum of individuals with immune response at least v, the proportion developing disease is no greater then c. Note that $\hat{v}_c$ is not well defined for c such that the estimated disease risk conditional on $S \geq S_j$ exceeds c for all $j = 1, \ldots, n$. In particular, let $\hat{\Pr}(Y = 1 | S \geq S_j) = \sum_{i=1}^n I(Y_i = 1, S_i \geq S_j)/\sum_{i=1}^n I(S_i \geq S_j)$ and $\hat{p}_{\min} = \min\{\hat{\Pr}(Y = 1 | S \geq S_j) : j = 1, \ldots, n\}$. For $c < \hat{p}_{\min}$, $\hat{v}_c$ is not well defined because the right side of (2) equals the minimum of the empty set. Examples from two simulated data sets are shown in Figure 1. These data were generated from logistic regression models of the form $\Pr[Y = 1 | S \geq v] = \text{expit}(\beta_0 + \beta_1 v)$ for different values of β_0 and β_1, where expit denotes the inverse of the logit function. In the left panel of Figure 1, $\hat{p}_{\min} = 0$ such that $\hat{v}_c$ is well defined for all $c \in [0, 1]$. On the other hand, in the right panel of Figure 1, $\hat{p}_{\min} > 0$ so $\hat{v}_c$ is only well defined for $c \in [\hat{p}_{\min}, 1]$. Note that $\hat{v}_c$ being well defined is separate from existence of v_c, as $\hat{v}_c$ not being well defined may be a result of sampling variation or insufficient sample size.
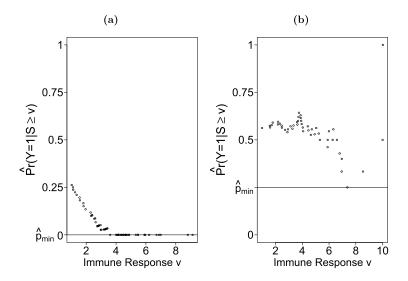

FIG. 1. *Panels (a) and (b) display nonparametric estimates of disease risk conditional on $S \geq v$ for each observed v from two simulated data sets. These estimates are denoted $\hat{\Pr}(Y = 1 | S \geq v)$. For each data set, the minimum of these estimates is denoted by $\hat{p}_{\min}$.*

The estimator $\hat{v}_c$ is compared with parametric methods that estimate $p(v) = \Pr(Y = 1 | S = v)$ using a regression model, denoted $\hat{p}(v)$. These parametric estimators of v_c are of the form

$$(3) \qquad \min\left\{ S_j : \frac{\sum_{i=1}^n I(S_i \geq S_j)\hat{p}(S_i)}{\sum_{i=1}^n I(S_i \geq S_j)} \leq c, \, j \in \{1, \ldots, n\} \right\}.$$

Following Storsaeter (Storsaeter et al. (1998)), $p(v)$ may be modeled by $p(v) = \mathrm{expit}(\beta_0 + \beta_1 v)$. Define the parametric estimator $\hat{v}_c^l$ by (3) with $\hat{p}(v) = \mathrm{expit}(\hat{\beta}_0 + \hat{\beta}_1 v)$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ are maximum likelihood estimators of β_0 and β_1.

Dunning's scaled logit model (Dunning (2006)) suggests another parametric estimator. Suppose individuals are either immune or susceptible to disease. Let λ denote the risk of disease given an individual is susceptible. Let $\pi(v)$ denote the probability an individual is susceptible given an immune response v and suppose $p(v) = \lambda \pi(v)$, where $\pi(v) = \mathrm{expit}(\beta_{0d} + \beta_{1d} v)$. Denote the parametric estimator corresponding to Dunning's scaled logit model by $\hat{v}_c^d$, defined as (3) with $\hat{p}(v) = \hat{\lambda} \, \mathrm{expit}(\hat{\beta}_{0d} + \hat{\beta}_{1d} v)$, where $\hat{\lambda}$, $\hat{\beta}_{0d}$, and $\hat{\beta}_{1d}$ are maximum likelihood estimators of λ, β_{0d} and β_{1d}.

For each of the three estimators described above, a corresponding $(1 - \alpha)$ CI for v_c can be computed using the bootstrap percentile method (Efron and Tibshirani (1993)) as follows. First, B independent bootstrap samples, each consisting of n copies of (S, Y), are drawn with replacement from $\{(S_1, Y_1), \ldots, (S_n, Y_n)\}$. Then for each of the B samples, the estimator for v_c is calculated. A $(1 - \alpha)$ CI for v_c is given by the $\alpha/2$ and $(1 - \alpha/2)$ percentiles of the B estimates of v_c.

2.2. *Threshold for zero risk.* Recall v_0 is the lowest immune response threshold above which an individual has no risk of disease when vaccinated. When v_0 exists, the distribution of S given $Y = 1$ is truncated with an upper truncation point at v_0 such that $f_S(v | Y = 1) = 0$ for $v \geq v_0$, where $f_S(v | Y = 1)$ is the conditional density of S given $Y = 1$. The parametric models discussed in Section 2.1 imply nonzero values for $\Pr(Y = 1 | S = v)$ for all v. As a result, these models are guaranteed to be misspecified when finite v_0 exists. While these parametric models could be extended to allow for zero-risk above some truncation point, the corresponding threshold estimators would not in general be expected to perform well when the parametric component of the model is misspecified. Below, three nonparametric estimators of v_0 are considered instead.

The first nonparametric estimator is $\hat{v}_0$, the estimator $\hat{v}_c$ from Section 2.1 when $c = 0$. Suppose there are m vaccinated individuals who develop disease (i.e., $Y = 1$) and let $S_{v(1)} < S_{v(2)} < \cdots < S_{v(m)}$ denote their ordered immune response values. Then $\hat{v}_0 = \min\{S_j : S_j > S_{v(m)}, \, j = 1, \ldots, n\}$.

The second nonparametric estimator considered uses Cooke's (Cooke (1979)) method for estimating an upper truncation point, defined as

$$\tilde{v}_0 = S_{v(m)} + \sum_{i=1}^{m-1} (i/m)^m (S_{v(i+1)} - S_{v(i)}).$$

This estimator is motivated by the general result that

$$(4) \qquad v_0 = E(S_{v(m)}) + \int_0^{v_0} F^m(s)\,ds,$$

where $F^m(s)$ is the CDF of $S_{v(m)}$. Due to independence, $F^m(s) = \{F(s)\}^m$ where $F(s)$ is the CDF of S given $Y = 1$. Plugging in the sample maximum and empirical CDF $\hat{F}(s)$ into the right side of (4) yields $\tilde{v}_0$. Cooke (Cooke (1979)) also derived an asymptotic $(1 - \alpha)$ CI for an upper truncation point, adopted here for v_0. This interval is denoted (L_0^c, U_0^c), where $L_0^c = S_{v(m)} + \{(\alpha/2)^{-1} - 1\}^{-1}(S_{v(m)} - S_{v(m-1)})$ and $U_0^c = S_{v(m)} + \{(1 - \alpha/2)^{-1} - 1\}^{-1}(S_{v(m)} - S_{v(m-1)})$.

Another approach to drawing inference about v_0 was proposed by Siber et al. (Siber et al. (2007)). Unlike the methods described above, this approach requires data from both vaccinated and control individuals. A key assumption of Siber et al.'s method is that $\Pr(Y = 1 | Z = z, S \le v_0) = \Pr(Y = 1 | S \le v_0)$ and $\Pr(Y = 1 | Z = z, S > v_0) = \Pr(Y = 1 | S > v_0)$ for $z = 0, 1$ (i.e., disease status is independent of treatment conditional on the immune response). In other words, Prentice's (Prentice (1989)) full mediation condition for a valid surrogate endpoint is assumed. Define vaccine efficacy (VE) as $\text{VE} = 1 - p_v/p_c$, where $p_v = \Pr(Y = 1 | Z = 1)$ and $p_c = \Pr(Y = 1 | Z = 0)$. Let $\hat{v}_0^s$ denote the estimator of v_0 proposed by Siber et al. To calculate $\hat{v}_0^s$, VE is first estimated using the sample proportions of disease for vaccine and control, denoted $\hat{p}_v$ and $\hat{p}_c$ respectively. Under the Prentice mediation assumption above, $\text{VE} = 1 - \Pr(S \le v_0 | Z = 1)/\Pr(S \le v_0 | Z = 0)$. Thus, if v_0 were known, VE could be estimated by $1 - \hat{\Pr}(S \le v_0 | Z = 1)/\hat{\Pr}(S \le v_0 | Z = 0)$, where $\hat{\Pr}(S \le v_0 | Z = z)$ denotes the proportion of individuals in treatment arm z with immune response less then or equal to v_0. Therefore, $\hat{v}_0^s$ is calculated by determining the immune response v such that $\hat{p}_v/\hat{p}_c = \hat{\Pr}(S \le v | Z = 1)/\hat{\Pr}(S \le v | Z = 0)$. Siber et al. also proposed a $(1 - \alpha)$ CI for v_0, denoted (L_0^s, U_0^s). Let $(L_{\text{VE}}, U_{\text{VE}})$ be a $(1 - \alpha)$ CI for VE. Then L_0^s is calculated by determining the immune response v such that $L_{\text{VE}} = 1 - \hat{\Pr}(S \le v | Z = 1)/\hat{\Pr}(S \le v | Z = 0)$. The upper limit U_0^s is calculated analogously using U_{VE}. This method has some disadvantages. It requires data on both vaccine and control and the assumption of Prentice's full mediation condition may not hold in practice. Furthermore, $\hat{v}_0^s$ may not exist or be unique.

3. Simulation studies.

3.1. *Inference on v_c for $c > 0$.*
This section reports the results of simulation studies evaluating the methods described in Section 2 for constructing point estimates and 95% CIs for specified risk thresholds. For inference on v_c for $c > 0$, data were simulated for n independent vaccinated individuals. Immune response S was generated from a gamma distribution with mean and variance 4. Disease outcome Y given S was generated from a Bernoulli distribution with $\Pr(Y = 1 | S = v)$

according to one of the four following models. The logit model was defined by $\Pr(Y = 1|S = v) = \text{expit}(\beta_0 + \beta_1 v)$ for $0 \leq v < \infty$. The probit model was defined by $\Pr(Y = 1|S = v) = \Phi(\beta_{0p} + \beta_{1p} v)$ for $0 \leq v < \infty$ where $\Phi(\cdot)$ is the CDF of the standard Normal distribution. The step model was defined by $\Pr(Y = 1|S = v) = \gamma$ for $0 \leq v \leq v_0$, $0 < \gamma < 1$ and $\Pr(Y = 1|S = v) = 0$ otherwise. Finally, the scaled logit model was defined by $\Pr(Y = 1|S = v) = \lambda \exp(\beta_{0d} + \beta_{1d} v)$ for $0 \leq v < \infty$ with $\lambda = 0.5$. Each model's parameters were determined by fixing the marginal disease risk when vaccinated, that is, p_v. For the logit, probit and scaled logit models, the slope parameter was set to -5 to reflect a higher immune response being associated with lower risk of disease. For each model, 1000 datasets were generated each with sample size $n = 12{,}500$. For each dataset, $\hat{v}_c$, $\hat{v}_c^l$ and $\hat{v}_c^d$ from Section 2 were calculated and corresponding 95% CIs for v_c were computed using the bootstrap percentile methods with $B = 2000$.

Table 1 presents the simulation results for each of the four models of $\Pr(Y = 1|S = v)$. The reported results include the empirical bias of each estimator, the average CI widths and the empirical CI coverage rates for each model.

In general, the empirical bias of $\hat{v}_c$ was negligible and the corresponding CIs had coverage rates approximately equal to the nominal level 0.95. In contrast, the parametric estimator $\hat{v}_c^l$ performed poorly when the assumed model was misspecified, with $\hat{v}_c^l$ having substantial bias and the associated CIs failing to achieve the nominal coverage level. While the parametric estimator $\hat{v}_c^d$ had low bias for all models, the associated CIs had coverage rates well below 0.95 in some scenarios when the assumed model was misspecified. In scenarios where the parametric models were correctly specified, CIs based on the nonparametric estimator tended to have similar width to CIs computed using the parametric estimator, indicating minimal loss of efficiency associated with using the nonparametric estimator.

3.2. *Inference on v_0.* For inference on v_0, data were simulated for n independent individuals, assigned to vaccine or control in a 1:1 allocation. Immune response S was generated from a gamma distribution with mean and variance 4 when vaccinated and with mean and variance 2 when assigned control. Disease outcome Y given S and Z was generated from a Bernoulli distribution using the logit, probit and step models. The logit model was defined by $\Pr(Y = 1|Z = 0, S = v) = \text{expit}(\gamma_0 + \gamma_1 v)$ for $0 \leq v < \infty$ and $\Pr(Y = 1|Z = 1, S = v) = \text{expit}(\beta_0 + \beta_1 v)$ for $0 \leq v \leq v_0$ with $\gamma_0 = \beta_0$ and $\beta_1 = -5$. The probit model was defined by $\Pr(Y = 1|Z = 0, S = v) = \Phi(\gamma_{0p} + \gamma_{1p} v)$ for $0 \leq v < \infty$ and $\Pr(Y = 1|Z = 1, S = v) = \Phi(\beta_{0p} + \beta_{1p} v)$ for $0 \leq v \leq v_0$ with $\gamma_{0p} = \beta_{0p}$ and $\beta_{1p} = -5$. Finally, the step model was defined by $\Pr(Y = 1|Z = z, S = v) = \omega$ for $0 \leq v \leq v_0$, $0 < \omega < 1$ and $\Pr(Y = 1|Z = z, S = v) = 0$ otherwise for $z = 0, 1$. Each model's parameters were determined by fixing the marginal disease risks when assigned control and when vaccinated, that is, p_c and p_v. For each model, 1000 datasets were generated each with sample size $n = 25{,}000$. For each dataset, $\hat{v}_0$, $\tilde{v}_0$ and

TABLE 1
Results of simulation study described in Section 2.1 based on 1000 *simulated datasets each with sample size* 12,500 *where Bias denotes empirical bias, Width denotes average CI width and Cov denotes empirical CI coverage rate. The proposed nonparametric estimator is denoted by $\hat{v}_c$, the parametric estimator based on a logistic regression model by $\hat{v}_c^l$ and the parametric estimator based on Dunning's scaled logit model by $\hat{v}_c^d$*

Model	p_v	c	v_c	$\hat{v}_c$			$\hat{v}_c^l$			$\hat{v}_c^d$		
				Bias	Width	Cov	Bias	Width	Cov	Bias	Width	Cov
Logit	0.01	0.001	1.31	0.00	0.30	0.93	0.00	0.23	0.94	−0.01	0.24	0.93
		0.005	0.82	0.00	0.20	0.95	0.00	0.19	0.95	0.00	0.19	0.96
		0.009	0.47	−0.02	0.38	0.96	−0.02	0.38	0.96	−0.02	0.38	0.96
	0.1	0.009	2.00	0.00	0.10	0.96	0.00	0.08	0.96	0.00	0.08	0.96
		0.01	1.98	0.00	0.10	0.96	0.00	0.08	0.96	0.00	0.08	0.96
		0.05	1.43	0.00	0.09	0.94	0.00	0.08	0.95	0.00	0.08	0.95
Probit	0.01	0.001	1.00	0.00	0.13	0.95	0.01	0.12	0.94	0.00	0.12	0.94
		0.005	0.73	0.00	0.14	0.95	−0.01	0.13	0.95	0.00	0.13	0.95
		0.009	0.47	−0.01	0.33	0.95	−0.01	0.33	0.95	−0.01	0.33	0.95
	0.1	0.009	1.82	0.00	0.06	0.95	−0.01	0.05	0.93	−0.01	0.05	0.93
		0.01	1.80	0.00	0.06	0.95	−0.01	0.05	0.93	−0.01	0.05	0.93
		0.05	1.41	0.00	0.07	0.96	0.00	0.07	0.96	0.00	0.07	0.96
Step	0.01	0.001	1.16	0.00	0.05	0.94	0.28	0.17	0.00	−0.01	0.04	0.88
		0.005	0.97	0.00	0.15	0.94	−0.10	0.20	0.42	−0.01	0.13	0.94
		0.009	0.61	−0.04	0.53	0.94	−0.15	0.41	0.73	−0.05	0.51	0.95
	0.1	0.009	2.39	0.00	0.03	0.94	0.70	0.14	0.00	−0.01	0.03	0.76
		0.01	2.38	0.00	0.04	0.94	0.64	0.14	0.00	−0.01	0.03	0.77
	0.1	0.05	1.95	0.00	0.11	0.94	−0.17	0.14	0.00	−0.01	0.09	0.92
Scaled	0.01	0.001	1.43	0.00	0.29	0.95	0.10	0.25	0.69	−0.01	0.25	0.95
		0.005	0.94	0.00	0.21	0.94	−0.03	0.22	0.91	0.00	0.20	0.94
		0.009	0.57	−0.04	0.47	0.94	−0.08	0.42	0.91	−0.04	0.47	0.94
	0.1	0.009	2.48	0.00	0.11	0.94	0.48	0.15	0.00	0.00	0.09	0.95
		0.01	2.45	0.00	0.10	0.95	0.45	0.14	0.00	0.00	0.09	0.95
		0.05	1.84	0.00	0.11	0.95	−0.10	0.13	0.13	0.00	0.11	0.95

$\hat{v}_0^s$ were calculated and 95% CIs for v_0 were computed using the bootstrap percentile method with $\hat{v}_0$ and $B = 2000$, and the methods from Cooke and Siber et al. Table 2 presents the simulation results.

In general, the difference in bias between $\tilde{v}_0$ and $\hat{v}_0$ was negligible. The bootstrap CIs corresponding to $\hat{v}_0$ never reached the nominal coverage level 0.95. Cooke's (Cooke (1979)) method generally performed well in terms of bias and CI coverage, except for the probit model simulations when $p_v = 0.01$, in which case the estimator was biased and the CI coverage was lower then 0.95. In general, the estimator $\hat{v}_0^s$ performed poorly when the assumed models were misspecified (logit and probit models), with $\hat{v}_0^s$ having substantial bias and the associated CIs failing to achieve the nominal coverage level.

TABLE 2

Results of simulation study described in Section 2.2 based on 1000 *simulated datasets each with sample size* 25,000 *and a* 1:1 *vaccine:control allocation where Bias denotes empirical bias, Width denotes average CI width and Cov denotes empirical CI coverage rate. The proposed nonparametric estimator is denoted by $\tilde{v}_0$, Siber's estimator by $\hat{v}_0^s$ and the maximum immune response value for the vaccinated and diseased sample by $\hat{v}_0$*

Model	(p_c, p_v)	v_0	$\hat{v}_0$ Bias	Width	Cov	$\tilde{v}_0$ Bias	Width	Cov	(L_0^c, U_0^c) Bias	Width	Cov	$\hat{v}_0^s$ Bias	Width	Cov	(L_0^s, U_0^s)
Logit	(0.10, 0.01)	2.00	−0.18	0.25	0.00	−0.13	4.33	0.93				−0.80	0.24	0.00	
	(0.30, 0.10)	2.00	0.00	0.01	0.65	0.00	0.08	0.94				0.45	0.19	0.00	
Probit	(0.10, 0.01)	2.00	−0.73	0.14	0.00	−0.70	2.80	0.78				−0.77	0.25	0.00	
	(0.30, 0.10)	2.00	0.00	0.01	0.34	0.00	0.16	0.96				0.46	0.19	0.00	
Step	(0.10, 0.01)	1.20	0.00	0.01	0.75	0.00	0.12	0.95				0.00	0.25	0.98	
	(0.30, 0.10)	2.47	0.00	0.00	0.84	0.00	0.04	0.95				0.00	0.20	0.96	

4. Extensions.

4.1. *Background.*   The methods in Section 2 assume each individual's immune response and disease status are observed. However, in some studies this assumption may not hold. Two examples are studies with nested subsampling of S by design and with right censoring of Y. Common subsampling designs are case-control and case-cohort designs. Here *case* refers to individuals who develop disease during the study, and *control* refers to individuals who are disease-free at the end of study follow-up. There are multiple types of case-control studies (Vandenbroucke and Pearce (2012)), and in this section we consider extensions of the methods above for the "cumulative case-control study" type, for which S is measured for all cases and for a random sample of controls (Cornfield (1951), Vandenbroucke and Pearce (2012)).

The extended methods also apply for a case-cohort sampling design, for which S is measured for all individuals randomly sampled into a subcohort at enrollment and also for all cases not sampled into the subcohort (Prentice (1986)). Our methods do not apply for case-control studies with risk-set sampling of S, which sample one or more control participants at each point in time when an event $Y = 1$ occurs (Vandenbroucke and Pearce (2012)). In rare event studies such as our motivating application of preventive vaccine efficacy trials, risk-set sampling provides negligible benefits compared to cumulative case-control sampling, and cumulative case-control or case-cohort studies are the norm for vaccine efficacy trials.

In addition, in some studies the time to disease is right censored for some individuals, for example, due to loss to follow-up. Methods for this scenario need to account for missing data from right censoring as well as to the subsampling design.

4.2. *Methods.* Consider the vaccinated sample from the trial detailed in Section 2.1. We describe the methods for cumulative case-control sampling. Among individuals with $Y = 0$, proportion p_0 are randomly selected, by Bernoulli or without replacement sampling, to form the control group, denoted G_0. Similarly, a proportion p_1 of individuals with $Y = 1$ are randomly selected to form the case group, denoted G_1. Denote the union $G_0 \cup G_1$ by G. For individual i, let $M_i = 1$ if $i \in G$ and $M_i = 0$ otherwise for $i = 1, \ldots, n$. For now, suppose Y is observed for all n individuals but S is observed only for individuals in G.

As in Section 2, point estimation and a corresponding $(1 - \alpha)$ CI for v_c are desired. The scenario where $c = 0$ is not considered in this section. To adjust for the missing immune response data, individuals are weighted by their probability of being selected into G conditional on Y using inverse probability weighting (Horvitz and Thompson (1952)). For individual i, let $p_0 = \Pr(i \in G | Y_i = 0)$ and let $p_1 = \Pr(i \in G | Y_i = 1)$. Denote the corresponding empirical proportions by $\hat{p}_0$ and $\hat{p}_1$ respectively. Then following Robins et al. (Robins, Rotnitzky and Zhao (1995)), the weight for individual i, denoted W_i, is defined as $W_i = 1/\hat{p}_0$ if $Y_i = 0$ and $W_i = 1/\hat{p}_1$ if $Y_i = 1$. The weighted nonparametric estimator $\hat{v}_c^w$ is defined as

$$\hat{v}_c^w = \min\left\{ S_j : \frac{\sum_{i=1}^n W_i I(M_i = 1) I(Y_i = 1, S_i \geq S_j)}{\sum_{i=1}^n W_i I(M_i = 1) I(S_i \geq S_j)} \leq c, \, j \in \{1, \ldots, n\} \right\}.$$

This nonparametric estimator is compared to parametric estimators of the form

$$(5) \quad \min\left\{ S_j : \frac{\sum_{i=1}^n W_i I(M_i = 1) I(S_i \geq S_j) \hat{p}(S_i)}{\sum_{i=1}^n W_i I(M_i = 1) I(S_i \geq S_j)} \leq c, \, j \in \{1, \ldots, n\} \right\},$$

where $p(v) = \Pr(Y = 1 | S = v)$ is estimated by a parametric regression model, denoted $\hat{p}(v)$. Estimators $\hat{v}_c^{wl}$ and $\hat{v}_c^{wd}$ are defined by (5) using the logit and scaled logit models respectively from Section 2.1 for $p(v)$. The regression parameters for $p(v)$ are estimated using weighted maximum likelihood with weight $W_i I(M_i = 1)$ for individual i. Except in the special case where the sampling probabilities are the same for the disease and nondiseased group, not weighting will in general lead to biased estimates of $p(v)$. The bootstrap percentile method is used to construct a $(1 - \alpha)$ CI for v_c.

The estimator $\hat{v}_c^w$ and corresponding bootstrap CI can also be used for a cumulative case-cohort design or for a two-phase sampling design, which is defined as an extension of the cumulative case-control design that uses stratified random sampling of cases and controls within strata of a discrete covariate V with K levels that is measured in all n participants (Breslow et al. (2009)). For the latter extension, the estimated inverse probability weights W_i are now defined using $\hat{p}_{yk}$ that are estimates of $\Pr(i \in G | Y_i = y, V_i = k)$, which again may be obtained as empirical fractions.

For the methods to provide valid point and confidence interval estimation of the full population-level parameters $p(v)$ and v_c (i.e., as if S had been measured in

all n participants), the bootstrap is used with sampling type (Bernoulli or without-replacement) the same as that used for the observed data set. Validity of the methods relies on consistent estimation of (p_0, p_1) or the p_{yk}, which is readily achieved in studies where the investigator designs which individuals to sample from (barring a substantial amount of happenstance missing data).

Now suppose that the n individuals are monitored starting at time 0 for incident disease. Let T_i denote the time until disease or *survival time*. Suppose survival times are subject to right censoring and let C_i denote the censoring time for individual i. Assume T_i is independent of C_i, and that we observe $\min\{T_i, C_i\}$ and $\delta_i = I(T_i \leq C_i)$, where $\delta_i = 0$ indicates censoring. Consider some time $t_0 > 0$ and suppose the outcome of interest is $Y_i = I(T_i \leq t_0)$, that is, disease by t_0. Let $V(t_0|S \geq v) = 1 - \Pr(Y = 1|S \geq v)$ denote the survival function at time t_0, conditional on $S \geq v$. In this scenario, $\hat{v}_c^w$ needs to be adjusted since it ignores the right censoring. The proposed nonparametric estimator $\hat{v}_c^{wk}$ is defined as

$$\hat{v}_c^{wk} = \min\{S_j : 1 - \hat{V}_w(t_0|S \geq S_j) \leq c, \ j \in \{1, \ldots, n\}\},$$

where $\hat{V}_w(t|S \geq v)$ is the weighted Kaplan–Meier estimator (Pepe and Fleming (1989)) of $V(t|S \geq v)$ with weight $W_i I(M_i = 1, S_i \geq v)$ for individual i. The bootstrap percentile method is used to construct a $(1 - \alpha)$ CI for v_c.

Note in the preceding paragraph that Y_i is defined for a particular time point t_0. Thus the corresponding immune response threshold v_c, that is, the target parameter of inference, is also (implicitly) defined in terms of t_0. More generally, we can consider the target parameter of interest to be a function of time, say $v_c(t)$, and the approach described above can be used to draw inference about $v_c(t)$ by evaluating $\hat{v}_{ck}$ over a grid of values of t.

4.3. *Simulation studies.*   The methods in Section 4.2 were examined in simulation studies with and without right censoring. Data were simulated for n independent vaccinated individuals. Immune response S was generated as in Section 3.1 and 1000 datasets were generated each with sample size $n = 12{,}500$.

First, the simulations without right censoring are discussed. For these simulations, Y was generated as in Section 3.1. Define $n_0 = \sum_{i=1}^{n} I(Y_i = 0)$. For each dataset, cumulative case-control sampling was done, where subsets G_0 and G_1 were selected without replacement of size $\lfloor p_0 n_0 \rfloor$ from those with $Y = 0$ and $\lfloor p_1(n - n_0) \rfloor$ from those with $Y = 1$ respectively, where $p_0 = 0.2$ and $p_1 = 1$ and $\lfloor \cdot \rfloor$ denotes the floor function. Then, $\hat{v}_c^w$, $\hat{v}_c^{wl}$ and $\hat{v}_c^{wd}$ were calculated as well as 95% bootstrap CIs with $B = 2000$. Simulation results are presented in Table 3.

The results were similar to those in Table 1. In general, the empirical bias of $\hat{v}_c^w$ was negligible and the corresponding CIs had coverage rates approximately equal to the nominal level 0.95. In contrast, the parametric estimator $\hat{v}_c^{wl}$ performed poorly when the assumed model was misspecified, with $\hat{v}_c^{wl}$ having substantial

TABLE 3

Results of first simulation study described in Section 4.3 based on 1000 simulated datasets each with sample size 12,500 where Bias denotes empirical bias, Width denotes average CI width and Cov denotes empirical CI coverage rate. The proposed weighted nonparametric estimator is denoted by $\hat{v}_c^w$, the weighted parametric estimator using logistic regression by $\hat{v}_c^{wl}$ and the weighted parametric estimator using Dunning's scaled logit model by $\hat{v}_c^{wd}$

Model	p_v	c	v_c	$\hat{v}_c^w$ Bias	Width	Cov	$\hat{v}_c^{wl}$ Bias	Width	Cov	$\hat{v}_c^{wd}$ Bias	Width	Cov
Logit	0.01	0.001	1.31	0.00	0.29	0.95	0.00	0.26	0.94	−0.02	0.26	0.94
		0.005	0.82	0.00	0.20	0.96	0.00	0.20	0.96	0.00	0.20	0.95
		0.009	0.47	−0.02	0.38	0.95	−0.02	0.39	0.95	−0.01	0.39	0.95
	0.10	0.009	2.00	0.00	0.10	0.96	0.00	0.10	0.95	0.00	0.10	0.94
		0.010	1.98	0.00	0.10	0.96	0.00	0.10	0.96	0.00	0.10	0.96
		0.050	1.43	0.00	0.09	0.97	0.00	0.10	0.96	0.00	0.09	0.96
Probit	0.01	0.001	1.00	0.00	0.14	0.96	0.00	0.15	0.95	−0.01	0.13	0.95
		0.005	0.73	0.00	0.14	0.96	−0.01	0.14	0.95	0.00	0.13	0.96
		0.009	0.47	−0.02	0.33	0.96	−0.02	0.33	0.96	−0.01	0.34	0.96
	0.10	0.009	1.82	0.00	0.06	0.95	−0.01	0.06	0.92	−0.01	0.06	0.92
		0.010	1.80	0.00	0.06	0.95	−0.01	0.06	0.93	−0.01	0.06	0.93
		0.050	1.41	0.00	0.08	0.98	0.00	0.08	0.97	0.00	0.08	0.97
Step	0.01	0.001	1.16	0.00	0.05	0.94	0.28	0.25	0.00	0.00	0.05	0.91
		0.005	0.97	0.00	0.15	0.95	−0.10	0.21	0.51	−0.01	0.16	0.95
		0.009	0.61	−0.04	0.50	0.96	−0.15	0.39	0.83	−0.05	0.49	0.96
	0.10	0.009	2.39	0.00	0.03	0.94	0.69	0.24	0.00	0.00	0.04	0.86
		0.01	2.38	0.00	0.04	0.97	0.64	0.23	0.00	0.00	0.04	0.88
		0.05	1.95	0.00	0.12	0.97	−0.16	0.16	0.01	0.00	0.12	0.96
Scaled	0.01	0.001	1.43	0.01	0.29	0.94	0.10	0.30	0.76	0.00	0.26	0.96
		0.005	0.94	0.00	0.21	0.95	−0.03	0.23	0.93	0.00	0.22	0.97
		0.009	0.57	−0.03	0.44	0.95	−0.07	0.41	0.94	−0.03	0.44	0.95
	0.10	0.009	2.48	0.00	0.11	0.95	0.48	0.24	0.00	0.00	0.10	0.95
		0.010	2.45	0.00	0.11	0.95	0.44	0.23	0.00	0.00	0.10	0.95
		0.050	1.84	0.00	0.12	0.97	−0.10	0.16	0.22	0.00	0.12	0.97

bias and the associated CIs failing to achieve the nominal coverage level. The parametric estimator $\hat{v}_c^{wd}$ had low bias for all models, however the associated CIs had coverage rates below 0.95 in some scenarios when the assumed model was misspecified. The widths of CIs corresponding to $\hat{v}_c^w$ were nearly identical to the widths of CIs corresponding to $\hat{v}_c$ in Table 1, indicating a minimal loss in efficiency due to case-control sampling.

Next, results from simulations with right censoring are presented. Survival times were generated using accelerated failure time (AFT) models, adapted from Chan et al. (Chan et al. (2002)). In particular, T was generated by $\log(T) = \beta_0 + \beta_1 S + \epsilon$ where S was generated from a gamma distribution with mean and variance 4 and ϵ was generated from either a standard Normal distribution or logistic distribution

TABLE 4
Results of second simulation study described in Section 4.3 *based on* 1000 *simulated datasets each with sample size* 12,500 *where Bias denotes empirical bias, Width denotes average CI width and Cov denotes empirical CI coverage rate. The proposed weighted nonparametric estimator is denoted by* $\hat{v}_c^{wk}$ *and the weighted nonparametric estimator which ignores censoring by* $\hat{v}_c^w$

Model	p_v	c	v_c	$\hat{v}_c^{wk}$			$\hat{v}_c^w$		
				Bias	Width	Cov	Bias	Width	Cov
Log-Normal	0.01	0.001	1.43	−0.01	0.40	0.93	−0.19	0.32	0.37
		0.005	0.91	−0.01	0.34	0.94	−0.47	0.46	0.01
		0.009	0.50	−0.03	0.49	0.96	−0.30	0.13	0.01
	0.10	0.009	2.19	0.00	0.17	0.95	−0.20	0.13	0.00
		0.01	2.16	0.00	0.17	0.96	−0.21	0.13	0.00
		0.05	1.49	0.00	0.15	0.96	−0.44	0.20	0.00
Logistic	0.01	0.001	2.41	0.01	1.10	0.93	−0.28	0.76	0.65
		0.005	1.32	0.01	0.58	0.96	−0.46	0.70	0.13
		0.009	0.64	−0.02	0.71	0.96	−0.44	0.16	0.02
	0.10	0.009	2.88	0.00	0.35	0.94	−0.27	0.27	0.06
		0.01	2.82	0.00	0.33	0.95	−0.27	0.26	0.03
		0.05	1.71	0.00	0.21	0.96	−0.44	0.23	0.00

(location parameter 0 and scale parameter 1). To reflect an increasing survival time for higher immune response values, β_1 was set to 2 and β_0 was determined after fixing p_v. Disease outcome Y was defined by $Y = I(T \leq t_0)$ with $t_0 = 40$. Censoring time C was generated as $\min\{E, t_0\}$ where E followed an exponential distribution with mean t_0. Sets G_0 and G_1 were generated using $p_0 = 0.2$ and $p_1 = 1$ respectively. Given c, for each dataset, $\hat{v}_c^{wk}$ and $\hat{v}_c^w$ were computed along with 95% bootstrap CIs with $B = 2000$.

Table 4 presents the simulation results for each AFT model. In general, the empirical bias of $\hat{v}_c^{wk}$ was negligible and the corresponding CIs had coverage rates approximately equal to the nominal level 0.95. In contrast, $\hat{v}_c^w$ generally had high bias and the corresponding CIs failed to achieve the nominal coverage level, which is expected since $\hat{v}_c^w$ does not account for the right censoring.

The simulation study above was repeated with sample size 6500. Results (not shown) were similar to those in Table 4, except the CIs had larger average widths and coverage of the CIs corresponding to $\hat{v}_c^{wk}$ was slightly below the nominal level (89% for the log-normal model, and 87% for the logistic model).

5. Application.

5.1. *Data.* In this section, the methods described above are used to estimate neutralizing antibody thresholds for virologically confirmed dengue risk levels using data from two recent dengue vaccine trials. With an estimated 390 million

dengue infections per year and anti-viral treatment for dengue unavailable, continued research on dengue vaccination is important (World Health Organization (2017a)). CYD14 and CYD15 were Phase 3 placebo-controlled studies aimed at evaluating the efficacy and safety of the dengue vaccine CYD-TDV in children. Following these two studies, CYD-TDV became the first dengue vaccine to be licensed, beginning with Mexico in 2015 (World Health Organization (2017b)).

CYD14 was conducted in five Asian-Pacific countries (Indonesia, Malaysia, Philippines, Thailand and Vietnam), with participants 2–14 years old (Capeding et al. (2014)). CYD15 was conducted in five Latin American countries (Columbia, Brazil, Mexico, Puerto Rico and Honduras), with participants 9–16 years old (Villar et al. (2015)). The protocols for these two studies were harmonized. For both studies, participants were randomized in a 2:1 ratio to CYD-TDV vaccine or placebo. Participants received doses at the start of the study (month 0), at month 6 and at month 12 with follow-up visits at months 13 and 25. The primary endpoint was time until virologically confirmed dengue disease from any of four serotypes between months 13 and 25 (referred to simply as "dengue" below). The primary objective was to estimate vaccine efficacy against dengue occurring between months 13 and 25 in the per-protocol population (those with dengue occurring between months 0 and 13 were excluded). Both studies used a case-cohort sampling design where a subset of participants was randomly selected for neutralizing antibody titers measurement before the first dose and at months 7, 13 and 25. All cases (dengue) also had antibody titers measured. The magnitude of neutralizing antibodies to each of the four serotypes represented in the CYD-TDV vaccine were measured as 50% neutralization titer at the $\log_{10}$ scale; thus participants in the case-cohort subset had four $PRNT_{50}$ values.

Pooled analysis of participants 9–16 years old from both trials indicated that antibody response was positively associated with VE and that vaccination was associated with decreased dengue risk (Moodie et al. (2018)) (here and in Section 5.2, *antibody response* refers to the average of the four serotype-specific antibody titers from the month 13 measurements in the case-cohort subset). In a pooled analysis of participants 9–16 years old from both trials, the estimated VE was 65.6% (95% CI of 60.7%, 69.9%) using a Cox regression model with vaccine status and study participation as covariates (Hadinegoro et al. (2015)).

5.2. *Results*.   Motivated by these results, risk thresholds when vaccinated were estimated using the CYD14, CYD15 and pooled CYD14 and CYD15 (CYD14/15) data. The participants included in the statistical analysis were those who did not develop dengue and were not right censored prior to month 13 and had titers measured, that is, the case-cohort subset. These participants were analyzed according to the treatment group to which they were randomized, irrespective of per-protocol criteria. For the pooled analysis, only participants between 9 and 16 years old were included because 9 and older is the indication for the licensed vaccine. When analyzing the CYD14 and CYD15 data individually, there were no age restrictions.

The analysis included 6454 participants in CYD14 with 104 cases, 13,376 participants in CYD15 with 183 cases and 16,623 participants in CYD14/CYD15 with 216 cases. For CYD14, CYD15, and CYD14/CYD15, 1145, 1458, and 2088 participants respectively had measured antibody data.

Let Y denote dengue status, with $Y = 1$ denoting presence of dengue between months 13 and 25. Figures 2(a)–2(c) display estimates of $\Pr(Y = 1|Z = 1, S \geq v)$ for each observed antibody response v in the vaccine arm for CYD14, CYD15 and CYD14/CYD15 respectively. Estimates were computed using the weighted Kaplan–Meier estimator in Section 4.2 to account for case-cohort sampling and right censored outcomes, with weights equal to the inverse of the sample proportions having an observed antibody response in the strata of individuals with and without dengue. These proportions were 1 among individuals with dengue for each dataset and were 0.164, 0.100, 0.114 for CYD14, CYD15 and CYD14/15 respectively among individuals without dengue. All three figures indicate decreasing dengue risk as antibody response increases. Figures 2(d)–2(f) show estimates and 95% CI for risk thresholds over a range of risk levels for CYD14, CYD15 and CY14/15. Due to the case-cohort sampling and right censoring, point estimates and CIs were constructed using $\hat{v}_c^{wk}$ as described in Section 4.2 with $B = 5000$.

6. Discussion. In this paper, nonparametric methods for inference on risk thresholds were presented and examined through simulation studies. Extensions of these methods for studies with subsampling of the immune response, such as case-cohort and cumulative case-control and two-phase studies, and with right censoring of the outcome, were also considered. Simulation studies were presented demonstrating that the proposed nonparametric estimators tended to have minimal bias and the corresponding bootstrap CIs typically cover at approximately the nominal level. In contrast, estimators and CIs based on parametric models or Siber et al. (Siber et al. (2007)) tended to not perform well empirically when the additional underlying assumptions did not hold.

The nonparametric methods were used for inference on neutralizing antibody risk thresholds for specified risk levels of virologically confirmed dengue for the CYD-TDV vaccine. In new populations adopting the CYD-TDV vaccine, these risk thresholds may be helpful for policymakers and public health officials in estimating the risk of dengue based solely on the distribution of antibody concentrations. Similarly, these risk thresholds may be helpful in indirect evaluation of new candidate dengue vaccines on the basis of antibody titer data alone by providing initial estimates of disease risk by titer thresholds; one application is informing Go/No-Go decisions for down-selecting vaccines for advancement to efficacy trials.

In some settings it may be of interest to draw inference about immune response risk thresholds within certain subgroups. In this case the proposed method can be used by restricting the analysis to participants with particular (baseline) covariate
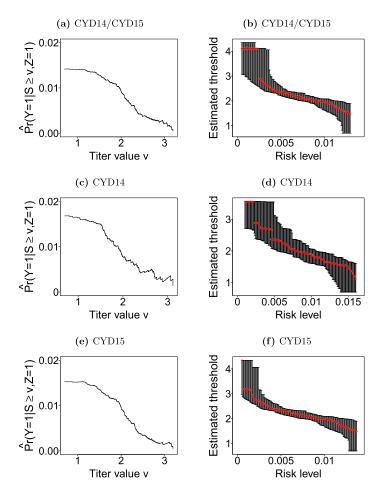
FIG. 2. *Panels* (a), (c) *and* (e) *show estimated disease risk conditional on* $S \geq v$ *by* $\log_{10}$ *titer at month* 13 *when vaccinated for* 9–16 *years olds in CYD*14 *and CYD*15 *pooled* (*CYD*14/15), *CYD*14 *all ages*, *and CYD*15 *all ages*, *respectively. These estimates were calculated for each observed antibody response* v *in the data and are denoted by* $\hat{\Pr}(Y = 1 | Z = 1, S \geq v)$. *Panels* (b), (d), *and* (f) *show estimates* (*red points*) *and* 95% *CIs for risk thresholds over a range of risk levels for CYD*14/15, *CYD*14 *and CYD*15 *respectively.*

values, resulting in estimates of covariate-specific risk thresholds. This would allow indirect evaluation of vaccine efficacy for subpopulations of interest or for new populations whose covariate distributions differ from those studied in the original efficacy trial(s). In particular, if a covariate that is correlated with both the immune response and disease has a different distribution in the two settings (e.g., if the new population is older), then the covariate-conditional thresholds would be needed for predicting vaccine efficacy in immune response threshold subgroups in the new population.

There are several possible avenues of future research related to the methods considered here. For example, unlike the frequentist approach used in this paper, Bayesian methods could be developed for inference about thresholds of risk. Advantages of a Bayesian approach would include (i) not relying on asymptotic approximations and (ii) allowing prior information on thresholds to be incorporated into the inference. On the other hand, a disadvantage of the Bayesian approach is that the results will depend on the (subjective) choice of the prior.

Another possible area of further research would be to extend the methods developed here to allow for diagnostic uncertainty. Throughout this paper it was assumed that the true disease status is observed for each individual. However, in some settings the presence of disease may be measured with error. Extensions of the proposed methods to allow for diagnostic uncertainty will naturally depend on the particular setting under consideration, for example, whether the sensitivity and specificity of the diagnostic instrument are known, whether a validation subsample is available where the true disease status is known, and so forth. Existing measurement error methods, such as multiple imputation (Cole, Chu and Greenland (2006)), could potentially be combined with the nonparametric approach developed here.

Future research could entail extending the nonparametric method presented here to the setting where it is assumed the risk threshold is a monotonic function of immune response. If this monotonicity assumption is correct, nonparametric methods that leverage this assumption would be expected to outperform the approach in this paper. That said, the nonparametric methods considered here have the appealing property of not relying on such a monotonicity assumption, and thus should provide valid inference regardless of whether such an assumption holds.

Future research could also examine developing confidence bands for a set of risk levels. The CIs presented here provide only pointwise coverage. Lastly, methods for testing for the existence of risk thresholds is a possible area of future research. There has been some development on methods for this type of inference for a risk level of zero (Chen et al. (2013)), however for general risk thresholds no such methods exist.

R code corresponding to this manuscript is available at http://github.com/kmdono02/Risk_Threshold (Donovan, Hudgens and Gilbert (2019)). The repository includes R functions to compute the proposed nonparametric estimators, as well as code to recreate some of the simulation studies in this paper.

SUPPLEMENTARY MATERIAL

R code for computation of nonparametric estimators for thresholds of risk (DOI: 10.1214/18-AOAS1237SUPP; .zip). We provide a repository which includes R functions to compute the proposed nonparametric estimators, as well as code to recreate some of the simulation studies in this paper.

REFERENCES

ANDREWS, N., BORROW, R. and MILLER, E. (2003). Validation of serological correlate of protection for meningococcal c conjugate vaccine by using efficacy estimates from postlicensure surveillance in England. *Clin. Diagn. Lab. Immunol.* **10** 780–786.

BLACK, S., NICOLAY, U., VESIKARI, T., KNUF, M., DEL GIUDICE, G., DELLA CIOPPA, G., TSAI, T., CLEMENS, R. and RAPPUOLI, R. (2011). Hemagglutination inhibition antibody titers as a correlate of protection for inactivated influenza vaccines in children. *Pediatr. Infect. Dis. J.* **30** 1081–1085.

BORROW, R., BALMER, P. and MILLER, E. (2005). Meningococcal surrogates of protection—-serum bactericidal antibody activity. *Vaccine* **23** 2222–2227.

BRESLOW, N., LUMLEY, T., BALLANTYNE, C., CHAMBLESS, L. and KULICH, M. (2009). Using the whole cohort in the analysis of case-cohort data. *Am. J. Epidemiol.* **169** 1398–1405.

CAPEDING, M. R., TRAN, N. H., HADINEGORO, S. R. S., ISMAIL, H. I. H. M., CHOTPITAYA-SUNONDH, T. and CHUA, M. N. (2014). Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: A phase 3, randomised, observer-masked, placebo-controlled trial. *Lancet* **384** 1358–1365.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2014). What would happen if we stopped vaccinations? Available at https://www.cdc.gov/vaccines/vac-gen/whatifstop.htm. Accessed June, 2018.

CHAN, I. S. F., LI, S., MATTHEWS, H., CHAN, C., VESSEY, R., SADOFF, J. and HEYSE, J. (2002). Use of statistical models for evaluating antibody response as a correlate of protection against varicella. *Stat. Med.* **21** 3411–3430.

CHEN, X., BAILLEUX, F., DESAI, K., QIN, L. and DUNNING, A. J. (2013). A threshold method for immunological correlates of protection. *BMC Med. Res. Methodol.* **13** 29.

COLE, S. R., CHU, H. and GREENLAND, S. (2006). Multiple-imputation for measurement-error correction. *Int. J. Epidemiol.* **35** 1074–1081.

COOKE, P. (1979). Statistical inference for bounds of random variables. *Biometrika* **66** 367–374. MR0548207

CORNFIELD, J. (1951). A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J. Natl. Cancer Inst.* **11** 1269–1275.

DONOVAN, K. M., HUDGENS, M. G. and GILBERT, P. B. (2019). Supplement to "Nonparametric inference for immune response thresholds of risk in vaccine studies." DOI:10.1214/18-AOAS1237SUPP.

DUNNING, A. J. (2006). A model for immunological correlates of protection. *Stat. Med.* **25** 1485–1497. MR2227290

DYE, C. (2014). After 2015: Infectious diseases in a new era of health and development. *Phil. Trans. R. Soc. B* **369** 20130426.

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903

HADINEGORO, S. R., ARREDONDO-GARCÍA, J. L., CAPEDING, M. R., DESEDA, C., CHOTPITAYASUNONDH, T. and DIETZE, R. (2015). Efficacy and long-term safety of a dengue vaccine in regions of endemic disease. *N. Engl. J. Med.* **373** 1195–1206.

HAYNES, B. F., GILBERT, P. B., McELRATH, M. J., ZOLLA-PAZNER, S., TOMARAS, G. D., ALAM, S. M., EVANS, D. T., MONTEFIORI, D. C., KARNASUTA, C. et al. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. *N. Engl. J. Med.* **366** 1275–1286.

HORVITZ, D. G. and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685. MR0053460

JÓDAR, L., BUTLER, J., CARLONE, G., DAGAN, R., GOLDBLATT, D. and KÄYHTY, H. (2003). Serological criteria for evaluation and licensure of new pneumococcal conjugate vaccine formulations for use in infants. *Vaccine* **21** 3265–3272.

JOKINEN, J. T. and ÅHMAN, H. (2004). Concentration of antipneumococcal antibodies as a serological correlate of protection: An application to acute otitis media. *J. Infect. Dis.* **190** 545–550.

MOODIE, Z., JURASKA, M., HUANG, Y., ZHUANG, Y., FONG, Y., CARPP, L. N., SELF, S. G., CHAMBONNEAU, L., SMALL, R. et al. (2018). Neutralizing antibody correlates analysis of tetravalent dengue vaccine efficacy trials in Asia and Latin America. *J. Infect. Dis.* **217** 742–753.

PEPE, M. S. and FLEMING, T. R. (1989). Weighted Kaplan–Meier statistics: A class of distance tests for censored survival data. *Biometrics* **45** 497–507. MR1010515

PRENTICE, R. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73** 1–11.

PRENTICE, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Stat. Med.* **8** 431–440.

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *J. Amer. Statist. Assoc.* **90** 106–121. MR1325118

SIBER, G. R., CHANG, I., BAKER, S., FERNSTEN, P., O'BRIEN, K. L. and SANTOSHAM, M. (2007). Estimating the protective concentration of anti-pneumococcal capsular polysaccharide antibodies. *Vaccine* **25** 3816–3826.

STORSAETER, J., HALLANDER, H. O., GUSTAFSSON, L. and OLIN, P. (1998). Levels of anti-pertussis antibodies related to protection after household exposure to *Bordetella pertussis*. *Vaccine* **16** 1907–1916.

VANDENBROUCKE, J. P. and PEARCE, N. (2012). Case–control studies: Basic concepts. *Int. J. Epidemiol.* **41** 1480–1489.

VILLAR, L., DAYAN, G. H., ARREDONDO-GARCÍA, J. L., RIVERA, D. M., CUNHA, R. and DESEDA, C. (2015). Efficacy of a tetravalent dengue vaccine in children in Latin America. *N. Engl. J. Med.* **372** 113–123.

WHITE, C. J., KUTER, B. J., NGAI, A., HILDEBRAND, C. S., ISGANITIS, K. L. and PATTERSON, C. M. (1992). Modified cases of chickenpox after varicella vaccination: Correlation of protection with antibody response. *Pediatr. Infect. Dis. J.* **11** 19–22.

WORLD HEALTH ORGANIZATION (2017a). Dengue and severe dengue. Available at http://www.who.int/mediacentre/factsheets/fs117/en/. Accessed June, 2018.

WORLD HEALTH ORGANIZATION (2017b). Dengue vaccine research. Available at http://www.who.int/immunization/research/development/dengue_vaccines/en/. Accessed June, 2018.

K. M. DONOVAN
M. G. HUDGENS
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA, 27599
USA
E-MAIL: kmdono02@ad.unc.edu
        mhudgens@email.unc.edu

P. B. GILBERT
FRED HUTCHINSON CANCER RESEARCH CENTER
SEATTLE, WASHINGTON, 98109
USA
E-MAIL: pgilbert@scharp.org