Accounting for subject-level heterogeneity in sieve analysis of vaccine efficacy

Jason Shao

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science

University of Washington

2017

Reading Committee:

Paul T. Edlefsen, Ph.D., Chair

Peter Gilbert, Ph.D.

Program Authorized to Offer Degree:
Department of Biostatistics

University of Washington

Abstract

Accounting for subject-level heterogeneity
in sieve analysis of vaccine efficacy

Jason Shao

Chair of the Supervisory Committee:
Paul T. Edlefsen, Ph.D.
Biostatistics

In randomized trials of preventative vaccines, sieve analysis tests whether vaccine efficacy differs by a characteristic of the disease endpoint. These methods often assume a leaky model, in which treatment proportionally reduces the hazard of each disease type homogeneously in all subjects. Significant biases can occur in estimation and testing when this assumption does not hold. To allow for unobserved heterogeneity in participant response to vaccination, we propose a frailty mixture model for sieve analysis which incorporates unobserved, subject-level, random effects into a competing risks survival analysis framework. We show that parameters in the model can be straightforwardly and reliably estimated using standard numeric optimization methods. In simulation studies, our approach performs favorably to existing methods in cases where the leaky vaccine assumption is not appropriate. Finally, we implement our method on existing clinical trial datasets and discuss the implications of our findings for the design and interpretation of vaccine efficacy studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Vaccines are a vital part of global strategies for the prevention and control of infectious diseases. However, many diseases for which there are vaccines currently under development are caused by pathogens displaying significant genotypic and phenotypic diversity, which can pose issues in the design of broadly protective vaccines. Many modern vaccine candidates are multivalent, containing genetic material from many strains of the pathogen. Ideally, a vaccine should induce an immune response to all the strains it contains, and the response should be broad enough to protect against disease caused by strains that are similar but not identical. However, this is not always the case, and an important aspect of the development of safe and efficacious vaccines is *sieve analysis*, or the study of whether the efficacy of a vaccine is dependent on measurable characteristics of the infecting pathogen.

This thesis focuses on sieve analysis in the context of Phase IIb and III clinical trials, in which uninfected volunteers are randomly assigned to receive either a candidate vaccine or a placebo treatment, and subsequently monitored for disease. Depending on the vaccine and pathogen, disease may be defined in various ways using clinical definitions, various biological markers, or a combination of both. In addition to recording the time from randomization to the disease endpoint among subjects who experience the disease, we may obtain other information about the pathogen itself. For example, in trials of vaccine against the dengue virus, it is possible to determine the serotype of the virus (DENV-1 thru DENV-4) through immunological testing on plasma samples taken from infected subjects. The sieve analysis would then aim to answer the question of whether vaccine efficacy differs by serotype. Genetic sequencing can also be performed on the viral samples — this information can be used to determine, for instance, whether efficacy (serotype-specific or overall) depends on genetically

defined regions, or epitopes.

1.1 Sieve analysis methods

Suppose, in the general setting of a vaccine trial, that the variation in the pathogen of interest circulating in the geographic region can be summarized exhaustively by K distinct strains, and that each disease endpoint is caused by exactly one strain. We refer to the strain-specific vaccine efficacy as $VE_j = 1 - RR_j$ for $j = 1, ..., K$. Here, RR_j is the relative risk, among vaccine versus placebo recipients, of clinically significant disease following exposure to strain j. In this setting, sieve analysis is generally concerned with the pairwise odds ratios $OR_{j,k} = RR_k/RR_j$, which quantify the relative vaccine efficacy between two distinct strains j and k. The sieve null hypothesis is $H_0 : OR_{j,k} = 1 \ \forall \ j, k$, and a rejection of H_0 at an appropriate level is evidence of a "sieve effect."

Gilbert et al. (1998) described sieve analysis in the setting of Phase I and IIa clinical trials against human immunodeficiency virus type 1 (HIV-1), in which the only data available are the frequency of viral strains observed among infected subjects in the placebo and vaccine groups. Table 1.1 shows an example of count data that might arise from such a trial.

	Infecting strains						
	1	2	$\cdots$	K	Total infected	Not infected	Sample size
Placebo	$n_{p,1}$	$n_{p,2}$	$\cdots$	$n_{p,K}$	$n_p = \sum_j n_{p,j}$	$n_{p,0} = N_p - \sum_j n_{p,j}$	N_p
Vaccine	$n_{v,1}$	$n_{v,2}$	$\cdots$	$n_{v,K}$	$n_v = \sum_j n_{p,j}$	$n_{v,0} = N_v - \sum_j n_{v,j}$	N_v

Table 1.1: Possible endpoint counts from a vaccine efficacy trial

The observed frequency distribution $(n_{p,1}, ..., n_{p,K})$ arises due to a combination of differences in prevalence and natural barriers to HIV-1 acquisition. Conceptually, the sieve refers to the vaccine-induced barrier to acquisition, which may result in a different frequency distribution $(n_{v,1}, ..., n_{v,K})$ among infected subjects in the vaccine group. The fact that these

distributions are only observed for the set of subjects who become infected is what distinguishes sieve analysis from typical statistical methods used on clinical trial data.

Gilbert et al. (1998) were able to show that, although direct estimation of the RR_j is not possible in this setting, the estimators $\widehat{OR}_{j,k} = \frac{n_{v,k}/n_{p,k}}{n_{v,j}/n_{p,j}}$ can be used as the basis for valid inference on a sieve effect. They proposed using a multinomial logistic regression (MLR) model to perform inference on the $OR_{j,k}$ and allow for covariate adjustment. Suppose that for each subject we observe a binary indicator of vaccination x, a vector of other covariates ξ, and the infecting strain type J (for convenience in notation, denote $J = 0$ if the subject was not infected). The MLR model specifies:

$$\Pr(J = j | J \neq 0, \boldsymbol{x}) = \frac{\exp(\beta_{0,j} + \beta_j x + \gamma_j \xi)}{\sum_{i=1}^{K} \exp(\beta_{0,j} + \beta_j x + \gamma_j \xi)} \tag{1.1}$$

If the condition $\beta_{0,j} = 0$ is imposed for all j (in order to ensure identifiability), the β_j and γ_j are interpretable as strain-specific log odds ratios for disease caused by strain j vs. strain 1. If no covariates besides vaccination status are present, the maximum likelihood estimates of β_j are exactly equal to $\widehat{OR}_{j,1}$, as defined previously.

In Phase IIb and III trials, time-to-event data is typically available for each subject. Gilbert (2000) expanded the sieve analysis methodology to incorporate time-to-event data using Cox proportional hazards models, in which disease caused by different strains are treated as competing risks using the framework described by Prentice et al. (1971). In a clinical trial with left-censoring, the data are independent, identically distributed observations (T, δ_i, J, x, ξ), where T is the length of follow-up, δ is a censoring indicator ($\delta = 1$ if the participant was infected during the trial) and J, x, and ξ are the same as before. We define the cause-specific hazard λ_j as follows:

$$\lambda_j(t|\boldsymbol{x}) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t, J = j | T \geq t, \boldsymbol{x})}{\Delta t} \tag{1.2}$$

Cox's proportional hazards model specifies $\lambda_j(t|\boldsymbol{x}) = \exp(\beta_j x + \xi)\lambda_{0,j}(t)$ for arbitrary baseline hazard functions $\lambda_{0,j}(t)$ for $j = 1, ..., K$. The maximum partial likelihood estimator

can be used to estimate the β_j treating disease caused by other strains as censoring, and a data duplication method can easily be implemented to test for differences among analogous terms in β_j and β_k (Lunn and McNeil, 1995). Here, the β_j have direct interpretation as log relative risk of disease caused by strain j between vaccinated and unvaccinated subjects.

1.2 Subject-level heterogeneity in vaccine protection

The sieve analysis methods discussed above operate under a number of assumptions that are necessary to maintain validity of inferences made on strain-specific odds ratios. Chief among these is the assumption that, for each strain j, the effect of the vaccine is to reduce the probability of disease-causing infection from a single exposure by the same multiplicative factor for all vaccine recipients. In other words, the vaccine mechanism is "leaky" to each strain, using terminology from Halloran et al. (1992).

Violations to the leaky vaccine assumption can occur if vaccine protection is heterogeneous. Suppose that a vaccine has the effect of reducing the probability of infection by strain j to zero for a randomly determined proportion $(1-\mu_j)$ of vaccine recipients, whereas for the remaining proportion μ_j, the probability is multiplied by some coefficient, θ_j. If $\theta_j = 1$, then the vaccine could be described as "all-or-none" against strain j. If $\mu_j = 1$, the vaccine falls under the "leaky" category. In all these cases, we might interpret the quantity $RR_j^* = \mu_j \cdot \theta_j$ to be a measure of the average vaccine-induced reduction in the risk of infection by strain j. It has been demonstrated extensively, outside the realm of sieve analysis, that violations of the leaky vaccine assumption can result in inaccurate estimation of RR_j^* if a leaky mechanism is assumed (Smith et al., 1984; Halloran et al., 1992).

In the context of sieve analysis, Gilbert (2001) showed that estimation of the odds ratios $OR_{j,k}^* = RR_j^*/RR_k^*$ using the MLR or Cox proportional hazards approaches could also result in significant bias in certain contexts. Specifically, when the proportion of unvaccinated subjects infected in a trial was above 70%, and when there was a three-fold difference in the prevalence of strain j and k in the geographic setting of the trial, a bias of up to 20% occurred in the estimation of $OR_{j,k}^*$. This corresponded with an inflated type-1 error rate in

simulations under which there was no true sieve effect.

In the typical conditions of a Phase IIb or III clinical trial, the actual rate of disease endpoints among unvaccinated subjects is likely to be under 10%, a context in which the biases were less extreme. Thus, Gilbert (2001) suggested that the typical sieve analysis approaches were generally adequate. However, it is still of interest to develop procedures for sieve analysis that are robust to violations of the leaky vaccine assumption, for several reasons. First, it is plausible that such methods might be needed in the case of a clinical trial with higher event rates than normal. Secondly, the development of sieve methods outside the leaky vaccine framework will help to gain insight into the interpretability of sieve effect estimators when the vaccine mechanism is unknown, which is true in most cases.

This thesis proposes a model-based maximum likelihood estimation procedure which accounts for subject-level heterogeneity in vaccine protection. The model is similar to one developed by Longini and Halloran (1996) for a single disease type, in which the authors treat the heterogeneity in susceptibility separately for vaccinated and unvaccinated subjects using frailty models, and examine the performance of the model on estimating the summary vaccine efficacy $1 - RR^*$. We demonstrate that the model expands easily to sieve analysis, and that a method of maximum likelihood estimation can be used to determine the strain-specific summary vaccine efficacy $1 - RR_j^*$. In Chapter 2 we will lay out the notation of the full proposed model and derive the likelihood functions. We will also present a few simplified special cases of the model, including one in which the vaccine mechanism is assumed to be "all-or-none," for the purposes of comparison. In Chapter 3, we will use simulation studies to show that maximum likelihood estimation of the odds ratios $OR_{j,k}^* = RR_k^*/RR_j^*$ under our model is possible with greatly reduced bias compared to the proportional hazards model. In Chapter 4, we will apply the proposed method to a recent clinical trial of CYD 14, a tetravalent dengue vaccine, in order to determine whether vaccine efficacy differed between the four main serotypes of dengue. Finally, in Chapter 5, we will discuss the limitations and advantages of the proposed model compared to proportional hazards models.

Chapter 2

METHODS

Consider, as before, a study population susceptible to disease caused by K distinct strains, or types. Let x be a binary indicator of vaccination. We define the type-specific hazard of disease by the jth strain as follows for $j = 1, ..., K$:

$$\lambda_j(t|x, z_j) = z_j \cdot (\theta_j)^x \cdot \lambda_{0,j}(t) \tag{2.1}$$

where $\lambda_{0,j}(t)$ is the type-specific baseline hazard function, θ_j is a non-negative multiplier that represents the effect of the vaccine on disease caused by strain j, and $z_j, ..., z_K$ are *unobserved* non-negative multipliers which account for subject-level heterogeneity in susceptibility. Note that θ_j is exponentiated by x here, so that the hazard for an unvaccinated subject is $z_j\lambda_{0,j}(t)$.

Let the z_j be unobserved instances of random variables Z_j, distributed as follows for $j = 1, ...K$:

$$\begin{gathered} Z_j = U_j \cdot W_j \\ \text{where } U_j \sim \text{Bernoulli}((\mu_j)^x) \text{ and } W_j \sim f_{x,j}(\cdot) \end{gathered} \tag{2.2}$$

such that a proportion $(1-\mu_j)$ of vaccinated subjects are immune to disease caused by strain j. Again, note the exponentiation by x, such that no unvaccinated subjects are immune to strain j. For the remainder of the subjects, the distribution $f_{x,j}(\cdot)$ is left unspecified, but in order for the model to be identifiable, it must be positive and have mean 1 for $x = 0$ or 1.

Here, it is important to distinguish our model from *frailty* models, in which the distribution of Z_j is identical for all subjects. In a frailty model, we would simply have $Z_j \sim f_j(\cdot)$, and inferential methods for parametric $f_j(\cdot)$ have been well-studied in the literature (Aalen, 1988; Hougaard, 1986). In our case, however, we note that the distribution of both components of Z_j depend on vaccination status. This is more similar to the frailty mixture model

proposed by Longini and Halloran (1996) in the case of a single disease type. Note that, because of the exponentiation of μ_j by x in (2.2), we are assuming that 100% of unvaccinated subjects are susceptible to strain j - this is not the case in the model presented by Longini and Halloran (1996). The implications of this assumption are discussed in Section 5.2.

For computational ease, Longini and Halloran (1996) suggested using a Gamma distribution $W_j \sim \text{Gamma}(\psi_{x,j}^{-1}, \psi_{x,j}^{-1})$, such that

$$f_{x,j}(w) = \frac{\psi_{x,j}^{-1}{}^{\psi_{x,j}^{-1}}}{\Gamma(\psi_{x,j}^{-1})} w^{(\psi_{x,j}^{-1}-1)} \exp(-\psi_{x,j}^{-1} w) \tag{2.3}$$

which has $\text{E}[W_j] = 1$ and $\text{Var}[W_j] = \psi_{x,j}$. A special case of this model is the "degenerate mixture model" (DMM), in which the $\psi_{x,j}$ approach 0, and $f_{x,j}(\cdot)$ becomes a point mass at 1. In this case we have simply $Z_j \sim \text{Bernoulli}((\mu_j)^x)$.

We note here that the frailty mixture model is very similar to cure rate models, which are typically used to discern the effect of a treatment on both the short-term risk of disease and the probability of lifelong immunity in long-term clinical trials more typically found in cancer research (Maller and Zhou, 1996). In short-term studies on a single disease type, however, Longini and Halloran (1996) note that the goal is simply to provide valid inference on the summary relative risk $RR^* = (1-\mu)\theta$. We extend that principle to the strains $j = 1, ..., K$. Note that these quantities is unaffected by the relationship between vaccination and the frailty distributions $f_{x,j}(\cdot)$, as long as the distributions are defined with mean 1, as previously stated.

We will now demonstrate how maximum likelihood estimation can be used to estimate the terms in the frailty mixture model for j strains.

2.1 Derivation of the likelihood functions

Let $(t_i, \delta_i, J_i, x_i)$ be the observed data for the ith subject in a study of N subjects. We use the notation of equation 2.1, adding the subscript i whenever applicable to indicate values for the ith subject. Additionally, t_i is the follow-up time, δ_i is an indicator of whether disease occurred during the follow-up time and J_i is the cause of failure (for notational convenience

we define $J_i = 0$ when $\delta_i = 0$). The general form of the likelihood function for time-to-event data under K competing risks is

$$L_N(\cdot|t_i, \delta_i, J_i, x_i) = \prod_{i=1}^{N} \left[\lambda_{J_i}(t_i|x_i, z_{i,j})^{\delta_i} S(t_i|x_i, z_{i,j})\right] \tag{2.4}$$

where $S(t_i|x_i, z_{i,j}) = \exp(-\int_0^{t_i} \sum_{j=1}^{K} \lambda_j(u|x_i, z_{i,j})du)$ is the overall survivor function. The above likelihood contains the $z_{i,j}$, which are unobserved, so it will be necessary to integrate the likelihood over the random distributions specified in (2.2).

First, we note that (2.4) factors into K cause-specific components (Prentice et al., 1971):

$$L_N(\cdot|t_i, \delta_i, J_i, x_i) = \prod_{j=1}^{K} \prod_{i=1}^{N} \lambda_j(t_i|x_i, z_{i,j})^{\delta_i \cdot I(J_i=j)} \exp\big(-\Lambda_j(t_i|x_i, z_{i,j})\big) \tag{2.5}$$

where $\Lambda_j(t_i|x_i, z_{i,j}) = \int_0^{t_i} \lambda_j(u|x_i, z_{i,j})du$ is the cause-specific cumulative hazard function. If we define the type-specific censoring indicators $\delta_{i,j} \equiv \delta_i \cdot I(J_i = j)$, we can write the likelihood for the jth cause on the ith subject as

$$L_{i,j}(\cdot|t_i, \delta_{i,j}, x_i, z_{i,j}) \equiv \lambda_j(t_i|x_i, z_{i,j})^{\delta_{i,j}} \exp\big(-\Lambda_j(t_i|x_i, z_{i,j})\big) \tag{2.6}$$

Taking advantage of this factorization, and assuming that the subjects are independent, we can work on integrating (2.6) over the distribution for the random variable Z_j, and then generalize to the full likelihood in (2.4). Let $\Lambda_{0,j}(t_i) = \int_{u=0}^{t_i} \lambda_{0,j}(u)du$. Note that because of (2.1) we have

$$\begin{aligned} \Lambda_j(t_i|x_i, z_{i,j}) &= \int_0^{t_i} z_{i,j}(\theta_j)^{x_i} \lambda_{0,j}(u)du \\ &= z_{i,j}(\theta_j)^{x_i} \int_0^{t_i} \lambda_{0,j}(u)du \\ &= z_{i,j}(\theta_j)^{x_i} \Lambda_{0,j}(t_i) \end{aligned} \tag{2.7}$$

so (2.6) becomes the following:

$$L_{i,j}(\cdot|t_i, \delta_{i,j}, x_i, z_{i,j}) = \big[z_{i,j}(\theta_j)^{x_i} \lambda_{0,j}(t_i)\big]^{\delta_{i,j}} \exp\big(-z_{i,j}(\theta_j)^{x_i} \Lambda_{0,j}(t_i)\big) \tag{2.8}$$

At this point, in order to obtain an expression of the likelihood free of the unobserved terms $z_{i,j}$, it is necessary to use the expected value of the random variable Z_j defined above in (2.2). We will do this separately for subjects with $\delta_{i,j} = 0$ or 1.

First, for a subject with $\delta_{i,j} = 0$, we have

$$\begin{aligned} L_{i,j}(\cdot|t_i, \delta_{i,j} = 0, x_i) =& \mathrm{E}_{Z_j}\left[\exp\left(-Z_j(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)\right] \\ =& \Pr(U_j = 0) * \mathrm{E}_{W_j}\left[\exp\left(-0 \cdot W_j(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)\right] + \\ & \Pr(U_j = 1) * \mathrm{E}_{W_j}\left[\exp\left(-1 \cdot W_j(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)\right] \\ =& (1 - (\mu_j)^{x_i}) + (\mu_j)^{x_i}\int_{w=0}^{\infty}\exp\left(-w(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)df_{x,j}(w) \end{aligned} \tag{2.9}$$

For a subject with $\delta_{i,j} = 1$, we have

$$\begin{aligned} L_{i,j}(\cdot|t_i, \delta_{i,j} = 1, x_i) =& \mathrm{E}_{Z_j}\left[Z_j(\theta_j)^{x_i}\lambda_{0,j}(t_i)\exp\left(-Z_j(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)\right] \\ =& \Pr(U_j = 0) * \mathrm{E}_{W_j}\left[0 \cdot W_j(\theta_j)^{x_i}\lambda_{0,j}(t_i)\exp\left(-0 \cdot W_j(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)\right] + \\ & \Pr(U_j = 1) * \mathrm{E}_{W_j}\left[1 \cdot W_j(\theta_j)^{x_i}\lambda_{0,j}(t_i)\exp\left(-1 \cdot W_j(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)\right] \\ =& (\mu_j)^{x_i}\int_{w=0}^{\infty} w(\theta_j)^{x_i}\lambda_{0,j}(t_i)\exp\left(-w(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)df_{x,j}(w) \end{aligned} \tag{2.10}$$

If we use the Gamma$(\psi_{x,j}^{-1}, \psi_{x,j}^{-1})$ definition of $f_{x,j}(\cdot)$ in (2.3), it is possible to obtain closed-form solutions to the integrals in (2.9) and (2.10), which results in the following combined expression:

$$\begin{aligned} L_{i,j}(\cdot|t_i, \delta_{i,j}, x_i) =& (1 - (\mu_j)^{x_i})(1 - \delta_{i,j}) + \\ & (\mu_j)^{x_i}\left((\theta_j)^{x_i}\lambda_{0,j}(t_i)\right)^{\delta_{i,j}}\left(1 + \psi_{x_i,j}(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\right)^{-(\psi_{x_i,j}^{-1} + \delta_{i,j})} \end{aligned} \tag{2.11}$$

In the case of the DMM, in which the $f_{x,j}(\cdot)$ are point masses at 1, we have the simpler expression:

$$L_{i,j}(\cdot|t_i, \delta_{i,j}, x_i) = (1 - (\mu_j)^{x_i})(1 - \delta_{i,j}) + (\mu_j)^{x_i}\big((\theta_j)^{x_i}\lambda_{0,j}(t_i)\big)^{\delta_{i,j}} \exp\big(-(\theta_j)^{x_i}\Lambda_{0,j}(t_i)\big) \tag{2.12}$$

In either case, if the $\lambda_{0,j}(t)$ are specified parametrically, we can use existing numerical methods to maximize the full likelihood. Going forward, we will be using the DMM for examples, and we make the additional simplifying assumption that the strain-specific baseline hazards are constant, so that $\lambda_{0,j}(t) = \lambda_{0,j}$ and $\Lambda_{0,j}(t) = t\lambda_{0,j}$, but note that it is possible to parameterize the baseline hazards in other ways (such as with a piecewise constant function, a Weibull distribution on the failure times, etc).

Taking the product of (2.12) over subjects i and strains j, the expression of the likelihood that we will be maximizing is :

$$L_N(\theta_j, \mu_j, \lambda_{0,j}|\cdot) = \prod_{i=1}^{N}\prod_{j=1}^{K}\left((1 - (\mu_j)^{x_i})(1 - \delta_{i,j}) + (\mu_j)^{x_i}\big((\theta_j)^{x_i}\lambda_{0,j}\big)^{\delta_{i,j}} \exp\big(-(\theta_j)^{x_i}t_i\lambda_{0,j}\big)\right) \tag{2.13}$$

2.2 Constrained optimization via transformation of the parameter space

Here, we consider the problem of constrained parameter spaces in the maximum likelihood estimation of the parameters of (2.12). The parameter μ_j is necessarily constrained to the interval $(0, 1]$. The strain-specific baseline hazard rate $\lambda_{0,j}$ is constrained to $(0, \infty)$. θ_j is also constrained to $(0, \infty)$, but if the assumption of "no harm" is made, it only takes values in the interval $(0, 1]$. This assumption implies that there is no segment of the population in which the vaccine causes harm, i.e. increases the risk of disease following exposure to strain j. This assumption helps to avoid pitfalls in the likelihood fitting procedure for frailty mixture models, and we discuss its validity in Section 5.3.

We used a quasi-Newton `BFGS` algorithm for optimization in the `optim()` function built into R v3.2.2 to provide maximum likelihood estimation of the parameters in (2.12). We chose to transform the constrained parameters to range in $(-\infty, \infty)$, instead of using barrier algorithms during numerical optimization, because we found that it was generally more reliable. For parameters restricted to $(0, \infty)$, we used the logarithmic transformation (as is typical in regression models for time-to-event data). For parameters restricted to $(0, 1]$, we chose to use the probit transformation. Thus, for the general model (without the no-harm assumption) we define the following:

$$\alpha_j = \Phi^{-1}(\mu_j) \qquad \beta_j = \log(\theta_j) \qquad \gamma_j = \log(\lambda_{0,j}) \tag{2.14}$$

In the no-harm model, we have instead

$$\alpha_j = \Phi^{-1}(\mu_j) \qquad \beta_j = \Phi^{-1}(\theta_j) \qquad \gamma_j = \log(\lambda_{0,j}) \tag{2.15}$$

2.3 Standard error estimation for inferences on the sieve odds ratios

A sieve effect, or differential VE, is defined as $OR_{j,k} \neq 1$, or $\log(OR)_{j,k} \neq 0$ for some $j, k \in \{1, ..., K\}$. We perform pairwise hypothesis testing by calculating the estimator for

$$\log(OR)_{j,k} = \log\left(\frac{\mu_j \theta_j}{\mu_k \theta_k}\right) \tag{2.16}$$

and performing a Wald test. The covariance matrix for the test is estimated by applying the delta method to the inverted negative Hessian matrix from numerical optimization of the likelihood parameters.

Let $\Theta = (\alpha_1, \beta_1, \gamma_1, ...\alpha_K, \beta_K, \gamma_K)$ be the vector of all parameters in the full likelihood (2.13), rewritten using the transformations in (2.15):

$$\begin{aligned} L_N(\Theta) = \prod_{i=1}^{N} \prod_{j=1}^{K} \Big(& (1 - \Phi(\alpha_j)^{x_i})(1 - \delta_{i,j}) + \\ & \Phi(\alpha_j)^{x_i} \big(\Phi(\beta_j)^{x_i} \exp(\gamma_j)\big)^{\delta_{i,j}} \exp\big(- \Phi(\beta_j)^{x_i} t_i \exp(\gamma_j)\big) \Big) \end{aligned} \tag{2.17}$$

The Hessian is the negative expected value of the matrix of second derivatives of $\log(L_N)$ evaluated at the estimated parameter values:

$$\mathcal{H}(\hat{\Theta}) = -\mathrm{E}\big[\nabla\nabla \log(L_N(\Theta))\big|_{\Theta=\hat{\Theta}} \tag{2.18}$$

where $\hat{\Theta} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\gamma}_1, ...\hat{\alpha}_K, \hat{\beta}_K, \hat{\gamma}_K)$ are the values of Θ that maximize (2.17).

Because (2.17) is a product over $j = 1, ..., K$, it follows that $\mathcal{H}(\hat{\Theta})$ is a block diagonal matrix with blocks of size equal to the number of parameters for strain j (in this case, three). For convenience of notation, let $\hat{\Theta}_{j,k} = (\hat{\alpha}_j, \hat{\beta}_j, \hat{\gamma}_j, \hat{\alpha}_k, \hat{\beta}_k, \hat{\gamma}_k)$ and $\mathcal{H}_{j,k}(\hat{\Theta}_{j,k})$ be the corresponding blocks of entries from $\mathcal{H}(\hat{\Theta})$:

$$\mathcal{H}_{j,k}(\hat{\Theta}_{j,k}) = -\mathrm{E}\begin{bmatrix} \frac{\partial^2 L_N}{\partial \alpha_j^2} & \frac{\partial^2 L_N}{\partial \alpha_j \partial \beta_j} & \frac{\partial^2 L_N}{\partial \alpha_j \partial \gamma_j} & 0 & 0 & 0 \\ \frac{\partial^2 L_N}{\partial \alpha_j \partial \beta_j} & \frac{\partial^2 L_N}{\partial \beta_j^2} & \frac{\partial^2 L_N}{\partial \beta_j \partial \gamma_j} & 0 & 0 & 0 \\ \frac{\partial^2 L_N}{\partial \alpha_j \partial \gamma_j} & \frac{\partial^2 L_N}{\partial \beta_j \partial \gamma_j} & \frac{\partial^2 L_N}{\partial \gamma_j^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\partial^2 L_N}{\partial \alpha_k^2} & \frac{\partial^2 L_N}{\partial \alpha_k \partial \beta_k} & \frac{\partial^2 L_N}{\partial \alpha_k \partial \gamma_k} \\ 0 & 0 & 0 & \frac{\partial^2 L_N}{\partial \alpha_j \partial \beta_k} & \frac{\partial^2 L_N}{\partial \beta_k^2} & \frac{\partial^2 L_N}{\partial \beta_k \partial \gamma_k} \\ 0 & 0 & 0 & \frac{\partial^2 L_N}{\partial \alpha_k \partial \gamma_k} & \frac{\partial^2 L_N}{\partial \beta_k \partial \gamma_k} & \frac{\partial^2 L_N}{\partial \gamma_k^2} \end{bmatrix}_{\Theta_{j,k}=\hat{\Theta}_{j,k}} \tag{2.19}$$

We know that $\hat{\Sigma}_{j,k} = \mathcal{H}_{j,k}(\hat{\Theta}_{j,k})^{-1}$ is an estimator for the covariance matrix $\mathrm{Var}[\hat{\Theta}_{j,k}]$. The maximum likelihood estimator for (2.16) is $\widehat{\log(OR)}_{j,k} = \log(\Phi(\hat{\alpha}_j)) + \log(\Phi(\hat{\beta}_j)) - \big(\log(\Phi(\hat{\alpha}_k)) + \log(\Phi(\hat{\beta}_k))\big)$. It follows from the delta method that

$$\widehat{\mathrm{Var}}[\widehat{\log(OR)}_{j,k}] = \boldsymbol{g}^T \hat{\Sigma}_{j,k} \boldsymbol{g} \tag{2.20}$$

where $\boldsymbol{g}(\Theta_{j,k})$ is the gradient of the function $\boldsymbol{f}(\Theta_{j,k}) = \log(\Phi(\alpha_j)) + \log(\Phi(\beta_j)) - \big(\log(\Phi(\alpha_k)) + \log(\Phi(\beta_k))\big)$ evaluated at $\hat{\Theta}_{j,k}$:

$$\boldsymbol{g}(\Theta_{j,k}) = \nabla \boldsymbol{f}(\hat{\Theta}_{j,k}) = \begin{pmatrix} \partial \boldsymbol{f}/\partial\alpha_j \\ \partial \boldsymbol{f}/\partial\beta_j \\ \partial \boldsymbol{f}/\partial\gamma_j \\ \partial \boldsymbol{f}/\partial\alpha_k \\ \partial \boldsymbol{f}/\partial\beta_k \\ \partial \boldsymbol{f}/\partial\gamma_k \end{pmatrix}_{\Theta_{j,k}=\hat{\Theta}_{j,k}} = \begin{pmatrix} \phi(\hat{\alpha}_j)/\Phi(\hat{\alpha}_j) \\ \phi(\hat{\beta}_j)/\Phi(\hat{\beta}_j) \\ 0 \\ -\phi(\hat{\alpha}_k)/\Phi(\hat{\alpha}_k) \\ -\phi(\hat{\beta}_k)/\Phi(\hat{\beta}_k) \\ 0 \end{pmatrix} \tag{2.21}$$

Here, $\phi(\cdot)$ is the derivative of $\Phi(\cdot)$, and is equal to the probability distribution function of the standard normal distribution.

Chapter 3

SIMULATION STUDY

3.1 Models

We evaluated the performance of three models to fit data simulated from clinical trials in which there are two circulating strains ($K = 2$):

- DMM: The degenerate mixture model described by Longini and Halloran (1996), applied to events caused by multiple strains. We estimate $\log(OR_{1,2}) = \log(\frac{\mu_1\theta_1}{\mu_2\theta_2})$ by maximizing the likelihood shown in (2.13).

- rDMM: A special case of the DMM is the "all-or-none" model, in which vaccine protection is assumed to be "all-or-none" to each strain. We estimate $\log(OR_{1,2}) = \log(\mu_1/\mu_2)$ by maximizing the likelihood shown in (2.13), with $\theta_j \equiv 1$ for $j = 1, 2$.

- PH: The proportional hazards model, in which vaccine protection is assumed to be "leaky" to each strain. We estimate $\log(OR_{1,2}) = \log(\theta_1/\theta_2)$ by maximizing the partial likelihood of a Cox proportional hazards model with strains treated as competing risks (Gilbert, 2001; Lunn and McNeil, 1995).

3.2 Simulation parameters

Here we look at simulation parameters, including sample size, proportion of subjects infected, and vaccine mechanism. We simulated data under a model with exponential event times for two circulating strains ($K = 2$). Subjects were randomly assigned to the vaccine and placebo groups in a 1:1 ratio. For $j = 1, 2$, the parameter μ_j determined the expected proportion of vaccinated subjects being partially susceptible to type j events (at a rate multiplied by θ_j),

with the remaining proportion $(1 - \mu_j)$ being completely immune to disease from type j . Simulated protection was independent against type 1 and type 2 events.

To determine whether the DMM, rDMM, and PH methods correctly controlled type 1 error, we simulated data for which VE_j was fixed at 75% ($RR_j^* = 0.25$) against both disease types, with $j = 1$ and $j = 2$. We varied the vaccine mechanism against events with type $J = j$ as follows:

- All-or-none (A): $\mu_j = 0.25, \theta_j = 1.0$
- Mixed (M): $\mu_j = 0.50, \theta_j = 0.50$
- Leaky (L): $\mu_j = 1.0, \theta_j = 0.25$

In all simulations, the prevalence of disease with $j = 2$ was set to be three times that of disease with $j = 1$. The study was halted at a predetermined time t_{max}, at which a certain proportion p of placebo recipients were expected to have experienced the event. In the data set, all subjects had a single follow-up time which was equal to the time of the first event, if disease caused by either strain occurred before t_{max}. Otherwise, the subject was considered censored at t_{max}.

In the simulated datasets, t_{max} was held constant at 36 months, while the baseline hazard rates $\lambda_{0,j}$ were adjusted to achieve different levels of p and maintain a prevalence ratio of 3:1 between strains 1 and 2. For the default simulation, p was set at 0.5, with a total sample size of $N = 500$ (250 vaccinated, 250 unvaccinated). Thus, the expected number of events among placebo subjects was $\mathrm{E}[n_p] = 125$. For time-to-event data, the expected number of events is a clearer indication of the impact of sample size than the actual number of samples. Thus, to examine the effect of $\mathrm{E}[n_p]$ and p on the methods, we varied $\mathrm{E}[n_p]$ from 125 to 500 and 2000 while keeping p constant at 0.5. We also kept $\mathrm{E}[n_p]$ constant at 125 while changing p to 0.1 and 0.9.

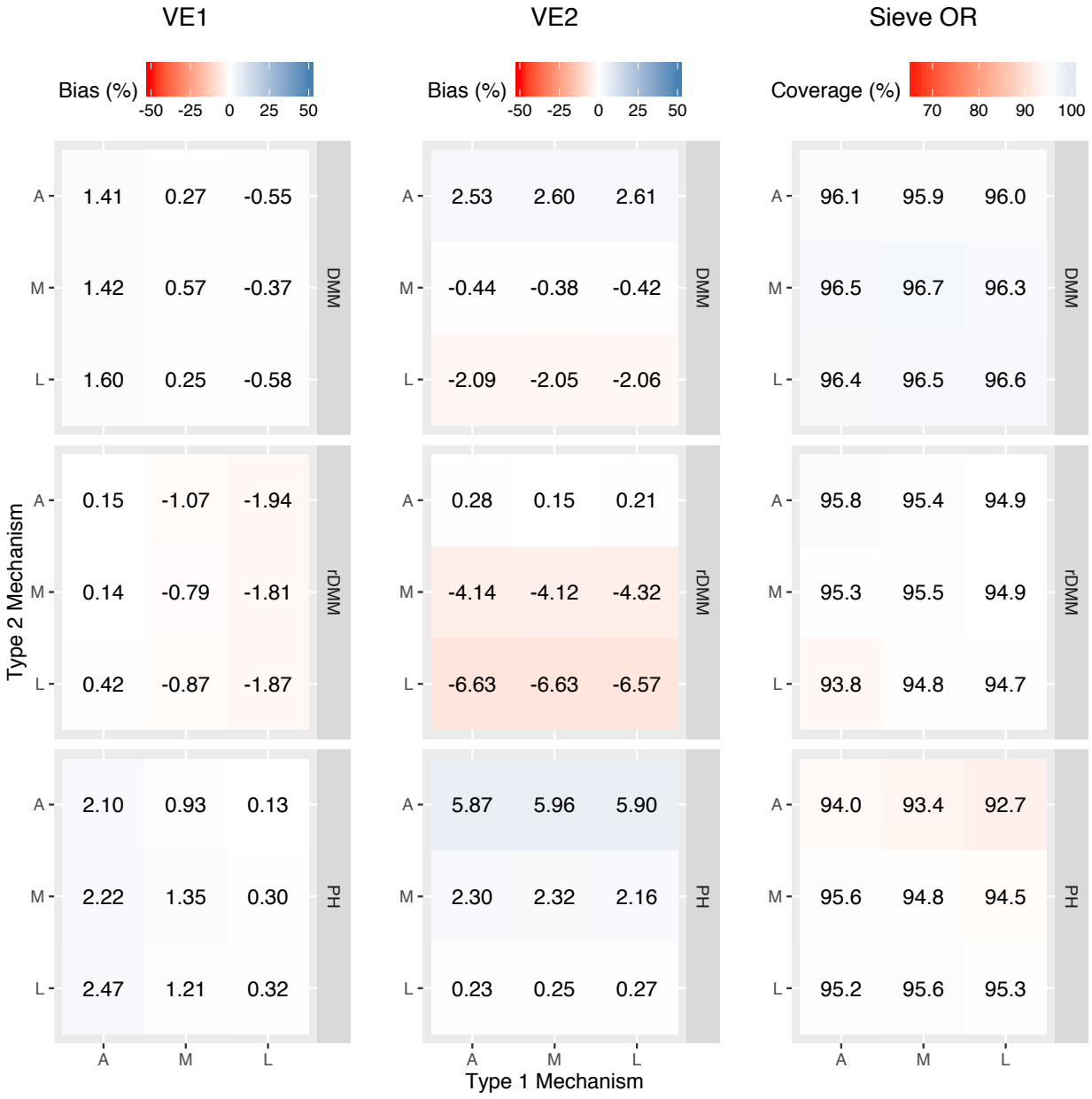
3.3 Bias and coverage

In Figure 3.1, we present nine panels of heatmaps displaying the bias and coverage of the three methods when applied to simulated datasets. Each heatmap cell represents values calculated from the results of applying the three methods - DMM (first row of panels), rDMM (second row of panels), and PH (third row of panels) - to 5,000 simulated datasets.

Within each heatmap, the true vaccine mechanism varies independently between all-or-none (A), mixed (M) and leaky (L) for disease type 1 on the horizontal axis, and type 2 on the vertical axis.

In the first and second columns of panels, the heatmap cells show the bias in estimation of VE_1 and VE_2 respectively, and are shaded according to the degree of bias. Negatively biased estimates are shaded in red, positively biased estimates in blue, with darker shading corresponding to higher magnitude of bias. Bias is defined as the percent deviation of the median estimators $\widehat{VE}_1^* = 1 - \hat{\mu}_1\hat{\theta}_1$ and $\widehat{VE}_2^* = 1 - \hat{\mu}_2\hat{\theta}_1$ from the true value of 0.75.

In the third column of panels, the heatmap cells show the 95% confidence interval coverage, defined as the percentage of 95% confidence intervals around the estimated $\log(OR_{1,2}) = \log(\frac{\hat{\mu}_1\hat{\theta}_1}{\hat{\mu}_2\hat{\theta}_1})$ (panel column 3) containing the true value of 0. Note that in this case, since the null hypothesis is true, coverage is equal to 1 minus the probability of falsely rejecting $H_0 : \log(OR_{1,2}) = 0$ at the $\alpha = 0.05$ level. Red shading represents undercoverage (anti-conservative) while blue shading represents overcoverage (conservative), with darker shading corresponding to higher deviation from the expected nominal coverage level of $\alpha = 0.05$.



$\mathrm{E}[\boldsymbol{n_p}] = \mathbf{125}$ and $\boldsymbol{p} = \mathbf{0.5}$

Figure 3.1: Heatmaps show estimated VE_1 (panel column 1), bias in estimated VE_2 (panel column 2), and coverage of 95% confidence intervals for $\log(OR_{1,2})$ (panel column 3) using three methods (panel rows). Within each heatmap, columns and rows indicate true simulated vaccine mechanism against type 1 and type 2 events, respectively. Heatmap cells represent median percentages across 5,000 simulated datasets.

Figure 3.1 indicates that, with a sample size of 500 (125 expected events in the placebo group), unbiased estimation of VE_j depends on correct specification of the vaccine mechanism in the PH model. As was previously shown by Gilbert (2001), using the PH model results in slight overestimation of vaccine efficacy when the true vaccine mechanism is all-or-none (A) or mixed (M). The overestimation is more noticeable for VE_2 (second column of panels), due to there being a greater number of strain 2 endpoints. We also observe the opposite pattern when the rDMM model is used, but the true vaccine mechanism is leaky (L) or mixed (M). In both cases, the inaccurate estimation of VE_1 and VE_2 results in under-coverage of the 95% confidence intervals, which leads to anti-conservative inference when a sieve effect is not present. The greatest amount of undercoverage occurred when the model was misspecified against the more prevalent strain 2, but was correctly specified against strain 1.

There is also some degree of bias in using the DMM (first row of panels) to estimate VE when the true mechanism is leaky (L) or all-or-none (A), even though the model should be correctly specified for all cells in the heatmaps. It's possible that this occurs due to the fact that the fact that the "leaky" and "all-or-none" models are on the boundary of the parameter space of the DMM. The implications of this are discussed in Section 5.4.

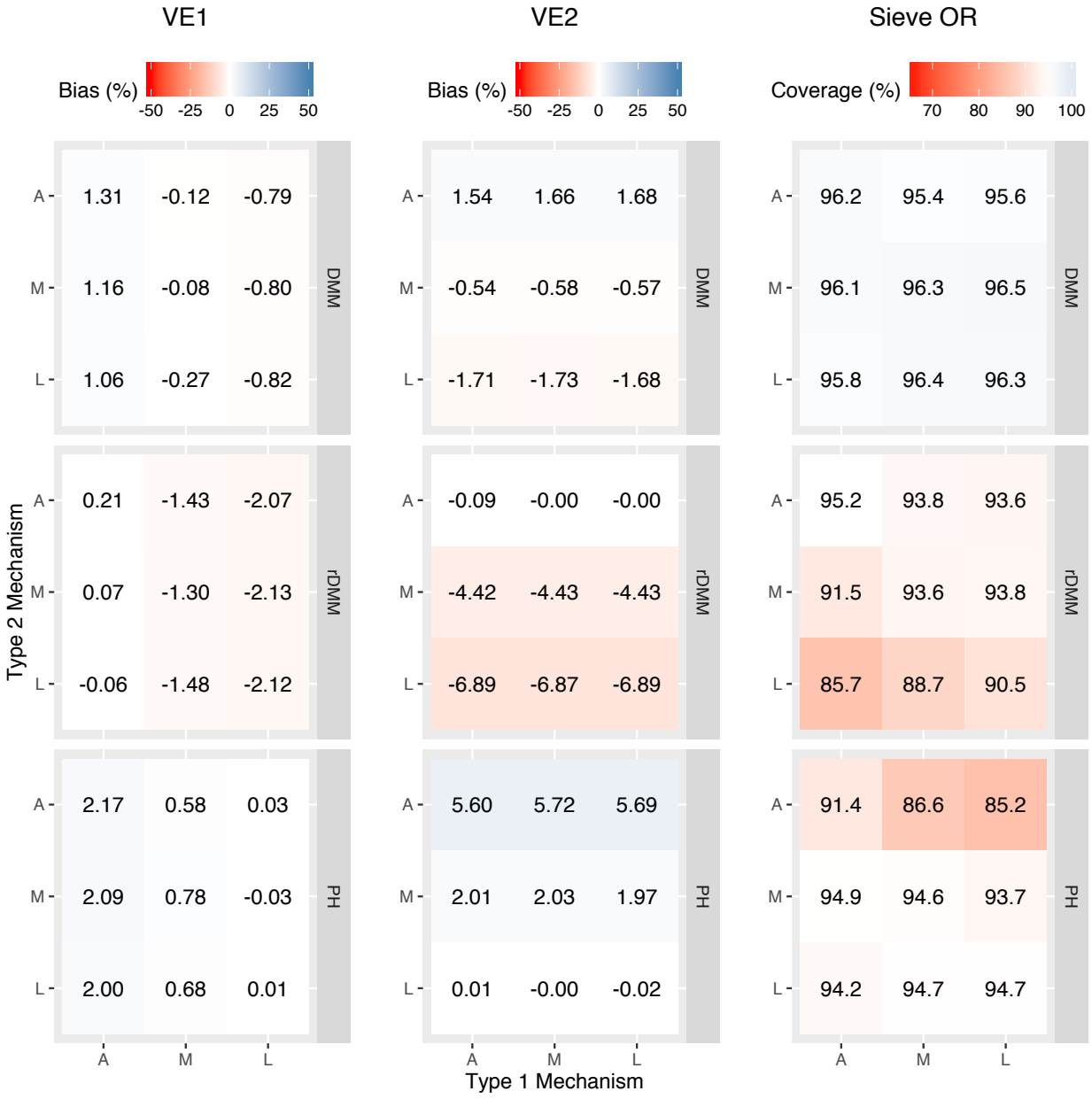
3.3.1 Effect of changing the sample size

Figures 3.2 and 3.3 show the effect of changing the sample size, and are set up similarly to Figure 3.1. The bias and undercoverage of misspecified models shown in Figure 3.1 are not very large for a small $\mathrm{E}[n_p]$ and moderate p. However, we see that when $\mathrm{E}[n_p]$ is increased to 500 or 2000, the bias in the misspecified models remains the same, but the coverage of the 95% confidence intervals decreases drastically (Figures 3.2 and 3.3). This indicates that the bias is not dependent on sample size, and larger sample size in the case of misspecified models produces more precise estimates of incorrect quantities. Additionally, there is some indication that increasing the sample size decreases the bias in estimation of VE when using the DMM, which was noted in the section above. Increasing the sample size also does not

induce any under-coverage in the confidence intervals using the DMM.

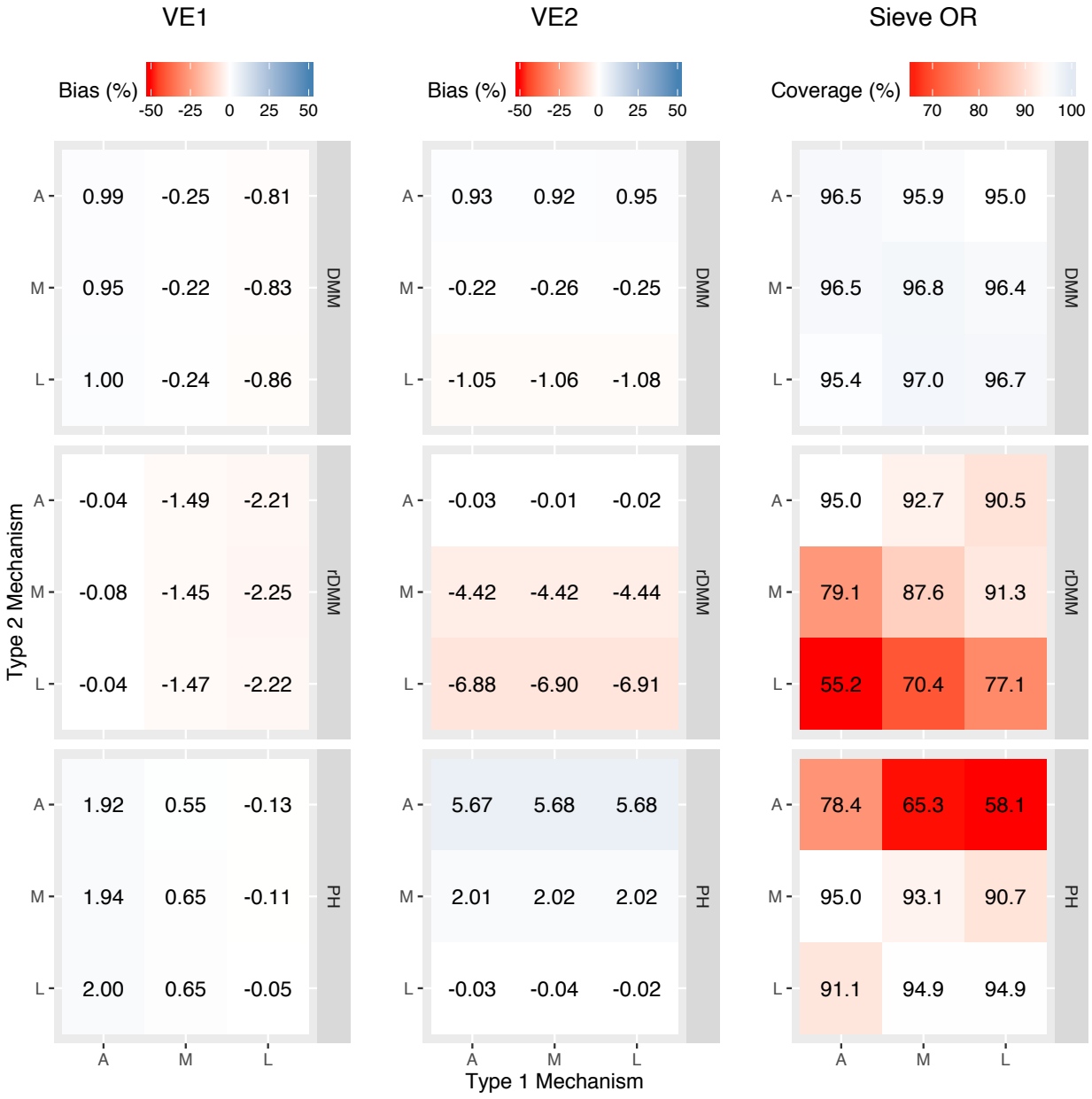
3.3.2 Effect of changing the expected event rate

Figures 3.4 and 3.5 show the effect of changing the expected event rate, and are also set up similarly to Figure 3.1. As shown in Figure 3.4, when the expected event rate for placebo subjects is increased to $p = 0.9$ (which is equivalent to $N = 278$ and $N_p = 139$ with $\mathrm{E}[n_p] = 125$), the bias when using rDMM and PH on misspecified vaccine mechanisms is much larger, and the 95% confidence interval coverage is correspondingly lower as well. Again, similarly to when sample size is increased, the full DMM is not affected by the increase in p, and performs very similarly throughout Figure 3.4. However, when p is lowered to 0.1 (Figure 3.5), we see that the bias and undercoverage when using rDMM and PH on misspecified vaccine mechanisms is negligible. This mirrors the findings in Gilbert (2001), and suggests that ignoring heterogeneity in vaccine protection may be acceptable in clinical trials where the event rate is expected to be low to moderate. The implications of this observation are discussed in Section 5.1.



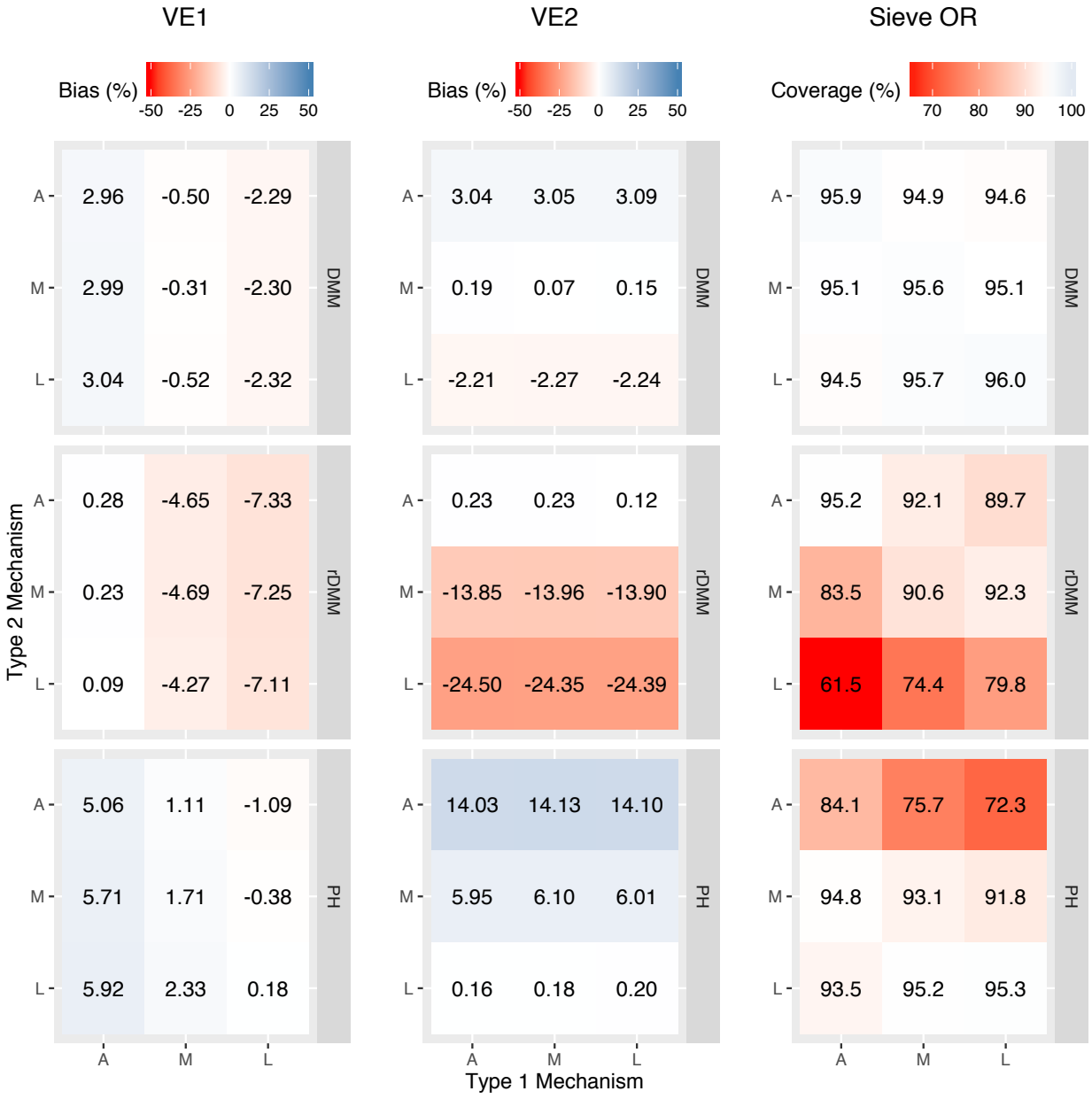
$\mathrm{E}[\boldsymbol{n_p}] = \mathbf{500}$ and $\boldsymbol{p} = \mathbf{0.5}$

Figure 3.2: Heatmaps show estimated VE_1 (panel column 1), bias in estimated VE_2 (panel column 2), and coverage of 95% confidence intervals for $\log(OR_{1,2})$ (panel column 3) using three methods (panel rows). Within each heatmap, columns and rows indicate true simulated vaccine mechanism against type 1 and type 2 events, respectively. Heatmap cells represent median percentages across 5,000 simulated datasets.



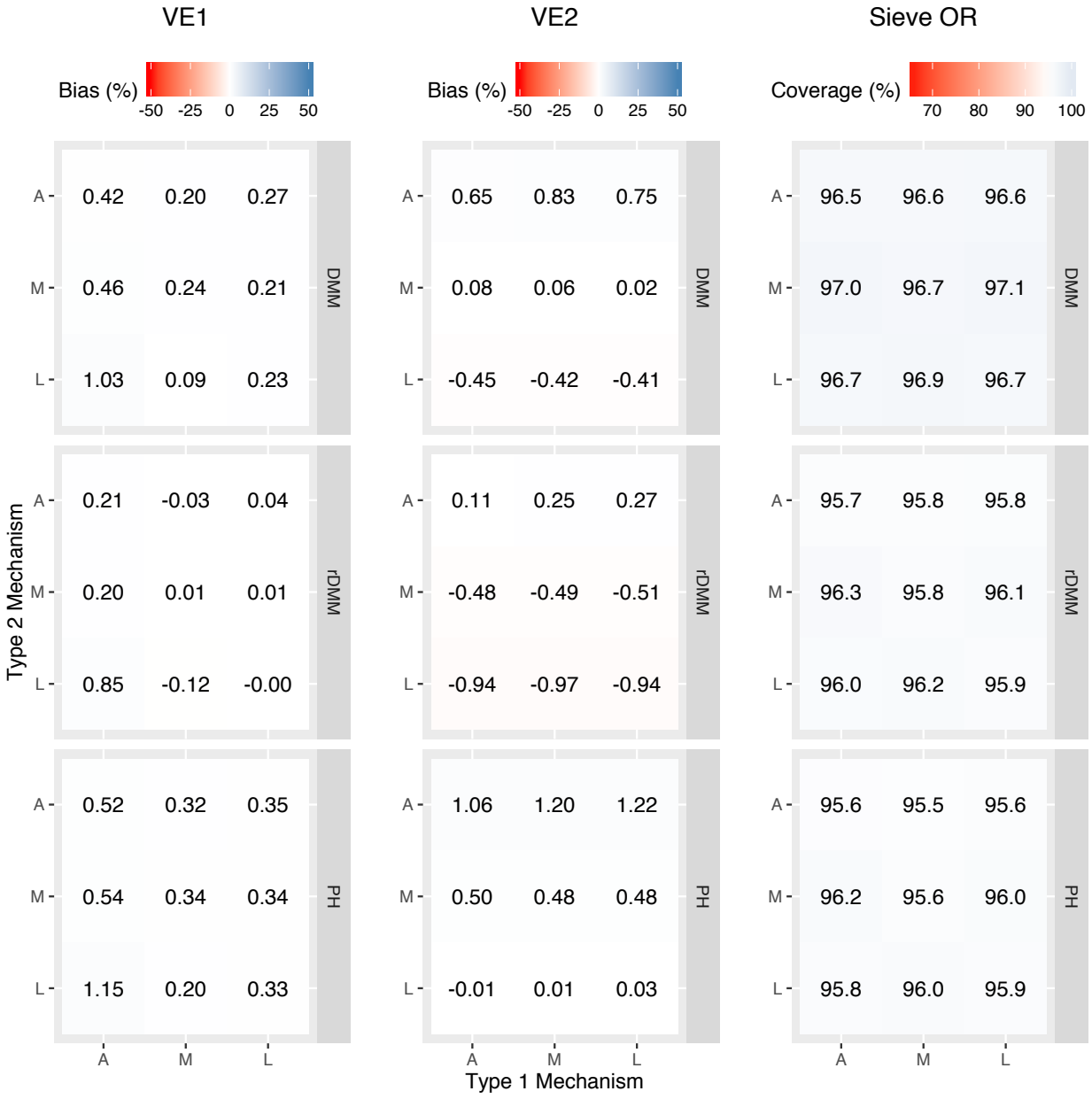
$\mathrm{E}[\boldsymbol{n_p}] = \mathbf{2000}$ and $\boldsymbol{p} = \mathbf{0.5}$

Figure 3.3: Heatmaps show estimated VE_1 (panel column 1), bias in estimated VE_2 (panel column 2), and coverage of 95% confidence intervals for $\log(OR_{1,2})$ (panel column 3) using three methods (panel rows). Within each heatmap, columns and rows indicate true simulated vaccine mechanism against type 1 and type 2 events, respectively. Heatmap cells represent median percentages across 5,000 simulated datasets.



$$\mathrm{E}[\boldsymbol{n_p}] = \mathbf{125} \text{ and } \boldsymbol{p} = \mathbf{0.9}$$

Figure 3.4: Heatmaps show estimated VE_1 (panel column 1), bias in estimated VE_2 (panel column 2), and coverage of 95% confidence intervals for $\log(OR_{1,2})$ (panel column 3) using three methods (panel rows). Within each heatmap, columns and rows indicate true simulated vaccine mechanism against type 1 and type 2 events, respectively. Heatmap cells represent median percentages across 5,000 simulated datasets.



$\mathrm{E}[\boldsymbol{n_p}] = \mathbf{125}$ and $\boldsymbol{p} = \mathbf{0.1}$

Figure 3.5: Heatmaps show estimated VE_1 (panel column 1), bias in estimated VE_2 (panel column 2), and coverage of 95% confidence intervals for $\log(OR_{1,2})$ (panel column 3) using three methods (panel rows). Within each heatmap, columns and rows indicate true simulated vaccine mechanism against type 1 and type 2 events, respectively. Heatmap cells represent median percentages across 5,000 simulated datasets.

3.4 Power as a function of effect size

We have shown that, in our simulation study setting, inference using the DMM provides rejection of the sieve null hypothesis at an appropriate level regardless of the vaccine mechanism against strains $j = 1, 2$. However, in settings which satisfy the leaky vaccine assumption, the PH model is also valid, and might have some advantages. The PH model has fewer parameters to estimate, since the μ_j are not present in the model at all, and the baseline hazards $\lambda_{0,j}(t)$ are estimated non-parametrically and are not part of the partial likelihood estimation at all (Cox, 1972). In contrast, the DMM has six parameters to estimate: μ_j, θ_j, and $\lambda_{0,j}$ for $j = 1, 2$. The increase in the amount of parameters estimated using the DMM might mean a decrease in the power to detect a sieve effect compared to a simpler model, provided the assumptions of vaccine mechanism are satisfied for both models. In settings where the true vaccine mechanism is "all-or-none" against each type, an analogous decline in power might be expected when comparing to the rDMM method (valid under the 'all-or'none' vaccine mechanism).

To examine the power of the DMM method, we simulated an additional 2500 datasets similar to the ones described above with $\mathrm{E}[n_p] = 500$, and varied the expected proportion of placebo subjects infected ($p = 0.1, 0.5$ or 0.9), and the true vaccine mechanism (leaky (L) or all-or-none (A) against both strain 1 and strain 2). We also varied the true effect size $\log(OR_{1,2})$ from -0.95 to 0.95 (in increments of 0.05). Similar to before, the type 2 prevalence was 3 times that of type 1. For values of $\log(OR_{1,2}) > 0$ (vaccine more effective against type 1 than type 2), we kept VE_1 constant at 0.75, and set $VE_2 = 1 - 0.25e^{\log(OR_{1,2})}$, such that VE_2 varied between 75% and 35%. For values of $\log(OR_{1,2}) < 0$ (vaccine more effective against type 2 than type 1), we kept VE_2 constant at 0.75 and calculated $VE_1 = 1 - 0.25e^{\log(OR_{1,2})}$.

Power was calculated as the proportion of 2500 datasets for which the sieve null hypothesis was rejected by a Wald test of $H_0 : \log(OR_{1,2}) = 0$ at the $\alpha = 0.05$ level using either the DMM or PH methods. When the true vaccine mechanism was all-or-none (A) against both type 1 and type 2, we also evaluated power for the rDMM method. Note that with an all-or-

none vaccine, the PH model does not test at the correct size, which is seen in the power curve being shifted to the right, compared to the correctly specified DMM and rDMM methods.

Figure 3.6 shows the resulting curves, plotting power against effect size. For the simulation parameters listed above, both the DMM and PH show close to 100% power to detect effect sizes near $\left|\log(OR_{1,2})\right| = 1.0$, and decrease to the size of the test (5%) as $log(OR_{1,2})$ decreased to zero, as expected. The decrease in power was slightly faster when the overall prevalence was lower ($p = 0.1$). The DMM method had noticeably less power than the simpler methods did when $\log(OR_{1,2})$ was in a moderate range, where the methods had power closer to the middle of the range between 5% and 100%, though the magnitude of the effect depended on the simulation parameters. To examine this phenomenon in more detail, Figure 3.7 shows the relative power of DMM vs the correctly specified simpler method (PH when the vaccine was leaky, and rDMM when the vaccine was all-or-none) as a function of the effect size $\log(OR_{1,2})$.

In general, when the expected prevalence was low ($p = 0.1$), the relative power of DMM compared to PH remained between 80% and 100% (Figure 3.7, Panel A), tending towards local minimums around $\log(OR_{1,2}) = -0.25$ or $+0.25$. The decrease in relative power was asymmetrical: DMM performed relatively worse for negative values of $\log(OR_{1,2})$ (when the vaccine was less efficacious against the less common type 1) than for positive values of the same magnitude. A similar trend was seen when $p = 0.5$, but the relative power decreased to around 60% for values of $\log(OR_{1,2}) = -0.25$, but only to around 80% for $\log(OR_{1,2}) = +0.25$ (Figure 3.7, Panel B). However, the trend was reversed when $p = 0.9$ (Panel C) or when the true vaccine mechanism was all-or-none (Panel D): relative power decreased to around 60% when $\log(OR_{1,2}) = +0.25$ but only to around 75% when $\log(OR_{1,2}) = -0.25$. A possible explanation for this observed asymmetry will be discussed in Section 5.4.

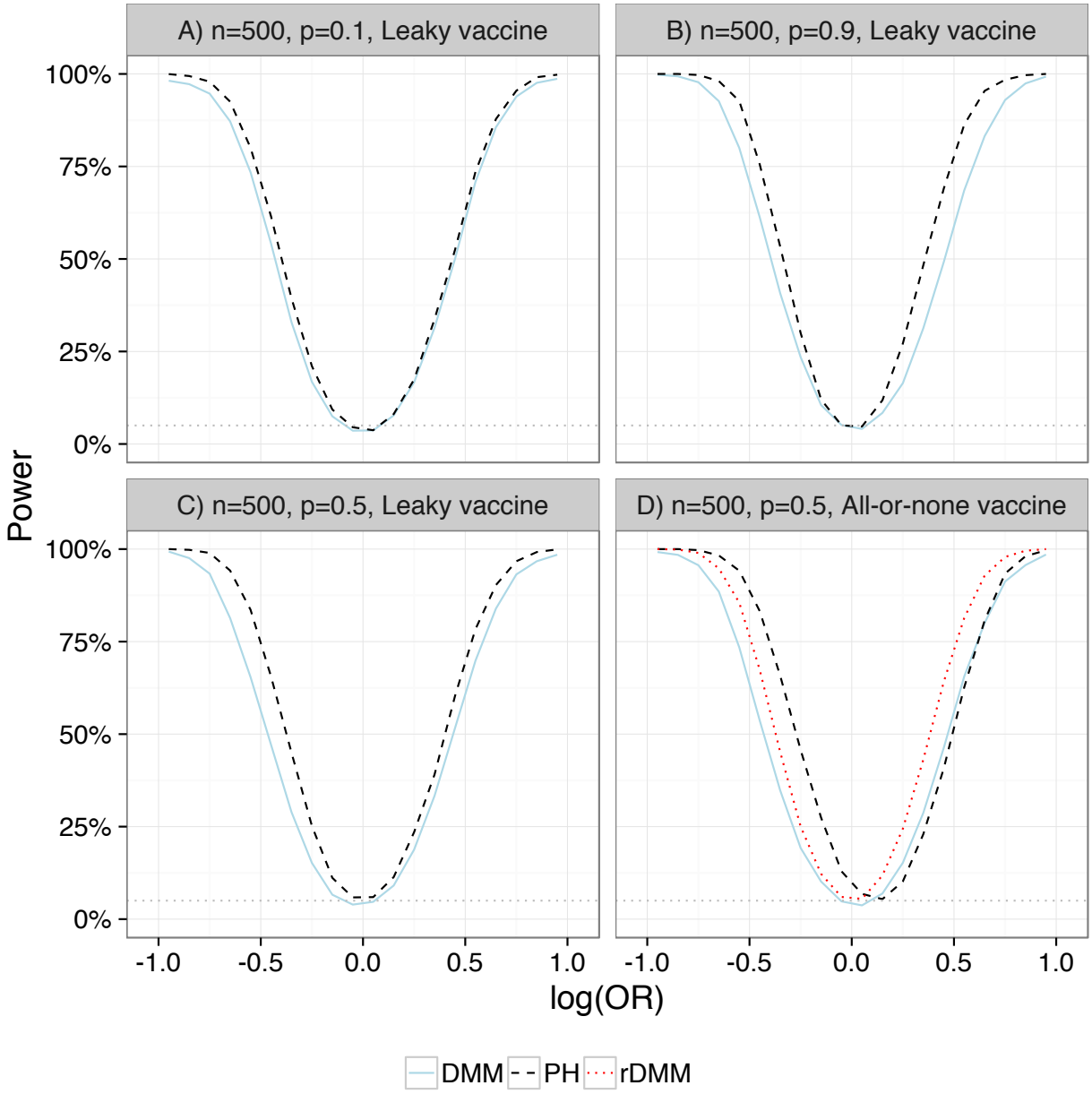


Figure 3.6: We simulated datasets with $\mathrm{E}[n_p] = 500$ and $p = 0.1$ with a leaky vaccine ($\mu_j = 1$ for all j) [Panel A], $p = 0.9$ with a leaky vaccine [Panel B], $p = 0.5$ with a leaky vaccine [Panel C], and $p = 0.5$ with a non-leaky vaccine model ($\theta_j = 1$ for all j) [Panel D]. For each value of $\log(OR)$ between -0.95 and 0.95 (in increments of 0.10), power was calculated as the proportion of 2,500 simulated datasets for which $H_0 : \log(OR_{1,2}) = 0$ was rejected at the $\alpha = 0.05$ level by a Wald test using the DMM or PH methods [Panels A, B, C]. In Panel D, we also evaluated power using the rDMM method. For reference, a horizontal dotted line is drawn at $\alpha = 0.05$.

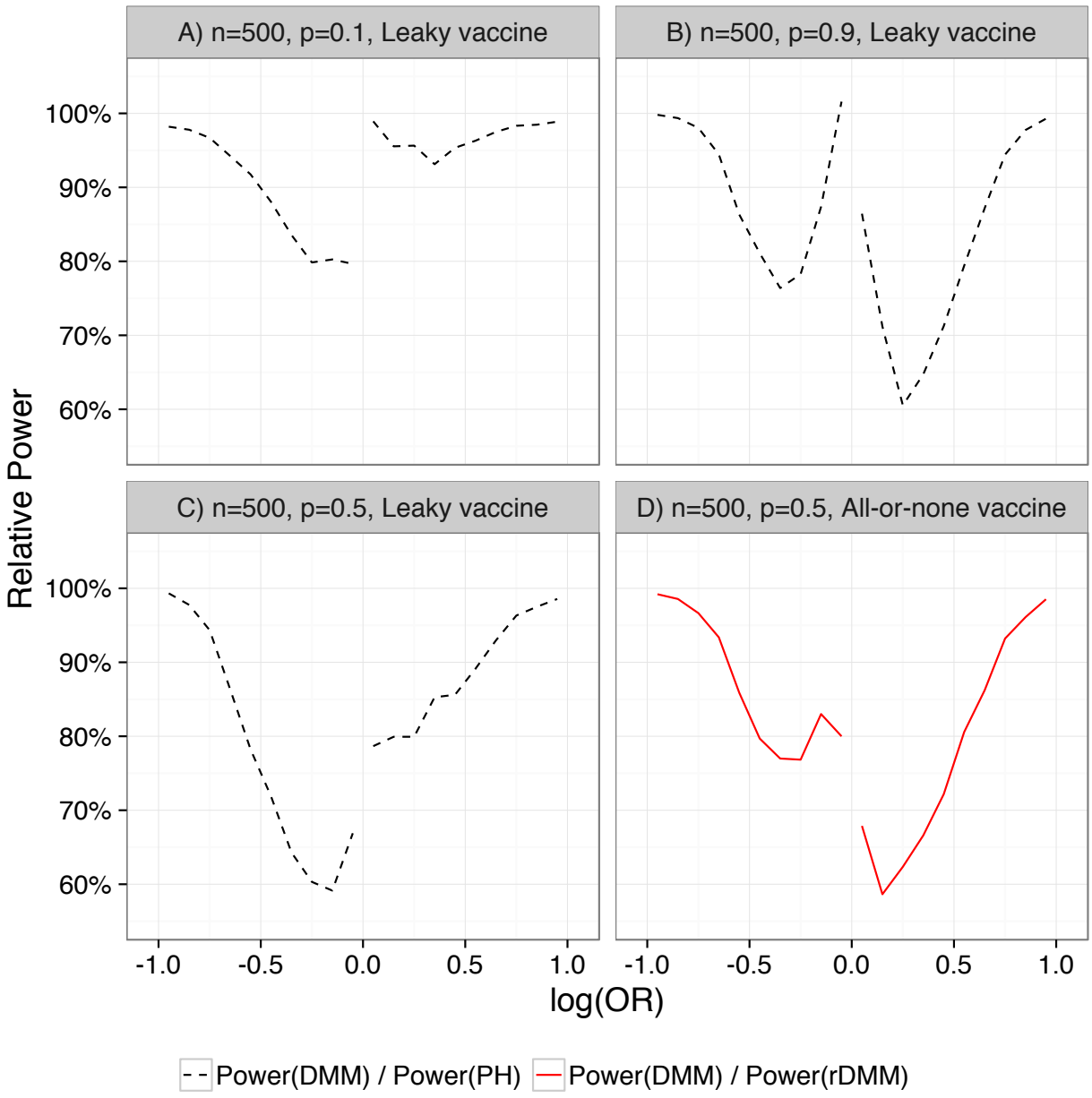


Figure 3.7: For simulated datasets as described in Figure 3.6, we calculated the relative power of DMM to detect various effect sizes versus the correctly specified simplified model (either PH when the vaccine was leaky, Panels A-C, or rDMM when the vaccine was all-or-none, Panel D).

Chapter 4

EXAMPLE USING CYD14 AND CYD15 DATA

Dengue fever is a mosquito-borne illness caused by four genetically distinct serotypes of viruses collectively known as dengue virus (DENV-1 thru DENV-4), and is considered endemic to many tropical and subtropical areas of the world where the disease vector, *Aedes aegypti*, is found. The four serotypes of dengue virus are genetically distinct, and their relative prevalence in a certain region can change quickly and unpredictably over time. Disease caused by one serotype produces lifelong immunity to that serotype, but only short-term immunity to the other serotypes. Thus, the development of a broadly-protecting vaccine is a key part to global strategies to contain the rapidly expanding disease, which causes over 500,000 hospitalizations each year, with approximately 2.5% of cases resulting in death.

In 2011, a recombinant live attenuated tetravalent dengue vaccine (Dengvaxia/CYD-TDV) entered Phase III testing in Southeast Asia (Capeding et al., 2014) and Latin America (Villar et al., 2015), with the CYD14 and CYD15 trials respectively. Both trials found that the vaccine was efficacious against virologically confirmed dengue fever (VCD); however, the vaccine was not equally efficacious against disease caused by all four serotypes. In particular, while VE against overall VCD was estimated at 54.8% (95% confidence interval: 46.8, 61.7) and 64.7% (58.7, 69.8) in the intention-to-treat analyses, efficacy against DENV-2 was much lower- only 34.7 % (10.4, 52.3) and 50.2% (31.8, 63.6) in CYD14 and CYD15, respectively. Analyses of sequences from VCD endpoint cases in CYD14 revealed that this disparity may have been due to the greater intra-serotype diversity of DENV-2, and suggests that CYD-TDV may have induced an immune response with limited cross-protective properties beyond the specific genotypes of DENV-1 thru DENV-4 found in the vaccine (Juraska et al., 2017).

Among the many sieve analyses performed by Juraska et al. (2017) were comparisons of VE across categorical variables using sieve methods relying on proportional hazards models

(PH). As a demonstration, we used the PH, DMM and rDMM methods to determine the extent to which VE varied among the four serotypes of DENV in both CYD14 and CYD15. Our aim is to compare the performance of the three models in the setting of a Phase III clinical trial.

4.1 Description of data

Both CYD14 and CYD15 were multicenter, randomized, observer-masked, placebo-controlled Phase 3 clinical trials. CYD14 was conducted among healthy children ages 2-14 years in five countries in Southeast Asia (Indonesia, Malaysia, Philippines, Thailand and Vietnam), while CYD15 was conducted among healthy children ages 9-16 years in five countries in Latin America (Colombia, Brazil, Mexico, Puerto Rico, and Honduras). In both studies, subjects were randomly assigned in a 2:1 ratio to receive three doses of vaccine or placebo (at 0, 6 and 12 months), stratified by study site and age group.

Data were collected via active surveillance starting on the day of the first vaccine dose - the families of study subjects were contacted weekly and reminded to visit the trial health-care center in case of acute fever of at least 38°C on 2 or more consecutive days. During center visits, blood samples were obtained within 5 days of fever onset, and tested for the presence of DENV using both a quantitative reverse-transcriptase polymerase-chain-reaction (RT-PCR) and an enzyme-linked immunosorbent assay (ELISA) to test for dengue non-structural protein 1 (NS1) antigen. Subjects were classified as cases of VCD if either test was positive, and both tests were serotype-specific for DENV1 thru DENV4. Almost all subjects had either 0 or 1 occurrences of the symptomatic VCD primary study endpoint in during their respective trial - for subjects with more than one event (fewer than 20 subjects in both studies), only the first one was used in the sieve analyses. For 5 placebo recipients in CYD14, 8 placebo recipients in CYD15, and 6 vaccine recipients in CYD15, multiple serotypes of DENV were isolated at a single clinic visit; for these subjects, a single serotype was randomly selected for the analysis. The resulting endpoint counts for both trials are shown in Table 4.1.

	CYD14			CYD15		
	Vaccine	Placebo	Total	Vaccine	Placebo	Total
DENV-1	116	119	235	96	105	201
DENV-2	94	70	164	80	81	161
DENV-3	30	43	73	54	104	158
DENV-4	39	70	109	32	83	115
Total VCD	279	302	581	262	373	635
No VCD	6567	3120	9687	13652	6567	20219
Total	6846	3422	9687	13914	6940	20854

Table 4.1: Counts of serotype-specific occurrences of symptomatic virologically-confirmed dengue (VCD) study endpoints in CYD14 and CYD15.

4.2 Results

We used identical procedures to analyze the CYD14 and CYD15 data, using the DMM and rDMM models with $K = 4$ to obtain point estimates and standard errors of VE_j for $j = 1, ..., 4$ corresponding to the DENV-1 thru DENV4 events shown in Table 4.1. For comparison, we also fit a proportional hazards model using the method described in Lunn and McNeil (1995). We performed tests for equality of the VE_j using the estimates of standard error derived in Section 2.3. More specifically, we first tested

$$H_{0,overall} : \log(OR_{1,2}) = \log(OR_{1,3}) = \log(OR_{1,4}) = 0$$

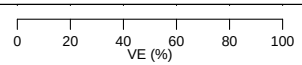
by comparing the Wald test statistic to the null distribution of $\chi^2_{(3)}$. If $H_{0,overall}$ was rejected at the $\alpha = 0.05$ level, we then performed pairwise tests of

$$H_{0,j,k} : \log(OR_{j,k}) = 0$$

for all $j \neq k$ by comparing the z-statistic to standard normal distribution. The results are shown in Tables 4.2 and 4.3.

A) Proportional Hazards Model (PH)

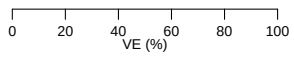
Endpoint	μ (Fixed)	$\hat{\theta}$	VE (%)	95% CI	p-value	Differential VE p-value DENV2	DENV3	DENV4
DENV1	1	0.48	52.39	(38.52, 63.14)	1.28×10^{-8}	0.118	0.214	0.018
DENV2	1	0.66	34.44	(10.67, 51.89)	7.48×10^{-3}	—	0.021	0.001
DENV3	1	0.34	66.02	(45.83, 78.68)	5.70×10^{-6}	—	—	0.464
DENV4	1	0.27	72.94	(59.96, 81.71)	6.13×10^{-11}	—	—	—



Wald test: $\chi^2_{(3)} = 13.7$, $p = 0.0033$

B) Degenerate Mixture Model (DMM)

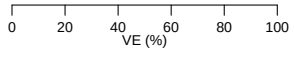
Endpoint	$\hat{\mu}$	$\hat{\theta}$	VE (%)	95% CI	p-value	Differential VE p-value DENV2	DENV3	DENV4
DENV1	0.52	1	52.01	(38.22, 62.72)	1.22×10^{-8}	0.125	0.251	0.020
DENV2	0.26	0.88	34.43	(10.72, 51.84)	7.37×10^{-3}	—	0.035	0.001
DENV3	0.47	0.64	65.9	(42.34, 79.83)	5.95×10^{-5}	—	—	0.508
DENV4	0.48	0.52	72.74	(59.16, 81.81)	2.94×10^{-10}	—	—	—



Wald test: $\chi^2_{(3)} = 12.9$, $p = 0.0050$

C) "All-or-None" Restricted DMM (rDMM)

Endpoint	$\hat{\mu}$	θ (Fixed)	VE (%)	95% CI	p-value	Differential VE p-value DENV2	DENV3	DENV4
DENV1	0.52	1	52.03	(38.27, 62.73)	1.15×10^{-8}	0.120	0.209	0.018
DENV2	0.34	1	34.26	(10.68, 51.62)	7.32×10^{-3}	—	0.021	0.001
DENV3	0.66	1	65.82	(45.61, 78.52)	5.88×10^{-6}	—	—	0.475
DENV4	0.73	1	72.59	(59.54, 81.43)	7.29×10^{-11}	—	—	—

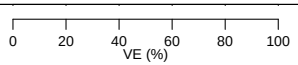


Wald test: $\chi^2_{(3)} = 13.7$, $p = 0.0034$

Table 4.2: Comparison of results from analyses of CYD14 data using the PH, DMM and rDMM models. Tables show point estimates of μ_j, θ_j and VE_j for $j = 1, ...4$, with 95% confidence interval and p-value for testing $VE_j = 0$. Forest plots show point estimates and 95% CIs of VE_j. Differential VE p-values are calculated using Z-tests $VE_j = VE_k$ for pairwise j, k. Omnibus differential VE p-values are calculated using a Wald test of $VE_j = VE_k$ for all j, k.

A) Proportional Hazards Model (PH)

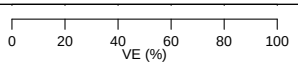
Endpoint	μ (Fixed)	$\hat{\theta}$	VE (%)	95% CI	p-value	Differential VE p-value DENV2	DENV3	DENV4
DENV1	1	0.45	55.24	(40.97, 66.07)	1.25×10^{-8}	0.718	0.009	0.001
DENV2	1	0.48	51.69	(34.20, 64.53)	3.93×10^{-6}	—	0.005	< 0.001
DENV3	1	0.25	74.66	(64.80, 81.76)	2.22×10^{-16}	—	—	0.277
DENV4	1	0.19	81.06	(71.52, 87.40)	1.33×10^{-15}	—	—	—



Wald test: $\chi^2_{(3)} = 19.6$, $p = 2.0 \times 10^{-4}$

B) Degenerate Mixture Model (DMM)

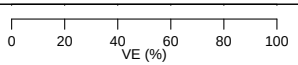
Endpoint	$\hat{\mu}$	$\hat{\theta}$	VE (%)	95% CI	p-value	Differential VE p-value DENV2	DENV3	DENV4
DENV1	0.14	0.52	55.1	(40.78, 65.95)	1.43×10^{-8}	0.717	0.010	0.001
DENV2	0.13	0.56	51.49	(33.93, 64.38)	4.44×10^{-6}	—	0.005	< 0.001
DENV3	0.26	0.34	74.5	(64.57, 81.64)	3.33×10^{-16}	—	—	0.272
DENV4	0.54	0.41	81.03	(71.00, 87.6)	1.68×10^{-14}	—	—	—



Wald test: $\chi^2_{(3)} = 18.9$, $p = 2.8 \times 10^{-4}$

C) "All-or-None" Restricted DMM (rDMM)

Endpoint	$\hat{\mu}$	θ (Fixed)	VE (%)	95% CI	p-value	Differential VE p-value DENV2	DENV3	DENV4
DENV1	0.55	1	54.8	(40.48, 65.67)	1.56×10^{-8}	0.717	0.010	0.001
DENV2	0.51	1	51.21	(33.64, 64.13)	4.80×10^{-6}	—	0.005	< 0.001
DENV3	0.74	1	74.3	(64.34, 81.47)	4.44×10^{-16}	—	—	0.259
DENV4	0.81	1	80.97	(71.41, 87.33)	1.33×10^{-15}	—	—	—



Wald test: $\chi^2_{(3)} = 19.8$, $p = 1.9 \times 10^{-4}$

Table 4.3: Comparison of results from analyses of CYD15 data using the PH, DMM and rDMM models. Tables show point estimates of μ_j, θ_j and VE_j for $j = 1, ...4$, with 95% confidence interval and p-value for testing $VE_j = 0$. Forest plots show point estimates and 95% CIs of VE_j. Differential VE p-values are calculated using Z-tests $VE_j = VE_k$ for pairwise j, k. Omnibus differential VE p-values are calculated using a Wald test of $VE_j = VE_k$ for all j, k.

The results of the analysis confirm the finding that vaccine efficacy differed significantly by serotype in both studies, with VE against DENV-2 being significantly lower than against DENV-3 and DENV-4, and VE against DENV-1 being significantly lower than against DENV-4 (both studies) and DENV-3 (CYD15 only).

In general, estimated vaccine efficacy against all serotypes, and associated standard errors, are very similar using all three models on the CYD14 and CYD15 data. Since fewer than 10% of placebo subjects were infected in either trial (Table 4.1), this is in line with the results in Figure 3.5 and 3.6 (Panel A), in which the DMM and rDMM methods performed very similarly to the PH method regardless of the true vaccine mechanism.

We note that while the point estimates of VE using the DMM and rDMM methods are all within 1% of estimates using the PH method, they are all also slightly closer to zero. This is in line with the results in Figures 3.1 to 3.5, in which we observed that the PH model tends to overestimate vaccine efficacy when the true mechanism is non-leaky. We would expect that the true vaccine mechanism against each serotype can be approximated somewhere on the spectrum between entirely leaky and entirely all-or-none, so lower point estimates are not surprising when using a model that allows for non-leaky vaccines (DMM) or enforces all-or-none vaccines (rDMM).

Chapter 5

DISCUSSION

5.1 Discussion of simulation results

In our simulation studies, we observed that the DMM tended to correctly estimate strain-specific vaccine efficacy regardless of the true vaccine mechanism. Although some degree of bias was present in the VE estimators, possibly due to an edge effect of the parameter space (more on this in Section 5.4), it did not seem to affect the coverage of the 95% CIs.

In contrast, proportional hazards models tended to overestimate strain-specific vaccine efficacy when the true mechanism was non-leaky. The degree of overestimation was greater for strains with higher relative prevalence, strains against which the true vaccine was closer to all-or-none (Figure 3.1), and when the overall disease prevalence was higher (Figures 3.4 and 3.5). The opposite (underestimation of strain-specific VE) was true for rDMM, which restricts the DMM to assuming an all-or-none vaccine mechanism. In all cases where the model was misspecified against at least one type, under- or overestimation of the VEs led to undercoverage of the 95% CIs (and thus an increase in type-1 error when no sieve effect was present). This trend increased with both overall disease prevalence (Figures 3.4 and 3.5) and sample size (Figures 3.2 and 3.3).

The DMM has the advantage of controlling type-1 error regardless of the true vaccine mechanism, but the larger number of coefficients estimated results in some loss of power to detect sieve effects when they were present. However, within the simulation parameters studied, the power of DMM to detect a sieve effect stayed above 60% of that seen when using simpler methods that correctly specified the vaccine model (PH when the vaccine was leaky, and rDMM when the vaccine was not). The relative performance of the DMM increased when the targeted effect size was larger, and when the overall prevalence was lower (Figure 3.7).

However, note that the DMM provides accurate results regardless of the vaccine mechanism, and that when the vaccine is a mixture of leaky and all-or-none effects, the DMM is the *only* model that provides inference at the correct rejection level.

In light of these trends, it appears that using the DMM instead of PH in sieve analyses of vaccine efficacy trials represents a trade-off - the DMM has slightly lower power to detect sieve effects with the same sample size and expected number of events, but has the correct type-1 error rate regardless of the true vaccine mechanism. The difference between the DMM and PH methods becomes less clear when the expected event rate in placebo subjects decreases to 10% or lower - for studies in which this is expected to be the case, the use of DMM in the form introduced here might be more limited to that of sensitivity analyses.

This is also reflected in the results of the sieve analysis of CYD14 and CYD15, which were similar when DMM was used compared to proportional hazards methods. Analyses using the rDMM, which assumes vaccines are all-or-none against each dengue serotype, produced similar results as well. This is not surprising, given that the event rate in these studies was less than 10%.

Despite its limited usefulness in short-term clinical trials, the DMM could still be a useful basis for methods in other types of trials, such as challenge studies, in which a larger proportion of subjects are expected to experience the disease in the course of the trial. However, the DMM still has several limitations as a method, which we discuss in the next few sections.

5.2 Relating the probability of susceptibility to model covariates

One of the potential limitations of the DMM is that it makes the assumption that unvaccinated subjects cannot have strain-specific immunity. However, although it was outside the scope of this study, it is possible to relate both the probability of susceptibility to strain j, as represented by the term μ_j in Equation 2.2, and the degree of susceptibility to strain j in non-immune subjects, as represented by the term θ_j, to model the effect of an arbitrary number of covariates via type-specific coefficient vectors. Such a model might look like the

following:

$$\lambda_j(t|x, z_j) = z_j \cdot \exp(\beta_{j,0} + \beta_j \boldsymbol{x}^T) \cdot \lambda_{0,j}(t) \tag{5.1}$$

$$\begin{gathered} Z_j = U_j \cdot W_j \\ \text{where } U_j \sim \text{Bernoulli}\big(\Phi(\alpha_{j,0} + \alpha_j \boldsymbol{x}^T)\big) \text{ and } W_j \sim f_{x,j}(\cdot) \end{gathered} \tag{5.2}$$

where $\boldsymbol{x}$ is now a vector of covariates including the vaccination status. This would increase the complexity of the model by allowing for unvaccinated subjects to be immune to the disease, and for the probability of immunity to vary with other covariates in the model. This is similar in some ways to the "cure rate" models seen in Maller and Zhou (1996).

5.3 Implications of the no-harm assumption

The no-harm assumption states that the effect of vaccination cannot be to increase the overall (average) susceptibility of subjects - specifically, θ_j is restricted to $[0, 1]$ instead of $[0, \infty)$ as it would usually be in the Cox PH model. We note here that this is actually a stronger assumption than it is in the case of a proportional hazards model - in the PH model, the no harm assumption applies marginally across all subjects in the study who are randomized to receive vaccine. However, in the DMM, the assumption is a conditional statement, applying specifically to the subgroup of subjects who receive vaccine but who are not completely immune (*i.e.* $Z_j \neq 0$ in Equation 2.2).

This assumption might not be defensible in certain situations. Violation of the no-harm assumption when using the Cox PH model is possible - for example, in the STEP study of an HIV-1 vaccine, the point estimate for VE was negative, which was a surprising apparently violation of the no-harm assumption (Buchbinder et al., 2008). However, even when the no-harm assumption might appear to be met when using the Cox PH model, it might be violated when using the DMM model. For example, it might be plausible that subjects who

do not respond to vaccination by becoming immune share a unobservable trait that also causes vaccination to increase susceptibility.

We utilized the no-harm assumption while fitting the DMM out of necessity - we found that if the value of θ_j was allowed to be arbitrarily high, that in some simulated datasets, fitted parameter sets with extremely large θ_j and μ_j extremely close to zero were not distinguishable from sets with values closer to the center of the parameter space in terms of overall model likelihood. While this may be revealing of an overall identifiability issue, a better solution might be to use penalized likelihood estimation to keep the estimated parameters away from extremely high and unlikely values.

5.4 Boundary effects when using DMM

We observed, in Figures 3.1 thru 3.5, that there are small amounts of bias observed in estimating strain-specific VE with the DMM when the true vaccine model was either leaky or all-or-none. We noted that the leaky and all-or-none vaccine models are actually boundary cases of the DMM. In the case of a truly leaky vaccine, using the DMM will result in more underestimation of the VE on average because geometrically, there are more combinations of θ_j and μ_j that result in a lower estimated VE than a higher one. The opposite is true when the true vaccine mechanism is all-or-none. In both of these cases, we would expect the bias to shrink when the sample size increases, which is indeed what happens (Figures 3.2 and 3.3). More simulation studies in which the true vaccine mechanism is very close to, but not completely leaky (or all-or-none) might be needed to clarify this phenomenon.

5.5 Future directions

In addition to the improvements suggested in Sections 5.2 and 5.3, there are several other areas in which improvements to the DMM could be made.

First, future work could examine the possibility of fitting a type-specific mixture model with a parameterized, nondegenerate frailty distribution to more accurately model subject-level heterogeneity in susceptibility. Similar to the frailty mixture models of Longini and

Halloran (1996) in the case of a single disease type, the parameters of the type-specific frailty distributions would be treated as nuisance parameters, but more work remains to be done in simulation studies to determine whether the inclusion of nondegenerate frailty distributions in the model has an effect on the estimation and testing of sieve effects.

We also noted earlier that in the scope of our study, we assumed that the subject-specific baseline hazard functions are constant. In reality, it would make more sense to model the hazard functions somehow (using a Weibull distribution, or other parameterization on the event times, or a piecewise constant function). However, we caution that increasing the complexity of the baseline hazard functions may cause some additional identifiability issues with modeling the frailty distributions, especially if the probability of susceptibility was depended on vaccination status and other covariates, as it does in the proposed model 5.2. Note that in this model, because the baseline value of $\mu_j = \Phi^{-1}(a_{j,0})$ is not necessarily equal to one, the presence of immune subjects with 'baseline' covariate values might complicate the estimation of a baseline hazard. Additionally, the presence of frailty distributions may cause some identifiability issues with estimation of baseline hazards (Elbers and Ridder, 1982).

Finally, we have assumed that the random distributions describing susceptibility (Equation 2.2) are independent between strains. Future study could focus on modeling the dependence between immunity to certain strains directly, or on expanding the generality of the method by including multivariate distributions for the frailty portion of the mixture models.

Bibliography

Aalen, O. (1988). Heterogeneity in survival analysis. *Statistics in Medicine*, 7:1121–1137.

Brunet, R., Struchiner, C., and Halloran, M. (1993). On the distribution of vaccine protection under heterogeneous response. *Mathematical Biosciences*, 116:111–125.

Buchbinder, S. P., Mehrotra, D. V., Duerr, A., Fitzgerald, D. W., Mogg, R., Li, D., Gilbert, P. B., Lama, J. R., Marmor, M., del Rio, C., et al. (2008). Efficacy assessment of a cell-mediated immunity hiv-1 vaccine (the step study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *The Lancet*, 372(9653):1881–1893.

Capeding, M. R. et al. (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. *The Lancet*, 384(9951):1358–1365.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: The identifiability of the proportional hazard model. *The Review of Economic Studies*, 49(3):403–409.

Gilbert, P. (2000). Comparison of competing risks failure time methods and time-independent methods for assessing strain variations in vaccine protection. *Statistics in Medicine*, 19(22):3065–3086.

Gilbert, P. (2001). Interpretability and robustness of sieve analysis models for assessing hiv strain variations in vaccine efficacy. *Statistics in Medicine*, 20:263–279.

Gilbert, P., Self, S., and Ashby, M. (1998). Statistical methods for assessing differential vaccine protection against human immunodeficiency virus types. *Biometrics*, 54(3):pp. 799–814.

Gilbert, P., Self, S., Rao, M., Naficy, A., and Clemens, J. (2001). Sieve analysis: methods for assessing from vaccine trial data how vaccine efficacy varies with genotypic and phenotypic pathogen variation. *Clinical Epidemiology*, 54:68–85.

Halloran, M., Haber, M., and Longini, I. (1992). Interpretation and estimation of vaccine efficacy under heterogeneity. *American Journal of Epidemiology*, 136(3):328–343.

Halloran, M., Longini, I., and Struchiner, C. (1996). Estimability and interpretation of vaccine efficacy using frailty mixing models. *American Journal of Epidemiology*, 144(1):165–173.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73(2):387–396.

Juraska, M., Magaret, C., Shao, J., Carpp, L., Fiore-Gartland, A., Benkeser, D., Girerd-Chambaz, Y., Langevin, E., Frago, C., Guy, B., Edlefsen, P., and Gilbert, P. (2017). Viral genetic diversity and protective efficacy of a tetravalent dengue vaccine in a phase 3 trial in asia. *Proceedings of the National Academy of Sciences (submitted)*.

Longini, I. and Halloran, M. (1996). A frailty mixture model for estimating vaccine efficacy. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(2):pp. 165–173.

Lunn, M. and McNeil, D. (1995). Applying cox regression to competing risks. *Biometrics*, 51(2):pp. 524–532.

Maller, R. and Zhou, X. (1996). *Survival Analysis with Long-Term Survivors*. Wiley and Sons.

Prentice, R., Kalbfleisch, J., Peterson, A., Flournoy, N., Farewell, V., and Breslow, N. (1971). The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4):pp. 541–554.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Smith, P., Rodrigues, L., and Fine, P. (1984). Assessment of the protective efficacy of vaccines against common diseases using case-control and cohort studies. *International Journal of Epidemiology*, 13(1):87–93.

Villar, L. et al. (2015). Efficacy of a tetravalent dengue vaccine in children in latin america. *New England Journal of Medicine*, 372(2):113–123.