Sieve Analysis: Statistical Methods for Assessing Genotype-Specific Vaccine Protection in HIV-1 Efficacy Trials with Multivariate and Missing Genotypes

Michal Juraska

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

Doctor of Philosophy

University of Washington

2012

Reading committee:

Peter Gilbert, Chair

Ying Qing Chen

Ross Prentice

Program Authorized to Offer Degree: Department of Biostatistics

University of Washington

Abstract

Sieve Analysis: Statistical Methods for Assessing Genotype-Specific Vaccine Protection in HIV-1 Efficacy Trials with Multivariate and Missing Genotypes

Michal Juraska

Chair of the Supervisory Committee:
Professor Peter B. Gilbert
Department of Biostatistics

The extensive diversity of the human immunodeficiency virus type 1 (HIV-1) poses a major challenge for the design of a successful preventive HIV-1 vaccine. Thus an important component of HIV-1 vaccine development is the assessment of the impact of HIV-1 diversity on vaccine protection against HIV-1 acquisition. Statistical methods to evaluate whether and how vaccine efficacy depends on genetic features of exposing viruses in data collected in randomized double-blinded placebo-controlled Phase IIb/III preventive HIV-1 vaccine efficacy trials are developed. To characterize exposing HIV-1 strains, their genetic distances to the multiple HIV-1 sequences included in the vaccine construct are measured, where the set of genetic distances is considered as the continuous multivariate 'mark' observable in infected subjects only. A mark-specific vaccine efficacy model is described in the framework of competing risks failure time analysis that allows improved efficiency of estimation, relative to current alternative approaches, by using the semiparametric method of maximum profile likelihood estimation in the vaccine-to-placebo mark density ratio model. In addition, the model allows to employ a more efficient estimation method for the overall hazard ratio in the Cox model. Mark data proximal to the time of HIV-1 acquisition,

that are of greatest biological relevance, are commonly subject to missingness due to the intra-host HIV-1 evolution. Two inferential approaches accommodating missing marks are proposed: (i) weighting of the complete cases by the inverse probabilities of observing the mark of interest (Horvitz and Thompson, 1952), and (ii) augmentation of the inverse probability weighted estimating functions for improved efficiency and model robustness by leveraging auxiliary information predictive of the mark (using the general theory of Robins, Rotnitzky, and Zhao (1994)). The missing-mark methods provide a general framework for parameter estimation in density ratio/biased sampling models in the presence of missing data. The proposed methodology can serve either to make inference about whether and how vaccine efficacy varies with *prespecified* genetic distance measures, or as an exploratory tool to identify distance definitions with the greatest decline in vaccine efficacy, characterizing potential correlates of immune protection and indicating pathways for improved HIV-1 vaccine design. The developed methods are applied to HIV-1 sequence data collected in the RV144 Phase III preventive HIV-1 vaccine efficacy trial.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

Peter Gilbert has inspired me with his deep blend of statistical and scientific knowledge, and his brilliant understanding of the interplay between data and larger questions within science. But I must thank you, Peter, for your generosity of time, effort and attention, and for how you treated me as a colleague and friend. Throughout this process, you have shared your wisdom, perspective and encouragement in ways crucial to my scholarship and growth as a researcher, and I will always remain grateful to you.

Growing up, I was fascinated by my father's stacks of math exams that he had brought home to grade. A high-school math teacher, he infused in me an early appreciation for math. My mother, a biochemist and immunologist, had wished that I would study a biomedical science. In my current study of biostatistics, I have finally honored their intellectual interests and have made them very happy (or so they tell me!). Thank you, mom and dad, for your unconditional love and care, and for always supporting me in my academic pursuits, even from faraway.

Parts of this dissertation were written in the summers of 2010 and 2011 on the Kansas farm of my dear friends, Lane and Sarah Senne. I met them when Lane and I were graduate students at Kansas State University in 2004-5. They welcomed me into their family, blessing me with a loving, supportive and restful place to retreat to, and do good work at, during pivotal points in my research. Their faith taught me that I can do all things through Christ who gives me strength. Thank you, Lane and Sarah, for your abiding friendship.

Patrick and Tammie Lurlay have similarly blessed me with their faithful fellow-

ship. Whenever I have needed a local refuge, I have always been welcome to their peaceful home on the Olympic Peninsula, where some of the theoretical portions of this dissertation were finished. Playing music with them in their light- and love-filled home was a boost to my mind and soul. They are a loving couple whose commitment to a balance of family life and work continue to inspire me. Thank you, Patrick and Tammie, for your joy, your music and your eternal perspective on what it means to cultivate grounded and godly relationships.

DEDICATION

To my parents Dalibor and Eugenia, and brothers Tomas and Juraj

Chapter 1
INTRODUCTION

The development of a safe and efficacious preventive human immunodeficiency virus type 1 (HIV-1) vaccine provides the best long-term solution to controlling the global HIV-1 pandemic. Yet, the remarkable degree of genotypic and phenotypic diversity within HIV-1, reflected by the presence of HIV-1 subtypes, circulating recombinant forms, and continual viral evolution within populations and infected individuals, presents a significant problem in the design of broadly protective vaccine prototypes. A modern vaccine candidate may protect against challenge by viral strains that are the same or genetically close to the strain(s) contained in the tested vaccine, however, if the breadth and potency of vaccine-induced immune responses are not sufficient, it may fail to protect against divergent HIV-1 strains. Thus, an important component of HIV-1 vaccine development, referred to as the *sieve analysis*, is the assessment of the impact of HIV-1 diversity on HIV-1 vaccine effects. This dissertation develops statistical methods to evaluate whether and how the protection against HIV-1 acquisition conferred by the vaccine depends on genetic features of the transmitting virus. Detection and characterization of such a dependence can help guide HIV-1 vaccine research toward development of a vaccine with a greater breadth of protection. Because of the rapid and incessant viral adaptation in response to the host immune activity, it is important to note a difference between the objective of sieve analysis – characterization of 'strain-specific' vaccine protection against HIV-1 acquisition – and the study of vaccine-induced postinfection effects on the early evolution of the transmitted virus.

The proposed statistical methods are designed to evaluate 'strain-specific' vac-

cine efficacy in data sets arising from randomized double-blind placebo-controlled Phase IIb or Phase III preventive vaccine efficacy trials in which uninfected volunteers at risk of acquiring HIV-1 are randomly allocated to receive a candidate vaccine or placebo and monitored for HIV-1 infection. In such trials, besides observing the time from randomization to HIV-1 infection diagnosis, for the volunteers who become infected during the trial, we can isolate the viral RNA from postinfection clinical sample(s) (plasma sample is the predominant sample type) and, by using sequencing techniques, recover information about the genetic sequence of the isolated strain(s). Due to the extensive HIV-1 sequence variation, there are billions of distinct viruses circulating in the population of individuals exposing participants of a vaccine trial. Of those, however, only viral strains that establish infection and are detectable by HIV-specific PCR assays can be observed. Hence, resulting from natural immunity and vaccination, the sieve represents an immunobiological barrier to infection that sifts out observable strains from the swarm of strains an individual is exposed to.

The HIV-1 sequence data measured in infected trial participants serve to characterize genetic divergence of the isolated HIV-1 from the strain(s) included in the vaccine. To maximize biological relevance and statistical power, it is important to specify the measure of genetic divergence of an exposing virus to reflect the relative chance that a vaccine-induced immune response will be able to react with and kill the exposing HIV-1. That is, the chosen genetic distance reflects a biological model of cross-reactivity of the vaccine-induced immune response, wherein the vaccine is hypothesized to stimulate a protective immune response to HIVs with small distances to the vaccine insert sequence(s) but not to HIVs with the largest genetic distances, with each increment in genetic distance making protective cross-reactivity less likely. Because of challenges involved in modeling cross-reactivity, multiple immunologically meaningful genetic distances are considered for evaluation, based on (i) different conceptual definitions, (ii) different HIV sequence regions such as V2 or the CD4 binding site, (iii) different methods to specify HIV envelope peptides that may contain anti-

body epitopes, (iv) different reference sequences inside the vaccine, and (v) different ways to accommodate the multiple HIV sequences measured from individual subjects. For instance, Rolland et al. (2011) analyzed a total of 20 genetic distance measures in the 'Step' HIV-1 vaccine efficacy trial.

The genetic distance of interest may take on a unique value for each infected subject, therefore it is natural to consider it as a continuous cause of failure referred to as a *mark* variable to denote that it is only meaningfully measured in those experiencing the failure. The discretization of continuous mark data is inadequate due to data coarsening and biological ambiguity in specifying cut points defining the discrete marks. In the following chapters we develop estimation and hypothesis testing procedures to evaluate whether and how vaccine efficacy depends on the continuous mark. To address multidimensionality of the mark, a simple approach to the analysis of mark-specific vaccine efficacy would be to consider a single distance measure and collapse the multivariate mark to the minimum distance making the assumption that it is sufficient for protection that the exposing virus is genetically close to the vaccine in terms of at least one of the prespecified distances. This approach, however, suffers from the following deficiencies: (i) it precludes to compare the levels of dependence of the vaccine effect on various distance definitions, and (ii) it is possible that protection against infection is provided only when the exposing virus is near to the vaccine in a way that requires to consider the joint distribution of the mark. Therefore, our approach to analyzing mark-specific vaccine efficacy allows to flexibly accommodate a multivariate mark.

The choice of the sequence data used to define the mark needs to be carefully considered taking into account the HIV-1 within-host evolution and the trial-specific HIV testing algorithm. The most relevant mark, based on the actually transmitted strain, is largely unobservable due to rapid HIV-1 evolution. In HIV-1 vaccine efficacy trials, participants are screened for HIV-specific antibodies at periodic intervals, e.g., every 3 or 6 months. Antibody-based immunoassays (for example, ELISA)

have a nearly perfect sensitivity when the HIV transmission event precedes antibody testing by at least 4 weeks, otherwise the HIV-specific antibodies are likely to remain undetected. Furthermore, due to the frequency of testing, the earliest positive antibody-based (Ab+) test results are obtained from blood specimens often drawn weeks or months after the HIV transmission event. Nevertheless, for each participant with an Ab+ test result, earlier collected blood specimens are assayed with the HIV-1 nucleic acid PCR test which has a nearly perfect sensitivity when the HIV-1 transmission occurs at least 1 week prior to testing. Consequently, the PCR assay allows to detect the presence of HIV-1 in earlier infected blood specimens that yield an Ab− test result. Based on this "look-back" procedure, we can classify infected trial participants into one of two groups according to whether their earliest PCR+ specimen is Ab− ('acute'-phase sample) or Ab+ ('post-acute'-phase sample). The acute-phase virus has been proven to well-approximate the transmitting strain (Keele et al., 2008), although some CD8+ T-cell escape may occur within a few weeks after HIV transmission (Goonetilleke et al., 2009). Defining the mark based on a post-acute strain that has undergone substantial evolution and exhibits a number of mutations may lead to erroneous conclusions about the relationship between vaccine efficacy and the exposing virus. One solution, therefore, is to consider marks defined by strains observed in the PCR+ and Ab− phase. The fraction of infected subjects observed in the acute phase primarily depends on the frequency of HIV testing. To approximate the fraction as a function of the HIV testing frequency, we can consider the following simplified scenario: for antibody-based assays, assume 100% sensitivity for transmission events at least 4 weeks prior to testing and 0% sensitivity for those 0–4 weeks prior to testing. For HIV nucleic acid PCR assays, assume 100% sensitivity for transmission events at least 1 week prior to testing and 30% average sensitivity for those 0–1 week prior to testing. Furthermore, assume 100% specificity for both types of assays. Consequently, a tested sample will be PCR+/Ab− with probability 1 if the transmission event occurs 1–4 weeks prior to testing or with the average probability

Table 1.1: Approximated fractions of infected subjects observed in the acute phase. The 'basic' schedule considers the specified HIV testing period throughout the study whereas the 'extended' schedule additionally considers 1-monthly HIV testing during the initial 6 months. We assume that 25% of transmission events occur during the introductory 6-month period.

	Acute-phase fractions (%)	
Testing period (months)	'Basic' schedule	'Extended' schedule
1	82.5	82.5
3	27.5	41.3
6	13.8	30.9

0.3 if the transmission event occurs 0–1 week prior to testing, i.e., the model considers a 3-week time window for a 'guaranteed' PCR+/Ab− test result and a 1-week time window for a 'partially guaranteed' PCR+/Ab− test result. If we additionally assume that the time of transmission is uniformly distributed, Table 1.1 summarizes approximated fractions of PCR+/Ab− samples for a 'basic' and an 'extended' HIV testing schedule. In the next planned efficacy trial, HIV testing will be conducted every month (Gilbert et al., 2011), allowing an increased number of infected subjects to be caught in the acute phase of infection.

A 'complete-case' analysis of mark-specific vaccine efficacy that ignores subjects with missing acute-phase marks may be severely biased and inefficient. Therefore, we extend our 'complete-case' inferential procedures to accommodate missing at random continuous multivariate marks. To the best of our knowledge, there is no alternative statistical method in the existing literature that allows to specify marks of the aforementioned characteristics. Other reasons for a missing acute-phase mark include a missing blood sample or a technical failure in the HIV sequencing procedure, and

thus the extended method is designed to allow separate models for different types of missingness.

The next chapters are arranged as follows. Chapter 2 introduces the semiparametric model for the analysis of mark-specific vaccine efficacy defined as one minus the mark-specific vaccine-versus-placebo hazard ratio of infection that accommodates multivariate marks, completely observed in all infected subjects. The mark-specific vaccine efficacy $VE(t, v)$ approximately measures the multiplicative effect of the vaccine to reduce the susceptibility to infection by strain v given exposure to strain v at time t (Gilbert, McKeague, and Sun, 2008). The estimation method takes advantage of the factorization of the mark-specific hazard ratio into the vaccine-versus-placebo mark density ratio and the ordinary marginal hazard ratio. The two factors are estimated separately - the former using the method of maximum profile likelihood estimation in the density ratio/biased sampling model (Qin, 1998) under the assumption of time and covariate independence, and the latter using the method of maximum partial likelihood estimation in the Cox model. Furthermore, we characterize the joint limiting distribution of the combined estimator for the Euclidean parameters in the density ratio/Cox model. In addition, we develop likelihood ratio and Wald tests of the null hypotheses of (i) no vaccine protection against any exposing virus, and (ii) uniform vaccine protection against all exposing strains, considering two- and one-sided alternative hypotheses. Finally, we propose a diagnostic Kolmogorov–Smirnov-type test of the conditional independence between failure time and a continuous mark given treatment.

In Chapter 3, we summarize results from a simulation study of finite-sample properties of the semiparametric maximum likelihood vaccine efficacy estimator in the presence of complete mark data. The simulation is designed to mimic 3-year Phase IIb and Phase III two-arm placebo-controlled HIV-1 vaccine efficacy trials. Considering models with univariate and bivariate marks, we study finite-sample bias, asymptotic and empirical standard errors, and coverage probabilities of Wald confidence inter-

vals. In addition, we evaluate size and power of the proposed likelihood ratio and Wald tests. Finally, we investigate robustness of the inferential methods (i) to violation of the model assumption of conditional independence between the failure time and a mark (we also examine size and power of the proposed diagnostic test of the validity of this assumption), and (ii) to violation of the proportional marginal hazards assumption.

In Chapter 4, we extend the Chapter 2 methodology to accommodate multivariate marks that are subject to missingness. This phenomenon commonly occurs for marks of greatest biological relevance as, for example, acute-phase marks discussed in Chapter 1. We consider two approaches to estimation of mark-specific vaccine efficacy in this setting: (i) weighting of the complete cases by the inverse of the probabilities of observing the mark of interest (Horvitz and Thompson, 1952), and (ii) augmenting of the inverse probability weighted estimating functions by exploiting potential correlation between the mark of interest and collected auxiliary data (following the general theory of Robins, Rotnitzky, and Zhao (1994)). Asymptotic properties of the estimators are derived.

We devote Chapter 5 to summarizing results from a simulation study of finite-sample properties of the proposed mark-specific vaccine efficacy estimators in the presence of missing marks. We evaluate finite-sample bias, asymptotic standard errors, relative efficiencies, and coverage probabilities of Wald confidence intervals under correctly specified missing mark models. We also investigate robustness of the estimation procedures to (i) mis-specification of the missing mark model, and (ii) violation of the missing at random assumption.

In Chapter 6, we conduct a sieve analysis in the RV144 HIV-1 vaccine efficacy trial introduced in Section 1.2, focused on the V1/V2 domain of the HIV-1 envelope gp120 region. Chapter 7 contains concluding remarks and a discussion of future research. Proofs of Theorem 2.1 and auxiliary lemmas are given in Appendix A. An exploratory RV144 sieve analysis considering other V1/V2 and envelope gp120 distance measures

is presented in Appendix B.

1.1 HIV-1 diversity

The enormous HIV-1 sequence variability presents one of the greatest challenges to the development of a vaccine candidate that can induce potent cross-reactive immune responses to worldwide circulating infecting HIV-1 strains. The viral diversity originates in the fact that the reverse transcriptase lacks a proofreading mechanism to confirm that the DNA transcript it produces is a precise copy of the RNA sequence. This phenomenon allows mutations, in particular nucleotide substitutions, insertions, and deletions, to arise owing to which HIV-1 gains the capability to evade the host immune system (mutational escape) – HIV-1 infected persons develop cellular and humoral immune responses to the infecting strains but over time the pressure exerted by the immune system leads to the selection of viral variants that escape responses by neutralizing antibodies (NAb) and CD8+ cytotoxic T lymphocytes. Although cross-reactive immune responses to heterologous strains have been observed (Deeks et al., 2006; Thakar et al., 2005), the breadth and potency of such responses are generally weak (McKinnon et al., 2005).

HIV-1 recombination contributes to further viral diversity. It occurs as a result of coinfection by two different strains that reproduce in the same host cell. The resultant recombinant strain is referred to as a circulating recombinant form (CRF) if it is identified in at least three infected individuals with no direct epidemiologic linkage, otherwise it is termed a unique recombinant form.

The global viral diversity is reflected by the presence of multiple HIV-1 subtypes (phylogenetically linked strains of approximately the same genetic distance from one another) and CRFs. The currently identified subtypes are labelled A, B, C, D, F, G, H, J, and K. Subtype B predominates in the Americas and Western Europe, subtype A in Eastern Europe and Russia, subtype C in southern Africa and India, and subtypes D, F, G, H, J, and K are most prevalent is central Africa. The nu-

cleotide sequence variation within subtypes is between 15 and 20%, whereas that between subtypes is typically between 25 and 35% (Hemelaar et al., 2006) with an increase in both levels of variation observed over time (Korber et al., 2001). CRFs are also of global importance. They typically emerge in regions where multiple subtypes co-circulate with high prevalence; recombination of existent CRFs has also been observed. Currently, 43 CRFs have been described; CRF01_AE and CRF02_AG are highly prevalent in Southeast Asia and West Africa, respectively; others are limited to smaller geographic regions. The implications of the global genetic diversity for vaccine design are unclear, and sieve analysis provides direct tools to gain insight into the effects of HIV-1 diversity on vaccine protection conferred by a candidate vaccine and subsequent guidance for improvement of the vaccine design.

1.2 HIV-1 vaccine development

In more than two decades of HIV-1 vaccine research, a number of vaccine strategies have been pursued. Initially, the vaccine field focused on the development of protein immunogens designed to induce neutralizing antibodies that bind to the trimeric envelope complex on the virion surface. VaxGen, Inc. conducted two Phase III efficacy trials of recombinant envelope glycoprotein (gp) 120 vaccines, AIDSVAX B/E (Vax003 trial) (Pitisuttithum et al., 2006) and AIDSVAX B/B (Vax004 trial) (Flynn et al., 2005) but the monomeric forms of gp120 failed to elicit NAb responses to prevent HIV-1 infection. In the Vax004 trial, HIV-1 RNA was isolated from the earliest postinfection plasma samples, and three full-length gp120 sequences were identified from each of 336 of 368 infected individuals which allows to conduct sieve analysis in this trial. The results of Vax003 and Vax004 have led to more sophisticated antibody-based vaccine strategies to design an immunogen that mimics the trimeric envelope structure and to more distinctly express neutralization epitopes in the conserved regions of gp120 to focus the immune response.

Although T-cell–mediated immune responses may not prevent HIV-1 infection,

they are believed to be an essential immune component in controlling HIV-1 replication after infection (Douek et al., 2006). Vaccine-induced cytotoxic T cell responses may lower viral load during acute infection (Moss et al., 1995) and provide protection against disease progression. A T-cell candidate vaccine using a mixture of recombinant adenovirus type 5 (rAd5) vectors expressing the HIV-1 *gag*, *pol*, and *nef* genes from subtype B was evaluated in two Phase IIb test-of-concept trials in the Americas (the Step study) (Buchbinder et al., 2008) and in South Africa (the Phambili study) (Gray et al., 2010). The Step trial was stopped after the first interim analysis, and subsequently the Phambili trial was discontinued with partial enrollment as the Step trial suggested a potentially increased risk of HIV-1 acquisition due to vaccination in subjects with prior exposure to rAd5.

The large diversity of antibody and T-cell epitopes has motivated vaccine strategies that consider the use of multisubtype consensus sequences (i.e., most recent common ancestor sequences) and/or a combination of immunogens from different subtypes or CRFs. Prime-boost vaccine regimens have been introduced to enhance the breadth and potency of vaccine-induced immune responses. This regimen strategy was used in the Thai RV144 trial (Rerks-Ngarm et al., 2009), a Phase III efficacy trial of the combination of the prime recombinant canarypox vector vaccine (ALVAC-HIV [vCP1521]), with the vector expressing HIV-1 *gag* and *pro* from subtype B together with CRF01_AE gp120, and the booster recombinant gp120 subunit vaccine (AIDSVAX B/E). In the modified intention-to-treat analysis (excluding 7 subjects who tested HIV-1-positive at baseline), the marginal vaccine efficacy to prevent HIV-1 infection within 42 months after the first vaccination was estimated as 31% (95% CI, 1% to 52%; 2-sided p-value = 0.04) which has generated great interest to understand how the vaccine protection may have depended on certain measures of the genetic divergence. Full length HIV-1 sequences were measured from 121 of the 125 infected subjects. Sieve analysis of the RV144 trial data using methodology developed in this dissertation is performed in Chapter 6.

Chapter 2

STATISTICAL METHODS FOR SIEVE ANALYSIS WITH COMPLETE DATA

2.1 Introduction

In this chapter, we develop statistical methods for sieve analysis of preventive HIV-1 vaccines in the presence of complete genetic sequence data. Chapter 4 extends the proposed methods to accommodate missing acute-phase sequence data for a fraction of HIV-infected trial participants.

The fundamental problem of sieve analysis – the absence of exposure data in infection-free trial participants – was first discussed in Gilbert, Self, and Ashby (1998). Collapsing genetic characteristics of the transmitting virus into a single unordered categorical variable, Gilbert, Self, and Ashby (1998) considered an inferential method for this quantity based on the multinomial logistic regression model and proposed a generalization of this model for a continuous viral distance. This work, however, is limited by treating HIV infection as a dichotomous variable, thus ignoring the time to infection. Prentice et al. (1978) proposed the Cox regression method for the analysis of failure times in the presence of finitely many causes of failure (discrete marks). Huang and Louis (1998) developed a nonparametric maximum likelihood estimator for the joint distribution function of the failure time and a continuous mark by representing the joint distribution function through the cumulative mark-specific hazard function. Gilbert, McKeague, and Sun (GMS) (2008) defined the mark-specific vaccine efficacy and proposed a nonparametric estimator for this quantity when the mark is univariate. Furthermore, GMS used the Nelson-Aalen-type estimation for the doubly cumulative mark-specific hazard function to develop nonparametric and

semiparametric procedures for testing of the null hypothesis of zero vaccine efficacy against any exposing virus and the null hypothesis that vaccine efficacy does not depend on the viral divergence.[1] Sun, Gilbert, and McKeague (2009) developed the mark-specific proportional hazards model which allows covariate adjustment and, given the assumption of proportional hazards is valid, may provide more powerful tests of the aforementioned null hypotheses than the GMS's nonparametric method.

In this chapter, we propose a more efficient method of estimation and hypothesis testing for the mark-specific vaccine efficacy in the framework of competing risks failure time analysis that accommodates multivariate marks – thus far assumed to be completely observed in each HIV-infected trial participant. Our approach utilizes the maximum profile likelihood estimation method in the semiparametric density ratio/biased sampling model developed by Qin (1998). Qin and Zhang (1997) proposed a Kolmogorov–Smirnov-type statistic to test the validity of the density ratio model. In parallel with Qin (1998), a similar method of maximum profile partial likelihood estimation was derived by Gilbert, Lele, and Vardi (1999) and Gilbert (2000) for semiparametric biased sampling models with K possibly biased samples, and Gilbert (2004) proposed several goodness-of-fit tests for the K-sample setting.

Our method allows to employ the estimation and testing procedure of Lu and Tsiatis (2008) for the marginal vaccine-to-placebo log hazard ratio γ that utilizes information on auxiliary variables predictive of the failure time (we implemented the Lu and Tsiatis method in the R `speff2trial` package). Their estimator for γ is more efficient than the maximum partial likelihood estimator (MPLE) and the associated Wald test of $H_0 : \gamma = 0$ is more powerful than the log-rank test without requiring assumptions other than those needed for the validity of the MPLE and the log-rank test.

[1] In settings with a univariate mark completely observed in all infected trial participants, these hypothesis tests provide an alternative approach to inference about mark-specific vaccine efficacy. We evaluate and compare their performance to that of our proposed testing procedures in a simulation study presented in Chapter 3.

The remainder of the chapter is organized as follows. In Section 2.2, we introduce the basic notation, describe the structure of the observed data, and discuss the plausibility of the stated assumptions. In Section 2.3, we introduce the estimand of interest – the mark-specific vaccine efficacy. In Section 2.4, we posit a semiparametric model for this quantity, discuss identifiability of Euclidean model parameters, and, for the density ratio part of the model, describe the maximum semiparametric likelihood estimation method. We derive asymptotic properties of the proposed estimator in Section 2.5. Finally, Section 2.6 describes our proposed tests of hypotheses about mark-specific vaccine efficacy. Here we also develop a diagnostic test to assess validity of the $T \perp\!\!\!\perp V|Z$ assumption. Section 2.7 contains concluding remarks.

2.2 Notation and assumptions

Let T denote the continuous time to failure and $V \in \mathbb{R}^s$, a continuous, possibly multivariate, mark variable. Without loss of generality, the support of each component of V is taken to be $[0, 1]$. Let C be the time to censoring. The observed right-censored failure time is $X = \min(T, C)$ with the failure indicator $\delta = I(T \leq C)$. In this chapter, the mark V is assumed to be always observed if $\delta = 1$; otherwise it is unobservable. Let Z denote the indicator of assignment to the treatment group (in vaccine trials, $Z = 1$ indicates vaccine and $Z = 0$ indicates placebo). Let $(X_i, \delta_i, V_i, Z_i)$, $i = 1, \ldots, n$, be i.i.d. replicates of (X, δ, V, Z). The observed data consist of the observations (X_i, V_i, Z_i) for individuals with $\delta_i = 1$ and the observations (X_i, Z_i) for those with $\delta_i = 0$.

We assume that C is conditionally independent of both T and V given Z, that is, $C \perp\!\!\!\perp T|Z$ and $C \perp\!\!\!\perp V|Z$. Additionally, we adopt the assumption $T \perp\!\!\!\perp V|Z$ that ensures identifiability of the Euclidean parameters in the density ratio model introduced in Section 2.4. The addition of the last assumption leads to the following equality of

conditional density functions:

$$f(v|T=t,Z=z)=f(v|T=t,Z=z,\delta=1). \tag{2.1}$$

As a consequence, the assumption $T \perp\!\!\!\perp V|Z$ allows to posit a parametric model for the vaccine-to-placebo mark density ratio using mark data in infected subjects only. We hypothesize that the parametric structure may result in an increased efficiency of vaccine efficacy estimation compared to the alternative approach in Sun, Gilbert, and McKeague (2009) where the dependence of the regression parameter on the mark is modeled nonparametrically. In the HIV vaccine field, the size of the pool of promising vaccine products is small. Therefore, if the objective of an HIV vaccine trial is to test the merit of a vaccine product or concept, an efficiency gain in estimation of mark-specific vaccine efficacy, and, subsequently, an increased control of type II errors are of paramount importance with regard to preventing a costly mistake of discontinuing clinical evaluation of an auspicious candidate as a result of a false negative error (Gilbert, 2010).

In the HIV vaccine trial setting, the assumption $T \perp\!\!\!\perp V|Z$ entails that, for example, vaccine recipients with infection time $T = 6$ months have the same distribution of the mark V as vaccine recipients with infection time $T = 2.5$ years, which may approximately hold given a limited shift in the HIV sequence distribution over the period of 2 years. An analogous statement comparing vaccine recipients with infection times, say, 50 years apart would be clearly incorrect due to the genetic shift of HIV. Hence, the fact that HIV vaccine efficacy trials only last between 3–5 years is crucial for the assumption to be approximately met.

To further assess its plausibility, we consider the impact of a potential selective mechanism of vaccine protection. For example, if the vaccine confers greater protection for individuals with stronger immune systems, we may anticipate that, as the study progresses, the group of at-risk vaccine recipients will have an increasing percentage of subjects with stronger immune systems. If the distribution of HIV strains

infecting subjects with stronger immune systems is different from that for subjects with weaker immune systems, then V may conditionally depend on T given treatment. HIV infection, however, is a rare event in HIV vaccine efficacy trials, typically occurring in $< 10\%$ of trial participants. Consequently, assuming no drop-out, $> 90\%$ of trial participants remain in the risk-set during the entire follow-up period of the trial which makes it plausible that the risk-set composition remains approximately unchanged as the time progresses, and, subsequently, that $T \perp\!\!\!\perp V|Z$ approximately holds.

It is of note that the assumption may be less plausible if the level of protection conferred by the vaccine wanes over the follow-up period. The reason is that, in the presence of a waning vaccine effect, vaccine recipients with larger failure times may have marks closer to zero. In the next planned efficacy trial, vaccine recipients will be immunized at months 0, 1, 3, 6, and 12 after randomization with minimal waning expected to occur by month 18 (Gilbert et al., 2011). Hence, for an analysis of vaccine efficacy based on infection data collected through month 18, it would be safe to assume that $T \perp\!\!\!\perp V|Z$.

In Section 2.6.1, we develop a diagnostic Kolmogorov–Smirnov-type test for assessing validity of the assumption $T \perp\!\!\!\perp V|Z$, and, in Section 3.3, we demonstrate some robustness properties of the proposed inferential procedures to violation of the assumption $T \perp\!\!\!\perp V|Z$.

2.3 Estimands

We define the conditional multivariate mark-specific hazard function as

$$\lambda(t, v|Z = z) = \lim_{h_1, h_{21}, \ldots, h_{2s} \searrow 0} \frac{P(T \in [t, t + h_1), V \in \prod_{i=1}^{s} [v_i, v_i + h_{2i}) | T \geq t, Z = z)}{h_1 h_{21} \cdots h_{2s}}, \tag{2.2}$$

which is a natural generalization of the cause-specific hazard function in the presence of finitely many causes of failure (Prentice et al., 1978). GMS defined the mark-specific

vaccine efficacy as

$$VE(t,v) = 1 - \frac{\lambda(t,v|Z=1)}{\lambda(t,v|Z=0)}. \tag{2.3}$$

Also, GMS point out that $\lambda(t,v|Z=z)$ is the product of many interpretable component parameters that are not identifiable from data collected in HIV vaccine efficacy trials. Assuming homogeneous susceptibility to HIV, infectiousness, contact rates with HIV-infected individuals, mark distribution in HIV-infected contacts, a strain-specific leaky vaccine model (Halloran et al., 1992), and the fact that HIV infection is a rare event in HIV vaccine efficacy trials, $VE(t,v)$ has an approximate interpretation as the multiplicative vaccine effect to reduce susceptibility to HIV infection given exposure to a strain with mark v at time t.

2.4 Estimation method

The mark-specific hazard function factors as

$$\lambda(t,v|Z=z) = f(v|T=t,Z=z)\,\lambda(t|Z=z), \tag{2.4}$$

where $f(\cdot|T=t,Z=z)$ is the conditional density function of V given $T=t$ and $Z=z$, and $\lambda(\cdot|Z=z)$ is the ordinary marginal hazard function not involving the mark. Subsequently, the mark-specific vaccine efficacy can be written as

$$\begin{aligned} VE(t,v) &\stackrel{\text{def}}{=} 1 - \frac{\lambda(t,v|Z=1)}{\lambda(t,v|Z=0)} \\ &= 1 - \frac{f(v|T=t,Z=1)}{f(v|T=t,Z=0)}\frac{\lambda(t|Z=1)}{\lambda(t|Z=0)}. \end{aligned} \tag{2.5}$$

The factorization in (2.5) is advantageous because the two ratios can be estimated separately.

2.4.1 Density ratio model

For the mark density ratio in (2.5), we consider the semiparametric density ratio model (Qin, 1998)

$$\frac{f(v|T=t,Z=1)}{f(v|T=t,Z=0)} = g(v,\phi) \tag{2.6}$$

where $\phi \in \mathbb{R}^d$ is the parameter of interest and $g(v, \phi)$ is a known time-independent weight function, continuously differentiable in ϕ. The nuisance parameter distribution $f(v|T = t, Z = 0)$ is assumed to be time-independent (as a consequence of assuming that $T \perp\!\!\!\perp V|Z$); otherwise is treated nonparametrically. Owing to the identity in (2.1), the parameter ϕ in (2.6) is estimable using mark data for subjects with $\delta = 1$ only. Using Bayes rule, model (2.6) implies that

$$\begin{aligned}\frac{g(v_1, \phi)}{g(v_2, \phi)} &= \frac{f(v_1|Z = 1)}{f(v_1|Z = 0)} \left\{ \frac{f(v_2|Z = 1)}{f(v_2|Z = 0)} \right\}^{-1} \\ &= \frac{P(Z = 1|v_1)}{P(Z = 0|v_1)} \left\{ \frac{P(Z = 1|v_2)}{P(Z = 0|v_2)} \right\}^{-1}\end{aligned}$$

for marks $v_1, v_2 \in [0, 1]$. Thus $g(v_1, \phi)/g(v_2, \phi)$ has the interpretation as the odds of being assigned vaccine for an individual infected with strain v_1, relative to that for an individual infected with strain v_2.

The common choice of the weight function is

$$g(v, \phi) = \exp\{\alpha + \widetilde{g}(v, \beta)\} \tag{2.7}$$

where $\phi = (\alpha, \beta^T)^T$ and $\widetilde{g}(v, \beta)$ is a polynomial function. This exponential form is popular because it yields model (2.6) which is equivalent to a retrospective logistic regression model specified as $\text{logit}\{P(Z = 1|v, \delta = 1)\} = \alpha^* + \widetilde{g}(v, \beta)$ where $\alpha = \alpha^* + \log\{(1 - p_z)/p_z\}$ and $p_z = P(Z = 1|\delta = 1)$ the probability of assignment to treatment among individuals with observed failure.

If $g(v, \phi)$ satisfies (2.7), the following proposition characterizes the necessary and sufficient condition for identifiability of ϕ in (2.6).

Proposition 2.1. *In the semiparametric density ratio model*

$$\frac{f(v|T = t, Z = 1)}{f(v|T = t, Z = 0)} = \mathrm{e}^{\alpha(t) + \widetilde{g}(v, \beta)},$$

T and V are conditionally independent given Z if and only if $\alpha(t)$ is constant.

Proof. First, assume that $T \perp\!\!\!\perp V|Z$. Then $f(v|T=t,Z=0) = f(v|Z=0)$. Thus, the equality

$$1 = \int_0^1 f(v|T=t,Z=1)\mathrm{d}v = \mathrm{e}^{\alpha(t)} \int_0^1 \mathrm{e}^{\widetilde{g}(v,\beta)} f(v|T=t,Z=0)\mathrm{d}v \tag{2.8}$$

yields

$$\alpha(t) = -\log \int_0^1 \mathrm{e}^{\widetilde{g}(v,\beta)} f(v|Z=0)\mathrm{d}v =: \alpha$$

where the above integral is a positive constant for any $\beta \in \mathbb{R}^{d-1}$.

Conversely, assume that $\alpha(t) \equiv \alpha$. By (2.8),

$$\int_0^1 \mathrm{e}^{\widetilde{g}(v,\beta)} f(v|T=t,Z=0)\mathrm{d}v = \mathrm{e}^{-\alpha},$$

thus, $f(v|T=t,Z=0)$ is necessarily independent of t, that is, $T \perp\!\!\!\perp V|Z=0$. Additionally, by (2.8), $1 = \int_0^1 f(v|T=t,Z=1)\mathrm{d}v$, and therefore also $f(v|T=t,Z=1)$ is independent of t, that is, $T \perp\!\!\!\perp V|Z=1$. It implies that $T \perp\!\!\!\perp V|Z$. □

Following Qin (1998) and denoting $f(v) = f(v|T=t,Z=0)$, we consider the semiparametric log likelihood

$$l(\phi) = \sum_{i\in I} \log f(V_i) + \sum_{i\in I_1} \log g(V_i,\phi) \tag{2.9}$$

where $I = \{i : \delta_i = 1\}$ and $I_1 = \{i : \delta_i = 1 \ \wedge Z_i = 1\}$ denote the index sets of observed failures in all trial participants and the vaccine group, respectively. Thus, the likelihood comprises only information about individuals who are observed to have failed. To maximize (2.9), we employ Anderson's (1972) Lagrange multiplier method; an alternative maximization method was proposed by Prentice and Pyke (1979).

To maximize (2.9) with respect to $f(\cdot)$, it is sufficient to consider probability distributions with jumps at the mark values V_i. Denoting $f(V_i)$ by p_i, $i \in I$, the log likelihood can be written as

$$l(\phi) = \sum_{i\in I} \log p_i + \sum_{i\in I_1} \log g(V_i,\phi) \tag{2.10}$$

with the nuisance parameters p_i, $i \in I$, subject to constraints

$$p_i \in [0,1], \quad \sum_{i \in I} p_i = 1, \quad \text{and} \quad \sum_{i \in I} p_i g(V_i, \phi) = 1. \tag{2.11}$$

The last constraint reflects that $f(v)g(v,\phi)$ is a density function. In order to maximize (2.10) as a function of p_i, subject to (2.11), we consider the Lagrange function

$$H = \sum_{i \in I} \log p_i - \varphi \left(\sum_{i \in I} p_i - 1 \right) - \lambda m \sum_{i \in I} p_i \left(g(V_i, \phi) - 1 \right) \tag{2.12}$$

with the Lagrange multipliers φ and λ, and $m = \sum_{i=1}^{n} \delta_i$ the total number of observed failures. For a fixed $i \in I$, differentiation of H, at p_i, yields

$$\frac{\partial H}{\partial p_i} = \frac{1}{p_i} - \varphi - \lambda m \left(g(V_i, \phi) - 1 \right). \tag{2.13}$$

We search p_i such that $\frac{\partial H}{\partial p_i} = 0$ for all $i \in I$. It follows that

$$0 = \sum_{i \in I} p_i \frac{\partial H}{\partial p_i} = m - \varphi.$$

Thus, $\varphi = m$, and from (2.13), we finally obtain

$$p_i = \left[m(1 + \lambda \left(g(V_i, \phi) - 1 \right)) \right]^{-1}. \tag{2.14}$$

Now the vector parameter of interest ϕ can be estimated by maximizing the profile log likelihood

$$l_p(\phi, \lambda) \propto \sum_{i \in I} \left\{ - \log \left[1 + \lambda \left(g(V_i, \phi) - 1 \right) \right] + Z_i \log g(V_i, \phi) \right\} \tag{2.15}$$

obtained by replacing the nuisance parameters p_i in (2.10) by (2.14). The partial differentiation of (2.15) with respect to ϕ and λ, respectively, yields the profile score functions

$$\begin{aligned} U_\phi(\phi, \lambda) &= \frac{\partial l_p(\phi, \lambda)}{\partial \phi} = \sum_{i \in I} \left(- \frac{\lambda \dot{g}(V_i, \phi)}{1 + \lambda \left(g(V_i, \phi) - 1 \right)} + Z_i \frac{\dot{g}(V_i, \phi)}{g(V_i, \phi)} \right) \\ U_\lambda(\phi, \lambda) &= \frac{\partial l_p(\phi, \lambda)}{\partial \lambda} = - \sum_{i \in I} \frac{g(V_i, \phi) - 1}{1 + \lambda \left(g(V_i, \phi) - 1 \right)} \end{aligned} \tag{2.16}$$

where $\dot{g}(u,\phi) = \mathrm{d}g(u,\phi)/\mathrm{d}\phi$. The maximum profile likelihood estimator $(\hat{\phi}^T, \hat{\lambda})^T$ for $(\phi^T, \lambda)^T$ is defined as the solution to the system $U_\phi(\phi,\lambda) = 0$ and $U_\lambda(\phi,\lambda) = 0$. The profile score functions (2.16) are identical to those in Prentice and Pyke (1979) for $g(v,\phi) = \exp\{\alpha + \beta v\}$.

Let $m_0 = \sum_{i=1}^n \delta_i(1 - Z_i)$ and $m_1 = \sum_{i=1}^n \delta_i Z_i$ be the numbers of failures in the placebo and treatment group, respectively, and let $m = m_0 + m_1$. Denote $\rho_{mi} = m_i/m$ and assume that $\rho_{mi} \to \rho_i > 0$, $i = 0, 1$, as $m \to \infty$. If ϕ_0 denotes the true value of ϕ, then $(\hat{\phi}^T, \hat{\lambda})^T$ is a consistent estimator for $(\phi^T, \rho_1)^T$, and $\sqrt{m}(\hat{\phi} - \phi_0, \hat{\lambda} - \rho_1)$ is asymptotically normally distributed as $m \to \infty$ (Qin, 1998).

2.4.2 Proportional hazards model

For the marginal hazard ratio in (2.5), we propose to use the Cox regression model

$$\frac{\lambda(t|Z=1)}{\lambda(t|Z=0)} = \mathrm{e}^{\gamma}. \tag{2.17}$$

Considering the partial score

$$U(\gamma) = \sum_{i=1}^{n} \delta_i \left(Z_i - \frac{\sum_{k=1}^n Z_k I(X_k \geq X_i)\mathrm{e}^{\gamma Z_k}}{\sum_{k=1}^n I(X_k \geq X_i)\mathrm{e}^{\gamma Z_k}} \right),$$

or its modification based on approximate likelihoods proposed by Breslow (1974) or Efron (1977) in the presence of tied failure times, we obtain the maximum partial likelihood estimator $\hat{\gamma}$ for γ as the solution to $U(\gamma) = 0$. Alternatively, the factorization in (2.5) allows us to employ the more efficient estimation method of Lu and Tsiatis (2008) for γ by leveraging auxiliary data predictive of the failure time (estimation can be carried out by using the function `speffSurv` in the R `speff2trial` package). If additional, possibly time-dependent, covariates $Z^*(t) = (Z_1^*(t), \ldots, Z_p^*(t))^T$ are measured, then the Cox model in (2.17) can be extended to $\lambda(t|Z=1, Z^*(t) = z^*(t)) = \lambda(t|Z=0, Z^*(t)=0)\mathrm{e}^{\gamma + \gamma^{*T} z^*(t)}$, and any estimation method for $(\gamma, \gamma^{*T})^T$ available for the Cox model can be employed. For example, using $Z^*(t) = (t, Zt)^T$ specifies

$\lambda(t|Z=1,t)/\lambda(t|Z=0,t) = \mathrm{e}^{\gamma+\gamma_2^* t}$, allowing the overall treatment effect to change over time.

2.5 Asymptotic properties of the proposed estimator

Henceforth we restrict attention to the density ratio model (2.6) with weight function $g(v,\phi) = \mathrm{e}^{\alpha+\widetilde{g}(v,\beta)}$, where $\phi = (\alpha,\beta^T)^T$ and $\widetilde{g}(v,\beta)$ is a polynomial function in v, and the marginal Cox model adjusted for the covariates $(Z, Z^{*T}(t))^T$. Although Theorem 2.1 stated below applies in this described setting, for expositional simplicity the results are presented for $\widetilde{g}(v,\beta)$ a quadratic form in v and the marginal Cox model (2.17) with lone covariate the treatment group indicator Z. In this special case, the mark-specific vaccine efficacy in (2.3) takes the form $1-\mathrm{e}^{\alpha+\widetilde{g}(v,\beta)+\gamma}$. Let $\theta = (\phi^T,\lambda,\gamma)^T$ denote the vector of Euclidean parameters in this model. Let the random map $\Psi_n(\theta)$ denote the set of all estimating functions for the vector parameter θ, i.e., $\Psi_n(\theta) = (U_\phi^T(\phi,\lambda), U_\lambda(\phi,\lambda), U(\gamma))^T$. The estimator $\hat{\theta}_n$ for θ is obtained as the solution to $\Psi_n(\theta)=0$.

Next, define the random processes

$$\eta_{n,\gamma}(t) = \bar{Z}_n(t;\gamma) = \frac{\frac{1}{n}\sum_{k=1}^n Z_k I(X_k \geq t)\mathrm{e}^{\gamma Z_k}}{\frac{1}{n}\sum_{k=1}^n I(X_k \geq t)\mathrm{e}^{\gamma Z_k}} = \frac{\xi_{n,\gamma}(t)}{\zeta_{n,\gamma}(t)}, \quad t\in[0,\tau], \tag{2.18}$$

and

$$\eta_{0,\gamma}(t) = \frac{E[ZI(X\geq t)\mathrm{e}^{\gamma Z}]}{E[I(X\geq t)\mathrm{e}^{\gamma Z}]} = \frac{\xi_{0,\gamma}(t)}{\zeta_{0,\gamma}(t)}, \quad t\in[0,\tau],$$

where τ is selected such that $P(T>\tau)>0$. We will use the notation $\eta_{\cdot,\gamma}$ or $\eta_{\cdot,\theta}$ to denote the processes indexed by the component γ or by the full parameter vector θ. For $(x,\Delta,v,z)\in[0,\tau]\times\{0,1\}\times[0,1]^s\times\{0,1\}$, define the map $\varphi = (\varphi_1^T,\varphi_2,\varphi_3)^T$ as

$$\varphi(\theta,\eta_{n,\theta}) = \begin{pmatrix} \Delta\left(-\frac{\lambda\dot{g}(v,\phi)}{1+\lambda(g(v,\phi)-1)} + z\frac{\dot{g}(v,\phi)}{g(v,\phi)}\right) \\ -\Delta\,\frac{g(v,\phi)-1}{1+\lambda(g(v,\phi)-1)} \\ \Delta(z-\eta_{n,\theta}(x)) \end{pmatrix}. \tag{2.19}$$

The dependence of φ on (x, Δ, v, z) is suppressed in the notation. Let Ψ be the limit of Ψ_n as $n \to \infty$ (and also $m \to \infty$) and let $\Psi(\theta_0) = 0$. Let $\mathbb{G}_P$ denote the zero-mean P–Brownian bridge process and define the class of functions $f_{\theta,t,r}(x, z) = z^r I(x \geq t)\mathrm{e}^{\gamma z}$ for $r = 0, 1$. The following theorem characterizes the asymptotic distribution of the estimator $\sqrt{n}(\hat{\theta}_n - \theta_0)$ as $n \to \infty$ (see Appendix A for a proof).

Theorem 2.1. *Let $\hat{\theta}_n$ be the solution to $\Psi_n(\theta) = 0$ and let $\Psi(\theta_0) = 0$. Then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} -\dot{\Psi}_{\theta_0}^{-1} Z$$

where

$$Z = \mathbb{G}_P \begin{pmatrix} \varphi_1(\theta_0, \eta_{0,\theta_0}) \\ \varphi_2(\theta_0, \eta_{0,\theta_0}) \\ \varphi_3(\theta_0, \eta_{0,\theta_0}) + p_\delta l_{\theta_0} \end{pmatrix}$$

with the map $\varphi = (\varphi_1^T, \varphi_2, \varphi_3)^T$ defined in (2.19), $p_\delta = P(\delta = 1)$, and

$$l_{\theta_0} = l_{\theta_0}(x, z) = \int \zeta_{0,\theta_0}^{-1}(t) \left(f_{\theta_0,t,1}(x, z) - \eta_{0,\theta_0}(t) f_{\theta_0,t,0}(x, z) \right) \mathrm{d}F(t|\delta = 1).$$

Here $F(t|\delta = 1)$ is the conditional cumulative distribution function of X given $\delta = 1$, and $\dot{\Psi}_{\theta_0}$ is the continuously invertible derivative of the map $\theta \mapsto \Psi(\theta)$ at θ_0 and has matrix form

$$\dot{\Psi}_{\theta_0} = \begin{pmatrix} \dot{\Psi}_{11} & \dot{\Psi}_{12} & 0 \\ \dot{\Psi}_{12}^T & \dot{\Psi}_{22} & 0 \\ 0 & 0 & \dot{\Psi}_{33} \end{pmatrix}$$

with entries

$$\begin{aligned} \dot{\Psi}_{11} = \int \Delta \Bigg\{ & - \left(\frac{\rho_1}{1 + \rho_1(g(v, \phi_0) - 1)} - \frac{z}{g(v, \phi_0)} \right) \ddot{g}(v, \phi_0) \\ & + \left(\frac{\rho_1^2}{[1 + \rho_1(g(v, \phi_0) - 1)]^2} - \frac{z}{g^2(v, \phi_0)} \right) \dot{g}(v, \phi_0) \dot{g}^T(v, \phi_0) \Bigg\} \mathrm{d}P(x, \Delta, \Delta v, z) \\ \dot{\Psi}_{12} = \int & \frac{-\Delta\, \dot{g}(v, \phi_0)}{[1 + \rho_1(g(v, \phi_0) - 1)]^2} \, \mathrm{d}P(x, \Delta, \Delta v, z) \end{aligned}$$

$$\dot{\Psi}_{22} = \int \frac{\Delta\,(g(v,\phi_0)-1)^2}{[1+\rho_1(g(v,\phi_0)-1)]^2}\,\mathrm{d}P(x,\Delta,\Delta v,z)$$
$$\dot{\Psi}_{33} = \int -\Delta\,\eta_{0,\theta_0}(x)\,(1-\eta_{0,\theta_0}(x))\,\mathrm{d}P(x,\Delta,\Delta v,z)$$

where $\ddot{g}(v,\phi) = \partial^2 g(v,\phi)/\partial\phi\partial\phi^T$.

Denote the column vector of functions $\widetilde{\varphi}(\theta,\eta_{0,\theta}) = \big(\varphi_1^T(\theta,\eta_{0,\theta}), \varphi_2(\theta,\eta_{0,\theta}),\ \varphi_3(\theta,\eta_{0,\theta}) + p_\delta l_\theta\big)^T$. The following corollary describes the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ as $n \to \infty$.

Corollary 2.1. *The asymptotic random vector* $\dot{\Psi}_{\theta_0}^{-1} Z$ *in Theorem 2.1 is normally distributed with zero mean and covariance matrix* $\Gamma = \dot{\Psi}_{\theta_0}^{-1}\Omega\dot{\Psi}_{\theta_0}^{-1}$ *where*

$$\Omega = P\widetilde{\varphi}(\theta_0,\eta_{0,\theta_0})\widetilde{\varphi}^T(\theta_0,\eta_{0,\theta_0}) - P\widetilde{\varphi}(\theta_0,\eta_{0,\theta_0})\,P\widetilde{\varphi}^T(\theta_0,\eta_{0,\theta_0}).$$

Let $\widehat{\Gamma}_n$ denote the empirical estimator for Γ obtained by replacing P by the empirical probability measure $\mathbb{P}_n$, θ_0 by $\hat{\theta}_n$, and η_{0,θ_0} by $\eta_{n,\hat{\theta}_n}$ in the definition of Γ. Corollary 2.1 leads to the construction of Wald confidence intervals (pointwise in v) for the components of θ, and, subsequently, for the parameter $VE(v) = 1 - \mathrm{e}^{\alpha+\widetilde{g}(v,\beta)+\gamma}$.

2.6 Hypothesis testing

For illustrating the proposed testing procedures, we consider the simplified weight function $g(v,\phi) = \mathrm{e}^{\alpha+\beta^T v}$, $\phi = (\alpha,\beta^T)^T$, which leads to the mark-specific vaccine efficacy function $VE(v) = 1 - \mathrm{e}^{\alpha+\beta^T v+\gamma}$. We develop likelihood ratio and Wald tests to evaluate the null hypothesis

$$H_0^0 : VE(v) = 0 \text{ for all } v \in [0,1]^s, \tag{2.20}$$

which states that the vaccine provides no protection against infection with any HIV strain. If H_0^0 is rejected, the question arises as to whether vaccine efficacy depends on the viral divergence; thus, we develop likelihood ratio and Wald tests for the null hypothesis

$$H_0 : VE(v) \equiv VE \text{ for all } v \in [0,1]^s. \tag{2.21}$$

Under models (2.6) and (2.17), the null hypothesis H_0^0 is equivalent to $\{\widetilde{H}_0^0 : \beta = 0 \text{ and } \gamma = 0\}$. The likelihood ratio test of $\widetilde{H}_0^0$ against the alternative hypothesis $\{\widetilde{H}_1^0 : \beta \neq 0 \text{ or } \gamma \neq 0\}$ uses Simes' procedure (Simes, 1986), in which the profile likelihood ratio test statistic for β in model (2.6) and the partial likelihood ratio test statistic for γ in model (2.17) are evaluated separately. P-values p_β and p_γ are obtained based on the fact that the likelihood ratio statistics are asymptotically χ_s^2 and χ_1^2 under $\widetilde{H}_0^0$, respectively. Simes' procedure rejects H_0^0 if either $\max(p_\beta, p_\gamma) \leq \alpha$ or $\min(p_\beta, p_\gamma) \leq \alpha/2$ where α is the nominal familywise level of significance. The Wald test of H_0^0 versus H_1^0 is based on the statistic $n(\hat{\beta}_n^T, \hat{\gamma}_n)\widehat{\Gamma}_{n,\beta\gamma}^{-1}(\hat{\beta}_n^T, \hat{\gamma}_n)^T$ where $\widehat{\Gamma}_{n,\beta\gamma}$ is the submatrix of $\widehat{\Gamma}_n$ pertaining to the components β and γ. Under H_0^0, the Wald test statistic is asymptotically χ_{s+1}^2. We additionally propose a weighted one-sided Wald-type test of H_0^0 based on the Z-statistic

$$Z = \frac{\sum_{i=1}^{s} \frac{\hat{\beta}_{n,i}}{\widehat{\text{var}}\,\hat{\beta}_{n,i}} - \frac{\hat{\gamma}_n}{\widehat{\text{var}}\,\hat{\gamma}_n}}{\sqrt{\widehat{\text{var}}\left(\sum_{i=1}^{s} \frac{\hat{\beta}_{n,i}}{\widehat{\text{var}}\,\hat{\beta}_{n,i}} - \frac{\hat{\gamma}_n}{\widehat{\text{var}}\,\hat{\gamma}_n}\right)}}, \tag{2.22}$$

which is designed to increase power to detect alternative hypotheses where both the marginal vaccine efficacy $VE = 1 - \frac{\lambda(t|Z=1)}{\lambda(t|Z=0)} = 1 - e^{\gamma} > 0$ and $VE(v)$ declines with all of the components of V (we refer to the latter property as the sieve effect). Under H_0^0, the test statistic (2.22) is $\mathsf{N}(0,1)$. The null hypothesis H_0 is equivalent to $\{\widetilde{H}_0 : \beta = 0\}$, and thus the density ratio model (2.6) alone serves to construct the likelihood ratio and Wald test statistics for $\widetilde{H}_0$.

2.6.1 Diagnostic test for $T \perp\!\!\!\perp V|Z$

By Proposition 2.1, the conditional independence between the failure time and the mark variables given treatment assignment is a necessary assumption for parameter identifiability in the time-independent density ratio model. We propose a diagnostic

test of the null hypothesis $\{\widetilde{K}_0 : T \perp\!\!\!\perp V\}$ based on the statistic

$$\sup_{t,v} \left|\hat{F}_{TV}(t,v) - \hat{F}_T(t)\hat{F}_V(v)\right|, \tag{2.23}$$

where $\hat{F}_{TV}(t,v)$ is the nonparametric maximum likelihood estimator of the joint distribution function of (T,V) developed by Huang and Louis (1998), $\hat{F}_T(t)$ is one minus the Kaplan-Meier estimator of the survival function of T, and $\hat{F}_V(v)$ is the empirical distribution function of the observed values of V. The estimator $\hat{F}_V(v)$ is justified because the distribution of $V|Z$ is identical to that of $V|(\delta = 1, Z)$ under the assumptions introduced in Section 2.2. The critical values for the distribution of (2.23) under $\widetilde{K}_0$ can be assessed using a bootstrap algorithm as follows:

1. Draw an independent sample (X_i^*, δ_i^*), $i = 1, \ldots, n$, from the original time-on-study data (X_i, δ_i), $i = 1, \ldots, n$, with replacement.
2. Independently of Step 1, draw a sample V_i^*, $i \in \{k : \delta_k^* = 1\}$, from the original mark data V_i, $i \in \{l : \delta_l = 1\}$, with replacement.
3. Compute the value of the test statistic based on the bootstrap data $(X_i^*, \delta_i^*, \delta_i^* V_i^*)$, $i = 1 \ldots, n$.
4. Repeat Steps (1)–(3) B times.
5. Estimate the α–quantile of the null distribution of (2.23) by the empirical α–quantile of the replicated values of the test statistic obtained in Steps (1)–(4).

The test of the overall null hypothesis $\{K_0 : T \perp\!\!\!\perp V|Z\}$ is based on Simes' procedure applied to the tests of $\widetilde{K}_0$ performed separately for the two groups $Z = 1$ and $Z = 0$.

2.7 Conclusions

The proposed methods provide a tool to conduct sieve analysis which grants insight into how vaccine effects depend on viral divergence. The parametric component in the density ratio model can result in the method's greater efficiency compared to

alternative approaches. The tradeoff for the efficiency gain is the addition of the $T \perp\!\!\!\perp V|Z$ assumption which, however, is testable and, as Section 3.3 suggests, the method is largely robust to its violation. For successful interpretation of sieve analysis results, the method requires a scientifically meaningful definition of the sequence distance. It is advised to focus the distance on sequence regions that may constitute immunogenic antibody epitopes in order to increase power to detect a potential sieve effect.

Chapter 3

SIMULATION STUDY OF THE MAXIMUM LIKELIHOOD ESTIMATOR FOR MARK-SPECIFIC VACCINE EFFICACY UNDER COMPLETE DATA

3.1 Introduction

Consider the density ratio model (2.6) with the weight function $g(v, \phi) = e^{\alpha+\beta^T v}$ where $\phi = (\alpha, \beta^T)^T$. Let $(\hat{\alpha}, \hat{\beta}^T)^T$ denote the estimator for $(\alpha, \beta^T)^T$ that maximizes the log profile likelihood (2.15). Also, consider the Cox regression model (2.17) for the marginal hazard ratio and let $\hat{\gamma}$ denote the maximum partial likelihood estimator for the log hazard ratio. Consequently, the mark-specific vaccine efficacy function takes the form $VE(v) = 1 - e^{\alpha+\beta^T v+\gamma}$. In this chapter we present a simulation study of finite-sample properties of the estimator $\widehat{VE}(v) = 1 - e^{\hat{\alpha}+\hat{\beta}^T v+\hat{\gamma}}$ under both validity and violation of the model assumptions.

We investigate finite-sample bias, standard errors of $\widehat{VE}(v)$, and coverage probability of Wald pointwise confidence intervals for $VE(v)$ in univariate and bivariate mark settings. Furthermore, we examine size and power of the Wald and likelihood ratio tests of H_0^0 in (2.20) and H_0 in (2.21), and compare them to alternative tests in Gilbert, McKeague, and Sun (2008) (henceforth GMS). To allow a data analyst to explore the validity of the $T \perp\!\!\!\perp V|Z$ assumption, we additionally evaluate size and power of the proposed diagnostic test in (2.23).

3.2 Assessment of the proposed methods under model validity

3.2.1 Data generation

The simulation setup aims to mimic 3-year Phase IIb and Phase III two-arm placebo-controlled HIV vaccine efficacy trials. We specify the failure times T to be exponential with rates λ and λe^{γ} in the placebo and vaccine group, respectively, with e^{γ} the marginal hazard ratio. The rate $\lambda = \log(0.85)/(-3)$ is chosen so that the 0.15 quantile of the failure time distribution in the placebo group equals 3 years. We specify the censoring times C to be $\mathsf{Uniform}(0, 15)$ in each group which implies a 20% chance of censoring by 3 years. The observed time on study X is defined as $\min(T, C, 3)$, i.e., the minimum of the times to infection, random censoring, and administrative censoring at 3 years. The vaccine-to-placebo assignment ratio is 1:1. Let Z denote the vaccine group indicator.

We consider a continuous mark variable V, conditionally independent of T given Z, with the support of each component taken to be $[0, 1]$. A univariate mark V for placebo and vaccine recipients is generated from distributions with density functions

$$f(v|Z = 0) = \frac{2e^{-2v}}{1 - e^{-2}} I(0 \leq v \leq 1) \tag{3.1}$$

and

$$f(v|Z = 1) = f(v|Z = 0)e^{\alpha+\beta v}, \tag{3.2}$$

respectively, where, for a given value of β, the value of the parameter $\alpha = \alpha(\beta)$ is defined as the solution to $\int_0^1 f(v|Z = 0)e^{\alpha+\beta v}dv = 1$. The distribution (3.1) is the exponential distribution with rate 2 standardized to the support $[0, 1]$. The distribution (3.2) is chosen to preserve the density ratio model. In simulation scenarios involving a bivariate mark $V = (V_1, V_2)^T$, the components V_1 and V_2 are generated as independent univariate marks from the distribution (3.1) for infected placebo recipients and (3.2) for infected vaccine recipients, the latter requiring to prespecify values β_1 and β_2 for the components V_1 and V_2, respectively. For both univariate and bivariate

marks, only the mark values for subjects with $T \leq 3$ and $\delta = 1$ or their subset would be observed in a real study and hence are used in the analysis.

We consider the sample sizes $N = 1481, 741,$ and 556 per arm so that the expected numbers of observed placebo infections by year 3 are $N_{pI} = 200, 100,$ and 75, respectively. The sample sizes N are calculated based on the relationship

$$\begin{aligned} N_{pI} &= N \times P(\delta = 1 | Z = 0) \\ &= N \times P(T \leq \min(C, 3) | Z = 0) \\ &= N \times \left[P(T \leq C, C < 3 | Z = 0) + P(T \leq 3, C \geq 3 | Z = 0) \right]. \end{aligned}$$

3.2.2 Specification of model parameters

Simulation scenarios with univariate marks are characterized by the following model parameter values:

(M1): $(\beta, \gamma) = (0, 0)$ where $VE(v) = 0$;

(M2): $(\beta, \gamma) = (0.3, -0.3)$ where $VE(v)$ decreases, $VE(0) = 0.3$, and $VE(1) = 0.1$;

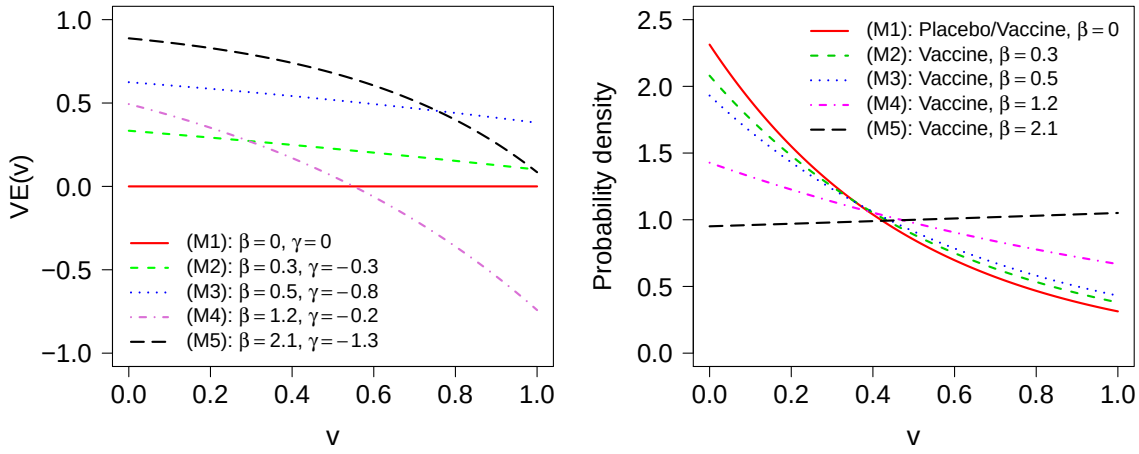
(M3): $(\beta, \gamma) = (0.5, -0.8)$ where $VE(v)$ decreases, $VE(0) = 0.6$, and $VE(1) = 0.4$;

(M4): $(\beta, \gamma) = (1.2, -0.2)$ where $VE(v)$ decreases, $VE(0) = 0.5$, and $VE(1) = -0.7$;

(M5): $(\beta, \gamma) = (2.1, -1.3)$ where $VE(v)$ decreases, $VE(0) = 0.9$, and $VE(1) = 0.1$.

Model (M4) represents a scenario with $\int_0^1 VE(v)\mathrm{d}v = 0$. Thus, in this case, for sufficiently large mark values, $VE(v)$ takes on negative values. From the immunological perspective, antibody-dependent enhancement of the infection risk (Mascola et al., 1993) is a phenomenon that may give rise to negative values of $VE(v)$. For scenarios (M1)–(M5), the corresponding mark-specific vaccine efficacy functions are depicted in Figure 3.1a. The density functions used to generate the mark values for observed infections in the placebo and vaccine group are displayed in Figure 3.1b.

Additionally, we investigate mark-specific vaccine efficacy models with bivariate marks which are characterized by the following parameter specifications:



(a) (b)

Figure 3.1: Simulation scenarios with a univariate mark variable: (a) mark-specific vaccine efficacy functions $VE(v) = 1 - \mathrm{e}^{\alpha+\beta v+\gamma}$, (b) density functions of the mark variable in the placebo/vaccine group.

(M6): $(\beta_1, \beta_2, \gamma) = (0, 0, 0)$ where $VE(v_1, v_2) = 0$;

(M7): $(\beta_1, \beta_2, \gamma) = (0.3, 0.2, -0.4)$ where $VE(v_1, v_2)$ decreases;

(M8): $(\beta_1, \beta_2, \gamma) = (0.5, 0.1, -0.4)$ where $VE(v_1, v_2)$ decreases;

(M9): $(\beta_1, \beta_2, \gamma) = (0.3, 0.2, -0.1)$ where $VE(v_1, v_2)$ decreases;

(M10): $(\beta_1, \beta_2, \gamma) = (1, 0.4, -0.9)$ where $VE(v_1, v_2)$ decreases.

Model (M9) represents a scenario where $\int_0^1\int_0^1 VE(v_1, v_2)\mathrm{d}v_1\mathrm{d}v_2 = 0$. In this case, mark values (v_1, v_2) that are sufficiently close to the point $(1, 1)$ yield negative values of $VE(v_1, v_2)$ which may occur, for example, in the presence of certain vaccine-induced enhancing antibodies as mentioned in the univariate case. The pertaining mark-specific vaccine efficacy functions $VE(v_1, v_2)$ are plotted in Figure 3.2. Table 3.1 summarizes the expected numbers of vaccine infections in scenarios (M1)–(M10).

Table 3.1: Expected numbers of placebo (N_{pI}) and vaccine (N_{vI}) infections in simulation scenarios (M1)–(M10)

	N_{vI}									
N_{pI}	(M1)	(M2)	(M3)	(M4)	(M5)	(M6)	(M7)	(M8)	(M9)	(M10)
200	200	152	94	167	58	200	138	138	183	85
100	100	76	47	83	29	100	69	69	91	43
75	75	57	35	63	22	75	52	52	69	32

3.2.3 Computer algorithm

The simulation was programmed in R 2.14.0. The Newton-Raphson method was used to find the root of the profile score equations (2.16). The procedure was insensitive to the choice of starting values; eventually the value 0 was used as a starting value for each component of the vector parameter ϕ. The strict concavity of the log profile likelihood (2.15) was checked graphically (see Figure 3.3).

3.2.4 Evaluated test statistics

For each simulation scenario, characterized by the choice of β, γ, and the sample size N, we generated data sets independently 1000 times to evaluate the performance of the proposed estimation and testing procedures for $VE(v)$. For the estimator $\widehat{VE}(v)$, we computed finite-sample bias, standard error measures, and coverage probabilities of Wald pointwise confidence intervals. For each set of replications, the bias is defined as the mean of the individual vaccine efficacy estimates minus the true value. The asymptotic standard error estimate, calculated as the mean of the replicated standard error estimates, is compared to the empirical standard error estimate defined as the sample standard deviation of the individual vaccine efficacy estimates.

We examined size and power of the Wald and likelihood ratio tests of H_0^0 and H_0.

The weighted Wald-type test (2.22) of H_0^0 is designed to be sensitive to alternative hypotheses with positive marginal vaccine efficacy and $VE(v)$ declining in v, for which it is conjectured to be more powerful than the standard log-rank test that assesses the marginal vaccine efficacy only and is efficient in this setting among tests that ignore the mark. Thus, size and power of the two-sided log-rank test is included for the purpose of comparison. Additionally, it is of interest to compare power of the proposed tests of H_0^0 and H_0 to that using the test procedures of GMS. In particular, we explored power of the GMS's tests based on the statistics denoted as $\hat{U}_3^1$ and $\hat{U}_4^1$ for two-sided testing of H_0^0 and $\hat{U}_2^{np}$ and $\hat{U}_2^{sp}$ for two-sided testing of H_0. The GMS's testing procedures were implemented by the authors in the R `cmprskContin` package. Finally, we investigated size and power of the Kolmogorov–Smirnov-type test (2.23) of the null hypothesis $\{T \perp\!\!\!\perp V|Z\}$.

Throughout the simulation study the nominal level of significance was taken to be 5% for each two-sided and 2.5% for each one-sided test. The coverage probability was computed as the proportion of 95% Wald-based confidence intervals which cover the true value $VE(v)$ for a fixed v.

3.2.5 Simulation results

Estimation: bias, standard errors, and coverage properties of confidence intervals

For a univariate mark, Figures 3.4–3.6 characterize finite-sample bias of $\widehat{VE}(v)$, asymptotic standard error estimates for $\widehat{VE}(v)$, and coverage probabilities of 95% pointwise Wald confidence intervals for $VE(v)$. In each simulation scenario, bias is the smallest in the central support region of the mark and slightly increases in the right tail of the mark distribution. As expected, bias decreases with a growing number of failures. The asymptotic and empirical standard error estimates are consistently in good accordance with one another. The coverage probabilities for all mark values approach the nominal confidence level with an increasing number of failures. For a bi-

variate mark, finite-sample performance of estimation of $VE(v_1, v_2)$ is summarized in Table 3.2 for the mark values $(v_1, v_2) = (0.1, 0.1)$, $(0.5, 0.5)$, $(0.5, 0.9)$, $(0.9, 0.5)$, and $(0.9, 0.9)$. The point and interval estimators for $VE(v_1, v_2)$ exhibit similar properties as in scenarios with a univariate mark.

Hypothesis tests: size and power

Tables 3.3 and 3.4 summarize the observed size and power of the Wald and likelihood ratio tests of H_0^0 and H_0 in scenarios with a univariate mark, and include a comparison with the alternative non- and semiparametric tests of GMS. Size of all considered tests is in good agreement with the nominal significance level. As for two-sided testing of H_0^0, power of the Wald and likelihood ratio test is comparable, and both tests outperform the $\hat{U}_3^1$ and $\hat{U}_4^1$ tests of GMS, with power gains up to 19%, for all specified values of the model parameters except for the scenario $(\beta, \gamma) = (0.3, -0.3)$. Verifying our conjecture, power of the weighted Wald-type test is uniformly higher than that of the log-rank test, with power gain reaching up to 27%.

Power of the Wald and likelihood ratio tests of H_0 is considerably higher than that of GMS's $\hat{U}_2^{np}$ and $\hat{U}_2^{sp}$ tests (Table 3.4). The power gain is as large as 36% compared to the more powerful of the $\hat{U}_2^{np}$ and $\hat{U}_2^{sp}$ tests. We conjecture that the power gain can be ascribed to the additional assumptions made by our methods compared to those of GMS. In the bivariate mark setting (Tables 3.5 and 3.6), the likelihood ratio tests of H_0^0 and H_0 are slightly more powerful than the respective Wald tests. Furthermore, the weighted Wald-type test attains higher power than the log-rank test, but the power gain is moderate.

Tables 3.7 and 3.8 suggest that the size of the diagnostic test (2.23) of the null hypothesis $\{T \perp\!\!\!\perp V | Z\}$ has nominal significance level. When performed in each treatment group separately, a similar performance is observed.

3.3 Robustness analysis of the proposed methods under model misspecification

3.3.1 Data generation

In this section, assuming the model $VE(v) = 1 - \mathrm{e}^{\alpha+\beta v+\gamma}$, $v \in [0,1]$, we examine robustness of the vaccine efficacy estimator and the Wald and likelihood ratio tests of H_0^0 and H_0 to violation of the $T \perp\!\!\!\perp V|Z$ assumption. To this end, we generate samples (T, V) with correlation ρ_{TV}. For an infected placebo recipient, we first generate $T = t$, and second generate V from the conditional distribution with density function

$$f(v|T = t, Z = 0; c) = \frac{c\, t\, \mathrm{e}^{-ctv}}{1 - \mathrm{e}^{-ct}}\, I(0 \le v \le 1), \quad c > 0,$$

where the parameter c governs the magnitude of ρ_{TV}. Using Proposition 2.1, for an infected vaccine recipient with $T = t$, we generate V from the conditional distribution with the density function $f(v|T = t, Z = 1; c) = f(v|T = t, Z = 0; c)\, \mathrm{e}^{\alpha(t)+\beta v}$, where $\alpha(t) = \alpha(t, \beta, c)$ is the solution to $\int_0^1 f(v|T = t, Z = 0; c)\, \mathrm{e}^{\alpha(t)+\beta v}\mathrm{d}v = 1$. The data generating mechanism implies that the true mark-specific vaccine efficacy is $VE(t, v) = 1 - \mathrm{e}^{\alpha(t)+\beta v+\gamma}$. The values of c are chosen such that ρ_{TV} varies over the 0.1–0.5 range. In each scenario, power to reject the null hypothesis $\{T \perp\!\!\!\perp V|Z\}$ using the diagnostic test (2.23) is evaluated.

3.3.2 Simulation results

Estimation: average deviation

Figure 3.7 depicts the Monte Carlo average deviation $\widehat{VE}(v) - \overline{VE}(v)$ where $\overline{VE}(v) = \int_0^3 VE(t, v) f(t)\mathrm{d}t \big/ \int_0^3 f(t)\mathrm{d}t$ and $f(t)$ is the density function of the failure time T. In words, $\overline{VE}(v)$ is the mean $VE(v)$ for observable mark data, i.e., marks V with corresponding failure times $T \le 3$ (see the simulation setup in Section 3.2.1). For scenarios where the magnitude of correlation is detected with moderate power, the deviation is minimal. Such scenarios are of particular interest because, in these cases, violation

of the assumption $T \perp\!\!\!\perp V|Z$ may frequently remain undetected. For scenarios with power to detect departures from $T \perp\!\!\!\perp V|Z$ greater than 90%, the deviation tends to increase for mark values in the right tail of the distribution, although this is fairly innocuous because these are the scenarios where the method would fail the diagnostics and hence would not be used.

Hypothesis tests: size and power

Tables 3.9 and 3.10 indicate that the likelihood ratio and Wald tests of H_0^0 and H_0 retain the correct size under violation of the $T \perp\!\!\!\perp V|Z$ assumption. This robustness property holds for any violation of this assumption, because under either null hypothesis H_0^0 and H_0, $\alpha(t) \equiv \alpha = 0$ for all $t > 0$, i.e., the estimand $VE(v)$ in the time-independent model coincides with the true $VE(t, v)$ characterizing the data generating mechanism regardless of the value of ρ_{TV}. In addition, Tables 3.9 and 3.10 show that power of the likelihood ratio and Wald tests is minimally sensitive to correlation between T and V and tends to be higher than that of GMS's tests.

Non-proportional hazard rates

Finally, we performed a robustness analysis with respect to violation of the proportional marginal hazards assumption in which we considered distributions of $T|Z = 0$ and $T|Z = 1$ characterized by the hazard functions $\lambda(t|Z = 0) \equiv \lambda$ and $\lambda(t|Z = 1) = \lambda e^{\gamma_0+\gamma_1 t}$; thus γ_1 governs the amount of variation in the marginal hazard ratio over time. We observed minimal differences in the patterns of average deviation $\widehat{VE}(v) - \overline{VE}(v)$, standard error estimates, and power relative to those presented for correctly specified models (thus results not reported). We conjecture that the similar performance occurs because $VE(t, v) = 1 - e^{\alpha+\beta v+\gamma_0+\gamma_1 t}$ does not include an interaction of V with T, and hence $VE(t, v)$ and $VE(v)$ have the same amount of change in v.

3.4 Conclusions

The simulation study demonstrates that the proposed estimation and hypothesis testing methods perform well in finite samples. The Wald and likelihood ratio tests of H_0^0 and H_0 tend to be more powerful, in certain instances with considerable power gains, than the GMS's tests. We conjecture that the power gain is achieved owing to additional parametric assumptions in our approach as compared to those in GMS. The weighted Wald-type test of H_0^0 detects one-sided alternatives of greatest scientific relevance with a larger power than the log-rank test, particularly in cases when $VE(v)$ declines in v yet the marginal vaccine efficacy equals zero. Furthermore, size of the Wald and likelihood ratio tests is robust to correlation between the failure time and a mark whereas power may be slightly sensitive to levels of correlation that lead to rejection of $T \perp\!\!\!\perp V|Z$ with high power when the infection rates are low. The diagnostic test of the null hypothesis $\{T \perp\!\!\!\perp V|Z\}$, with critical values assessed using the bootstrap method, has adequate size.

3.5 Tables and figures

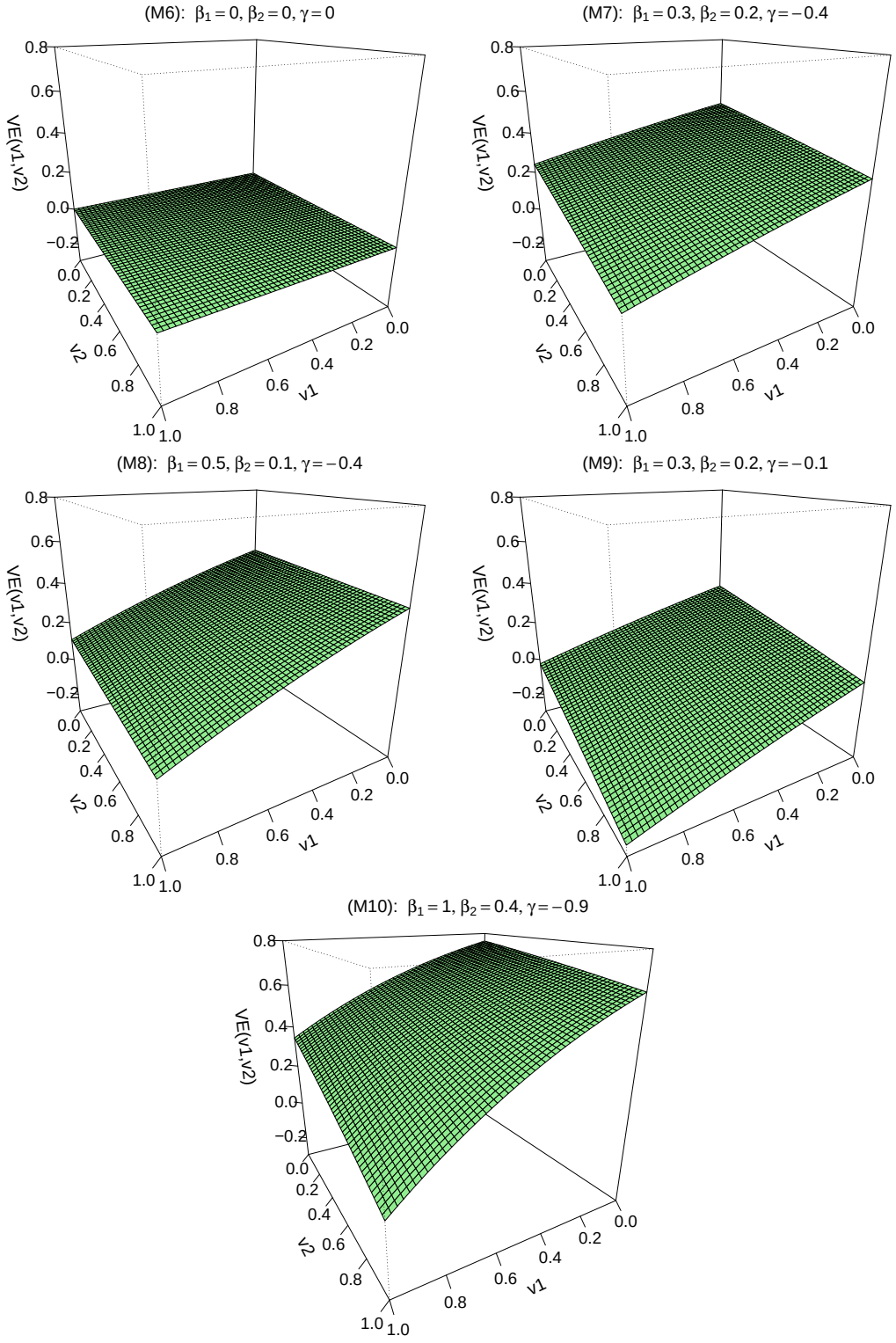


Figure 3.2: Mark-specific vaccine efficacy functions in simulation scenarios (M6)–(M10)

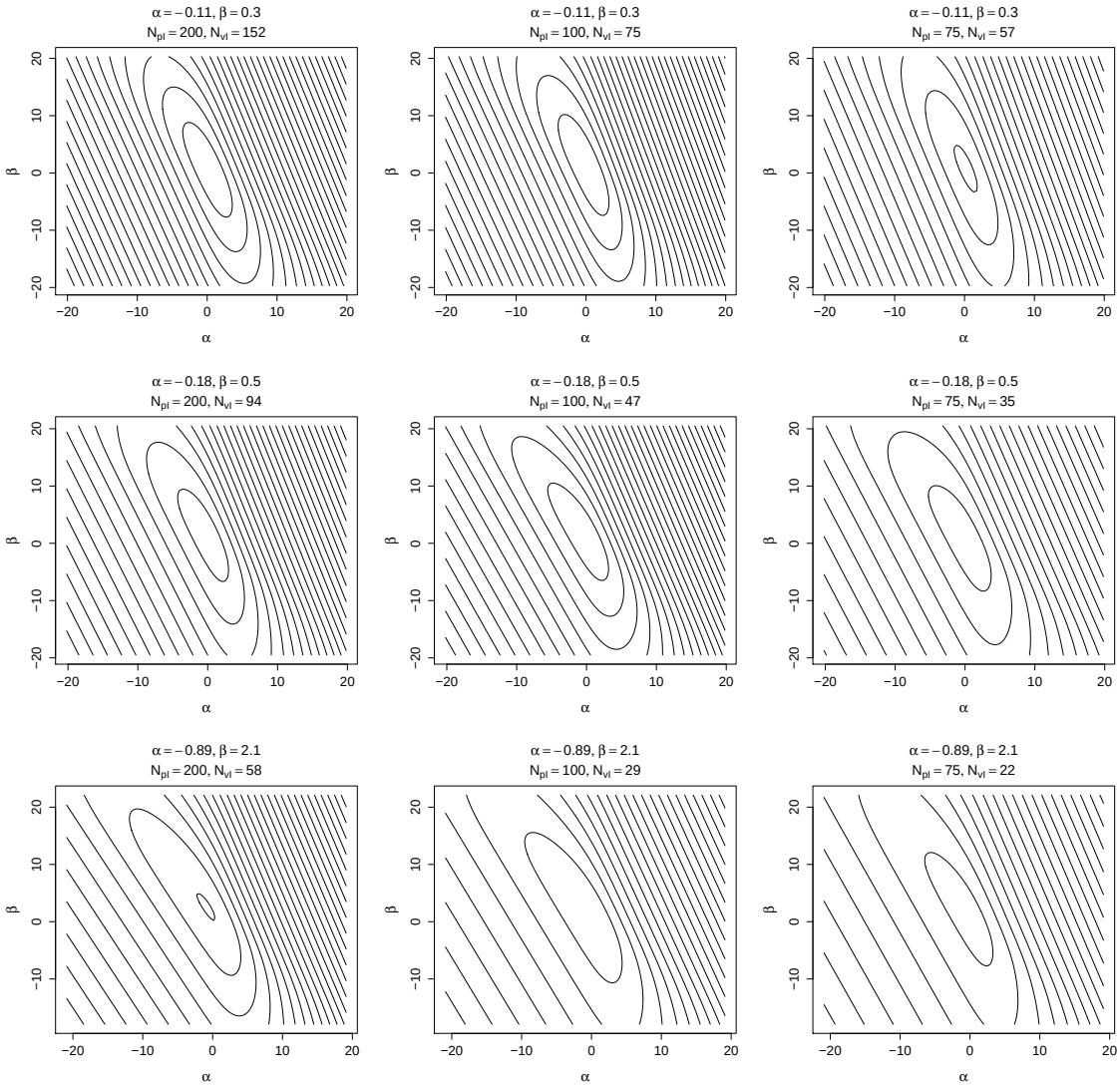


Figure 3.3: Contour plots of the log profile likelihood surface $l_p(\alpha, \beta)$ in (2.15) for data sets generated from density ratio models with univariate marks. N_{pI} and N_{vI} denote the expected numbers of observed infections in the placebo and vaccine group, respectively.

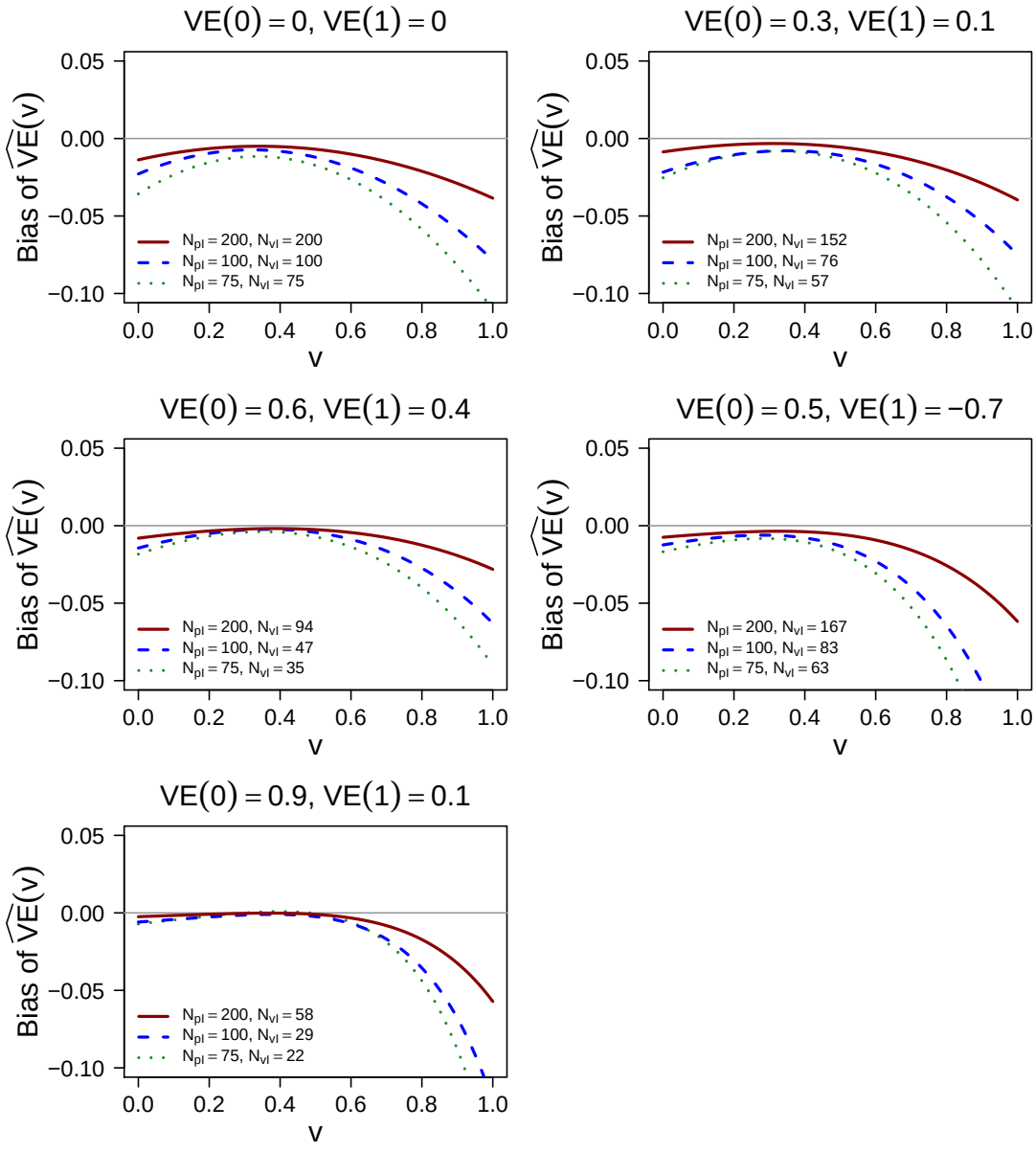


Figure 3.4: Finite-sample bias of estimation of $VE(v) = 1 - \mathrm{e}^{\alpha+\beta v+\gamma}$ in simulation scenarios (M1)–(M5) with a univariate mark. N_{pI} and N_{vI} are the expected numbers of observed placebo and vaccine infections, respectively.

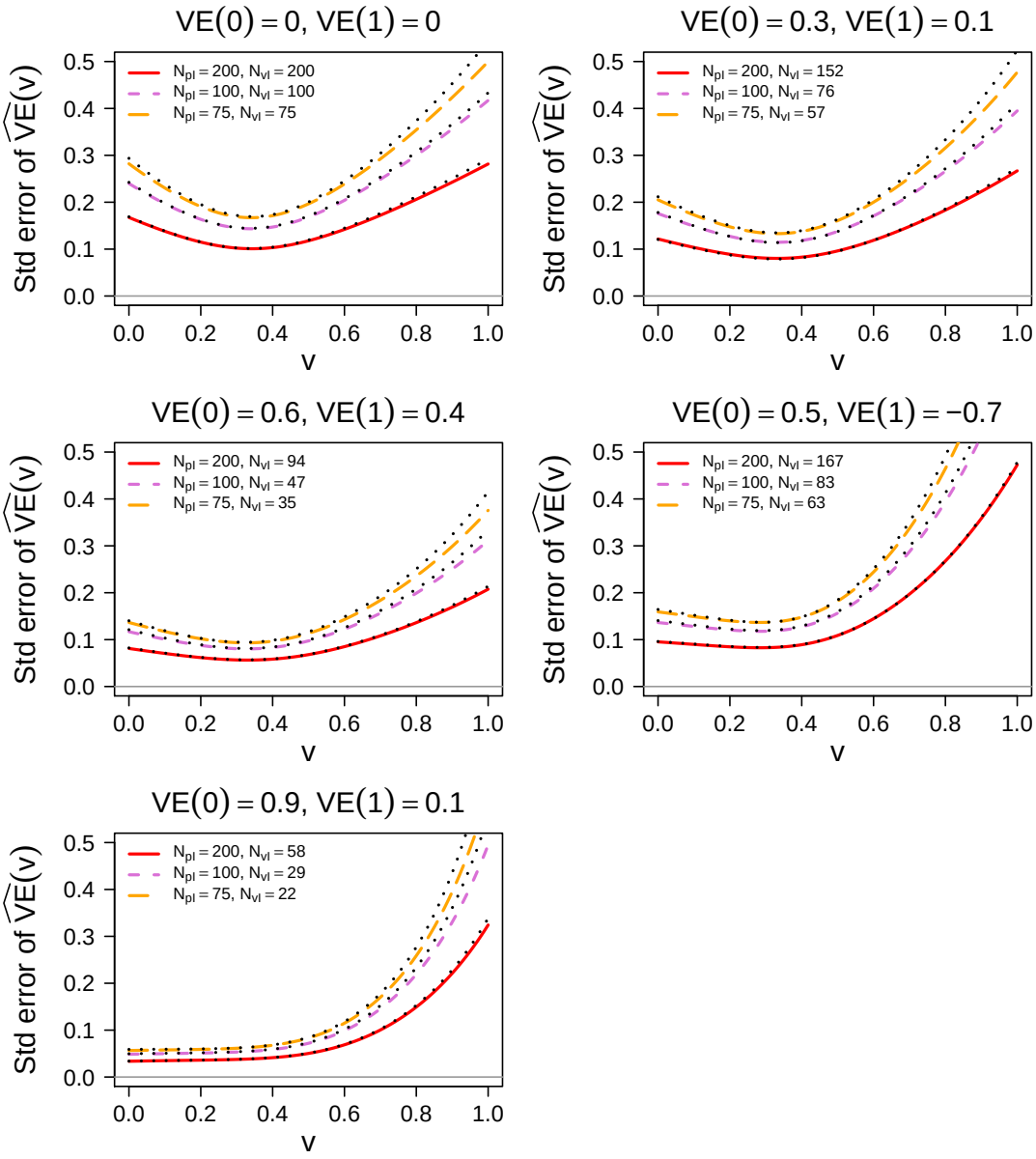


Figure 3.5: Asymptotic versus empirical (black dotted) standard error estimates of $\widehat{VE}(v)$ in simulation scenarios (M1)–(M5) with a univariate mark. N_{pI} and N_{vI} are the expected numbers of observed placebo and vaccine infections, respectively.

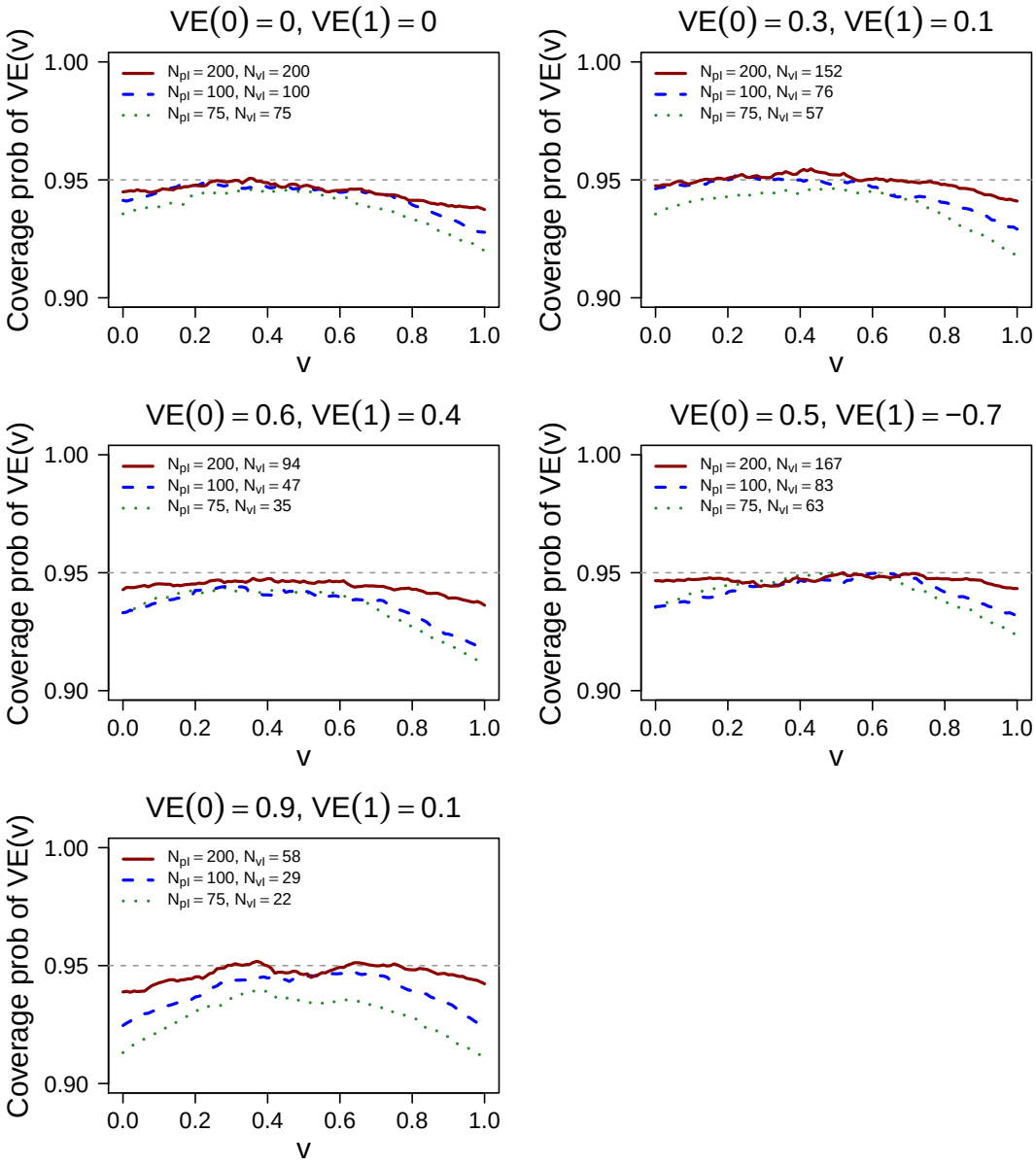


Figure 3.6: Coverage probabilities of 95% pointwise Wald confidence intervals for $VE(v)$ in simulation scenarios (M1)–(M5). N_{pI} and N_{vI} are the expected numbers of observed placebo and vaccine infections, respectively.

Table 3.2: Bias and standard error estimates for $\widehat{VE}(v_1, v_2)$, and coverage probabilities of 95% Wald confidence intervals for $VE(v_1, v_2)$ in simulation scenarios (M6)–(M10)

Model	N_{pI}	N_{vI}	v_1	v_2	Bias	Mean SE	Emp SE	p_{cover}
(M6)	200	200	0.1	0.1	−0.019	0.17	0.17	0.94
			0.5	0.5	−0.012	0.13	0.13	0.95
			0.5	0.9	−0.037	0.25	0.26	0.95
			0.9	0.5	−0.028	0.25	0.24	0.95
			0.9	0.9	−0.053	0.34	0.34	0.93
	100	100	0.1	0.1	−0.032	0.24	0.26	0.94
			0.5	0.5	−0.017	0.19	0.20	0.94
			0.5	0.9	−0.061	0.37	0.38	0.93
			0.9	0.5	−0.061	0.37	0.40	0.94
			0.9	0.9	−0.104	0.50	0.54	0.92
	75	75	0.1	0.1	−0.040	0.28	0.30	0.95
			0.5	0.5	−0.015	0.22	0.23	0.94
			0.5	0.9	−0.079	0.44	0.48	0.92
			0.9	0.5	−0.072	0.43	0.46	0.93
			0.9	0.9	−0.137	0.60	0.68	0.91
(M7)	200	138	0.1	0.1	−0.007	0.11	0.11	0.94
			0.5	0.5	−0.006	0.10	0.10	0.95
			0.5	0.9	−0.030	0.21	0.22	0.95
			0.9	0.5	−0.026	0.22	0.22	0.94
			0.9	0.9	−0.055	0.32	0.33	0.94
	100	69	0.1	0.1	−0.026	0.16	0.17	0.95
			0.5	0.5	−0.005	0.15	0.15	0.93
			0.5	0.9	−0.042	0.31	0.34	0.92
			0.9	0.5	−0.043	0.32	0.34	0.92
			0.9	0.9	−0.088	0.48	0.53	0.91
	75	52	0.1	0.1	−0.027	0.19	0.20	0.94
			0.5	0.5	−0.015	0.17	0.18	0.94
			0.5	0.9	−0.078	0.38	0.41	0.91
			0.9	0.5	−0.072	0.39	0.41	0.92
			0.9	0.9	−0.155	0.60	0.66	0.88
(M8)	200	138	0.1	0.1	−0.001	0.11	0.11	0.94
			0.5	0.5	−0.009	0.10	0.11	0.94
			0.5	0.9	−0.043	0.21	0.22	0.94
			0.9	0.5	−0.035	0.24	0.25	0.94
			0.9	0.9	−0.079	0.35	0.36	0.93
	100	69	0.1	0.1	−0.014	0.16	0.16	0.94
			0.5	0.5	−0.015	0.15	0.16	0.94
			0.5	0.9	−0.059	0.31	0.33	0.93
			0.9	0.5	−0.074	0.36	0.37	0.93
			0.9	0.9	−0.134	0.52	0.55	0.92

Continued on the next page

Table 3.2 continued

Model	N_{pI}	N_{vI}	v_1	v_2	Bias	Mean SE	Emp SE	p_{cover}
	75	52	0.1	0.1	−0.014	0.18	0.18	0.93
			0.5	0.5	−0.017	0.18	0.18	0.95
			0.5	0.9	−0.079	0.37	0.40	0.92
			0.9	0.5	−0.100	0.43	0.46	0.94
			0.9	0.9	−0.187	0.64	0.74	0.92
(M9)	200	183	0.1	0.1	−0.013	0.14	0.14	0.96
			0.5	0.5	−0.009	0.13	0.14	0.93
			0.5	0.9	−0.029	0.27	0.27	0.94
			0.9	0.5	−0.036	0.28	0.30	0.93
			0.9	0.9	−0.061	0.41	0.42	0.94
	100	91	0.1	0.1	−0.024	0.20	0.21	0.95
			0.5	0.5	−0.014	0.19	0.20	0.94
			0.5	0.9	−0.068	0.40	0.44	0.93
			0.9	0.5	−0.068	0.41	0.45	0.93
			0.9	0.9	−0.138	0.62	0.73	0.91
	75	69	0.1	0.1	−0.035	0.24	0.24	0.94
			0.5	0.5	−0.023	0.22	0.24	0.93
			0.5	0.9	−0.081	0.47	0.55	0.93
			0.9	0.5	−0.111	0.50	0.57	0.92
			0.9	0.9	−0.196	0.76	0.96	0.91
(M10)	200	85	0.1	0.1	−0.004	0.06	0.06	0.94
			0.5	0.5	−0.003	0.08	0.08	0.96
			0.5	0.9	−0.022	0.18	0.17	0.95
			0.9	0.5	−0.027	0.21	0.21	0.94
			0.9	0.9	−0.058	0.35	0.34	0.94
	100	43	0.1	0.1	−0.006	0.09	0.10	0.92
			0.5	0.5	−0.001	0.11	0.11	0.93
			0.5	0.9	−0.040	0.26	0.27	0.92
			0.9	0.5	−0.055	0.32	0.33	0.92
			0.9	0.9	−0.129	0.54	0.59	0.91
	75	32	0.1	0.1	−0.006	0.11	0.12	0.89
			0.5	0.5	−0.006	0.13	0.13	0.94
			0.5	0.9	−0.068	0.32	0.34	0.91
			0.9	0.5	−0.103	0.40	0.44	0.93
			0.9	0.9	−0.230	0.70	0.83	0.90

Table 3.3: Power of tests of the null hypothesis $\{H_0^0 : VE(v) = 0 \text{ for all } v \in [0,1]\}$ against (i) the alternative hypothesis $\{H_1^0 : \text{non } H_0^0\}$, and (ii) the alternative hypothesis $\{H_2^0 : VE > 0 \text{ and } VE(v) \text{ is a decreasing function in } v\}$. As for testing H_0^0 versus H_1^0, the proposed Wald and likelihood ratio tests are compared with the alternative nonparametric tests $\hat{U}_3^1$ and $\hat{U}_4^1$ (Gilbert, McKeague, and Sun, 2008) based on comparing nonparametric maximum likelihood estimates of the cumulative mark-specific hazard functions $\Lambda(t, v|Z = k) = \int_0^v \int_0^t \lambda(s, u|Z = k)\mathrm{d}s\,\mathrm{d}u$, $k = 0, 1$, for the vaccine and placebo arms. Furthermore, power of the one-sided weighted Wald-type test (2.22) of H_0^0 versus H_2^0 is compared with that of the two-sided log-rank test of equal failure distributions in the vaccine and placebo arm. The nominal significance level is taken to be 5% for each two-sided test and 2.5% for each one-sided test.

					H_0^0 vs H_1^0				H_0^0 vs H_2^0	
Model	VE(0)	VE(1)	N_{pI}	N_{vI}	Wald	LRatio	$\hat{U}_3^1$	$\hat{U}_4^1$	WtWald	Logrank
(M1)	0.0	0.0	200	200	0.04	0.04	0.04	0.04	0.02	0.04
			100	100	0.05	0.04	0.06	0.06	0.03	0.05
			75	75	0.05	0.05	0.06	0.04	0.03	0.05
(M2)	0.3	0.1	200	152	0.74	0.74	0.77	0.79	0.83	0.80
			100	76	0.42	0.44	0.80	0.80	0.54	0.52
			75	57	0.31	0.31	0.38	0.38	0.41	0.39
(M3)	0.6	0.4	200	94	1.00	1.00	1.00	1.00	1.00	1.00
			100	47	0.99	0.99	0.79	0.80	1.00	0.99
			75	35	0.97	0.97	0.98	0.97	0.98	0.98
(M4)	0.5	−0.7	200	167	0.93	0.90	0.50	0.89	0.74	0.47
			100	83	0.63	0.58	0.28	0.57	0.45	0.28
			75	63	0.53	0.51	0.24	0.48	0.37	0.20
(M5)	0.9	0.1	200	58	1.00	1.00	0.98	0.98	1.00	1.00
			100	29	1.00	1.00	0.80	0.80	1.00	1.00
			75	22	1.00	1.00	1.00	1.00	1.00	1.00

Table 3.4: Power of two-sided tests of the null hypothesis $\{H_0 : VE(v) \equiv VE$ for all $v \in [0,1]\}$. The proposed Wald and likelihood ratio tests are compared with the alternative non- and semiparametric tests $\hat{U}_2^{np}$ and $\hat{U}_2^{sp}$ (Gilbert, McKeague, and Sun, 2008) based on comparing the nonparametric maximum likelihood estimate $\hat{\Lambda}(t, v|Z = 1) - \hat{\Lambda}(t, v|Z = 0)$ with a non- and semiparametric estimate of $\Lambda(t, v|Z = 1) - \Lambda(t, v|Z = 0)$ under H_0. For each test, the nominal significance level is set to 5%.

Model	VE(0)	VE(1)	N_{pI}	N_{vI}	Wald	LRatio	$\hat{U}_2^{np}$	$\hat{U}_2^{sp}$
(M1)	0.0	0.0	200	200	0.04	0.04	0.04	0.05
			100	100	0.04	0.05	0.04	0.04
			75	75	0.06	0.06	0.03	0.04
(M2)	0.3	0.1	200	152	0.12	0.13	0.08	0.08
			100	76	0.08	0.08	0.06	0.04
			75	57	0.06	0.07	0.06	0.04
(M3)	0.6	0.4	200	94	0.18	0.18	0.09	0.09
			100	47	0.12	0.12	0.10	0.07
			75	35	0.12	0.12	0.10	0.07
(M4)	0.5	−0.7	200	167	0.89	0.89	0.66	0.65
			100	83	0.60	0.61	0.36	0.35
			75	63	0.53	0.54	0.24	0.26
(M5)	0.9	0.1	200	58	0.97	0.97	0.78	0.76
			100	29	0.79	0.80	0.44	0.36
			75	22	0.67	0.66	0.36	0.32

Table 3.5: Power of tests of the null hypothesis $\{H_0^0 : VE(v) = 0 \text{ for all } v \in [0,1]^2\}$ against (i) the alternative hypothesis $\{H_1^0 : \text{non } H_0^0\}$, and (ii) the alternative hypothesis $\{H_2^0 : VE > 0 \text{ and } VE(v) \text{ is a decreasing function in } v\}$. Power of the one-sided weighted Wald-type test (2.22) of H_0^0 versus H_2^0 is compared with that of the two-sided log-rank test of equal failure distributions in the vaccine and placebo arm. The nominal significance level is taken to be 5% for each two-sided test and 2.5% for each one-sided test.

			H_0^0 vs H_1^0		H_0^0 vs H_2^0	
Model	N_{pI}	N_{vI}	Wald	LRatio	WtWald	Logrank
(M6)	200	200	0.05	0.05	0.03	0.06
	100	100	0.06	0.07	0.03	0.06
	75	75	0.06	0.07	0.02	0.06
(M7)	200	138	0.90	0.93	0.96	0.95
	100	69	0.58	0.65	0.74	0.73
	75	52	0.45	0.53	0.61	0.60
(M8)	200	138	0.92	0.93	0.97	0.95
	100	69	0.62	0.67	0.75	0.72
	75	52	0.48	0.55	0.66	0.63
(M9)	200	183	0.19	0.20	0.19	0.13
	100	91	0.12	0.12	0.10	0.08
	75	69	0.09	0.10	0.09	0.08
(M10)	200	85	1.00	1.00	1.00	1.00
	100	43	1.00	1.00	1.00	1.00
	75	32	0.99	1.00	1.00	1.00

Table 3.6: Power of two-sided tests of the null hypothesis $\{H_0 : VE(v) \equiv VE$ for all $v \in [0,1]^2\}$. For each test, the nominal significance level is set to 5%.

Model	N_{pI}	N_{vI}	Wald	LRatio
(M6)	200	200	0.04	0.05
	100	100	0.05	0.06
	75	75	0.06	0.07
(M7)	200	138	0.10	0.11
	100	69	0.08	0.08
	75	52	0.07	0.08
(M8)	200	138	0.19	0.19
	100	69	0.11	0.12
	75	52	0.09	0.10
(M9)	200	183	0.14	0.14
	100	91	0.07	0.08
	75	69	0.07	0.08
(M10)	200	85	0.52	0.52
	100	43	0.26	0.28
	75	32	0.20	0.21

Table 3.7: Size of the supremum test of the null hypothesis $\{\widetilde{K}_0 : T \perp\!\!\!\perp V\}$ in the placebo and vaccine arm with $\alpha = 0.05$ and of the overall test of the null hypothesis $\{K_0 : T \perp\!\!\!\perp V|Z\}$ using Simes' procedure and $\alpha_{\text{familywise}} = 0.05$.

Model	VE(0)	VE(1)	N_{pI}	N_{vI}	Placebo	Vaccine	Overall
(M1)	0.0	0.0	200	200	0.04	0.05	0.05
			100	100	0.04	0.04	0.04
			75	75	0.05	0.03	0.04
(M2)	0.3	0.1	200	152	0.05	0.04	0.04
			100	76	0.05	0.05	0.04
			75	57	0.04	0.05	0.04
(M3)	0.6	0.4	200	94	0.04	0.04	0.04
			100	47	0.04	0.05	0.05
			75	35	0.03	0.04	0.04
(M4)	0.5	−0.7	200	167	0.06	0.05	0.06
			100	83	0.04	0.05	0.03
			75	63	0.04	0.05	0.04
(M5)	0.9	0.1	200	58	0.06	0.05	0.05
			100	29	0.05	0.04	0.04
			75	22	0.04	0.05	0.05

Table 3.8: Size of the supremum test of the null hypothesis $\{\widetilde{K}_0 : T \perp\!\!\!\perp (V_1, V_2)\}$ in the placebo and vaccine arm with $\alpha = 0.05$ and of the overall test of the null hypothesis $\{K_0 : T \perp\!\!\!\perp (V_1, V_2)|Z\}$ using Simes' procedure and $\alpha_{\text{familywise}} = 0.05$.

Model	N_{pI}	N_{vI}	Placebo	Vaccine	Overall
(M6)	200	200	0.04	0.04	0.04
	100	100	0.05	0.04	0.05
	75	75	0.04	0.04	0.04
(M7)	200	138	0.05	0.04	0.05
	100	69	0.04	0.05	0.04
	75	52	0.05	0.04	0.03
(M8)	200	138	0.04	0.05	0.05
	100	69	0.04	0.04	0.04
	75	52	0.05	0.04	0.04
(M9)	200	183	0.05	0.04	0.04
	100	91	0.05	0.06	0.05
	75	69	0.04	0.03	0.03
(M10)	200	85	0.05	0.05	0.05
	100	43	0.05	0.04	0.04
	75	32	0.04	0.05	0.04

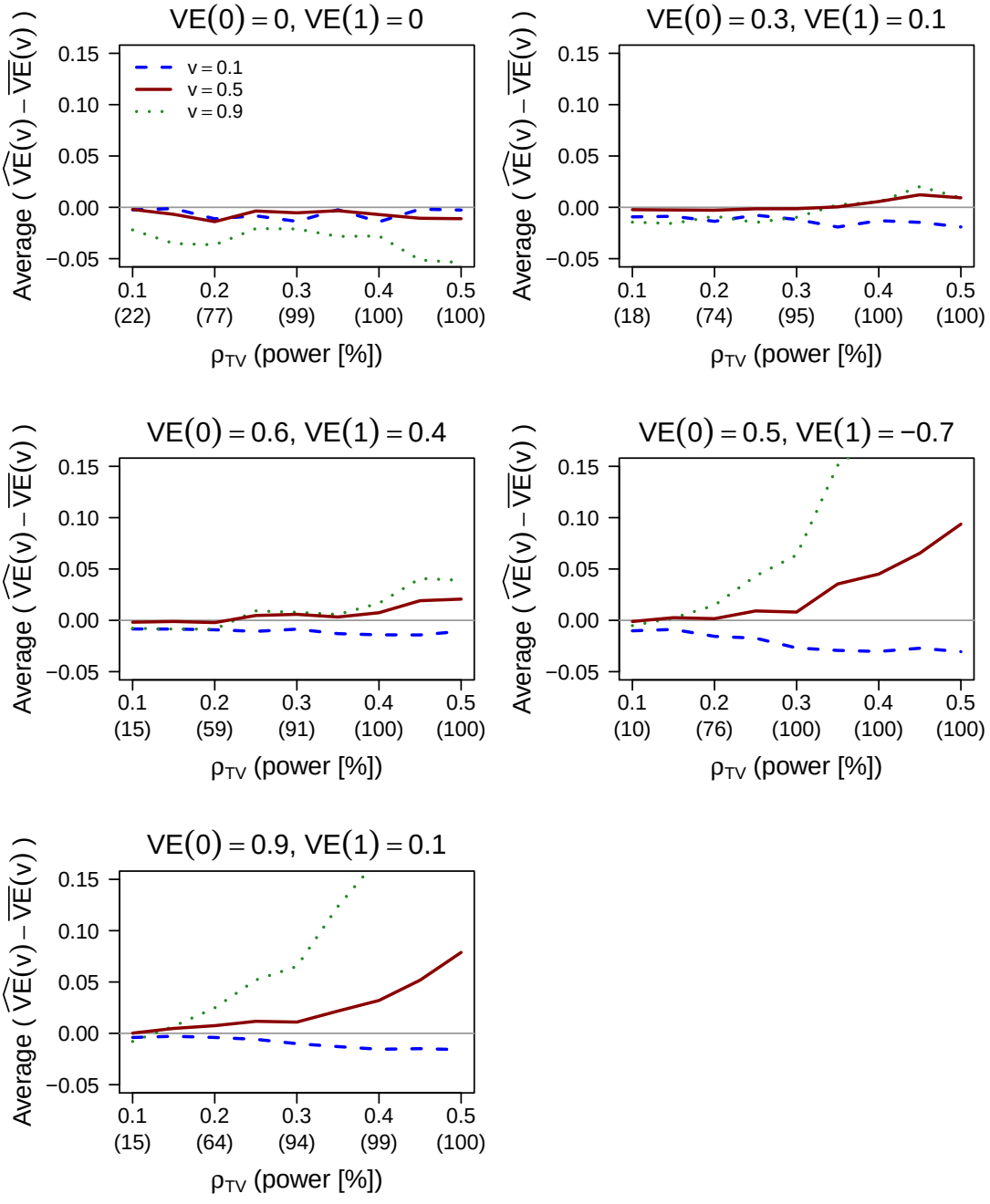


Figure 3.7: Average deviation $\widehat{VE}(v) - \overline{VE}(v)$ at $v = 0.1$, 0.5, and 0.9 as a function of the magnitude of correlation between T and V in scenarios (M1)–(M5) with $N_{pI} = 200$ (the $T \perp\!\!\!\perp V|Z$ assumption is violated). Power to reject $\{K_0 : T \perp\!\!\!\perp V|Z\}$ using the supremum test of independence is displayed in parentheses.

Table 3.9: Size and power of tests of the null hypothesis $\{H_0^0 : VE(v) = 0 \text{ for all } v \in [0,1]\}$ against (i) the alternative hypothesis $\{H_1^0 : \text{non } H_0^0\}$, and (ii) the alternative hypothesis $\{H_2^0 : VE > 0 \text{ and } VE(v) \text{ is a decreasing function in } v\}$ under violation of the $T \perp\!\!\!\perp V|Z$ assumption in scenarios with $N_{pI} = 200$. In each scenario, results for two levels of correlation between T and V (ρ_{TV}) are presented that lead to rejection of $\{T \perp\!\!\!\perp V|Z\}$ with moderate and high power using the supremum test of independence. The nominal significance level is taken to be 5% for each two-sided test and 2.5% for each one-sided test. (See Table 3.3 for details about the presented tests.)

				Power	H_0^0 vs H_1^0				H_0^0 vs H_2^0	
Model	VE(0)	VE(1)	ρ_{TV}	SupTest	Wald	LRatio	$\hat{U}_3^1$	$\hat{U}_4^1$	WtWald	Logrank
(M1)	0.0	0.0	0.15	0.36	0.04	0.05	0.04	0.07	0.02	0.04
			0.30	0.99	0.06	0.06	0.06	0.03	0.02	0.05
(M2)	0.3	0.1	0.15	0.40	0.74	0.73	0.84	0.83	0.82	0.79
			0.30	0.94	0.71	0.71	0.78	0.79	0.80	0.79
(M3)	0.6	0.4	0.15	0.37	1.00	1.00	1.00	1.00	1.00	1.00
			0.35	0.98	1.00	1.00	1.00	1.00	1.00	1.00
(M4)	0.5	−0.7	0.15	0.47	0.93	0.92	0.50	0.87	0.80	0.48
			0.25	0.97	0.91	0.89	0.43	0.87	0.74	0.47
(M5)	0.9	0.1	0.15	0.35	1.00	1.00	1.00	1.00	1.00	1.00
			0.35	0.99	1.00	1.00	1.00	1.00	1.00	1.00

Table 3.10: Size and power of two-sided tests of the null hypothesis $\{H_0 : VE(v) \equiv VE \text{ for all } v \in [0,1]\}$ under violation of the $T \perp\!\!\!\perp V|Z$ assumption in scenarios with $N_{pI} = 200$. In each scenario, results for two levels of correlation between T and V (ρ_{TV}) are presented that lead to rejection of $\{T \perp\!\!\!\perp V|Z\}$ with moderate and high power using the supremum test of independence. The nominal significance level is taken to be 5%. (See Table 3.4 for details about the presented tests.)

				Power				
Model	VE(0)	VE(1)	ρ_{TV}	SupTest	Wald	LRatio	$\hat{U}_2^{np}$	$\hat{U}_2^{sp}$
(M1)	0.0	0.0	0.15	0.36	0.05	0.05	0.03	0.04
			0.30	0.99	0.06	0.06	0.04	0.03
(M2)	0.3	0.1	0.15	0.40	0.13	0.13	0.13	0.12
			0.30	0.94	0.10	0.10	0.10	0.09
(M3)	0.6	0.4	0.15	0.37	0.18	0.19	0.12	0.12
			0.35	0.98	0.15	0.15	0.15	0.17
(M4)	0.5	−0.7	0.15	0.47	0.90	0.91	0.68	0.66
			0.25	0.97	0.88	0.88	0.67	0.64
(M5)	0.9	0.1	0.15	0.35	0.98	0.97	0.83	0.79
			0.35	0.99	0.92	0.92	0.72	0.70

Chapter 4

STATISTICAL METHODS FOR SIEVE ANALYSIS WITH MISSING MARK DATA

4.1 Introduction

Missing marks of interest present a common problem in preventive HIV-1 vaccine trials. In Chapter 1, we describe the missing mark mechanism commonly encountered in the process of collecting genetic sequence data from infected vaccine trial participants that arises as a consequence of rapid and continual HIV-1 evolution. A convenient approach in this setting is performing a complete-case analysis of mark-specific vaccine efficacy, i.e., an analysis of a complete data set excluding records with missing marks in the density ratio model. The complete-case analysis, however, may be inefficient since data on subjects with missing marks are not used in the mark density ratio estimation; the lack of efficiency may be severe if the missingness rate is high. Moreover, the complete-case analysis may provide a misleading statistical inference if the missingness mechanism is not completely at random. Therefore, in this chapter, we extend the proposed inferential methods for mark-specific vaccine efficacy that accommodate missing marks. We consider two approaches to estimation for the coefficients of the density ratio model: (i) weighting of the complete cases by the inverse of the probabilities of observing the mark of interest (Horvitz and Thompson, 1952), and (ii) augmenting of the inverse probability weighted (IPW) estimating functions by leveraging potential correlation between the mark and auxiliary data to "impute" the expected profile score vectors for subjects with both complete and incomplete mark data (using the general theory of Robins, Rotnitzky, and Zhao (1994)). To the best of our knowledge, the problem of parameter estimation in density ratio/biased

sampling models with missing data has not been addressed in the existing literature.

In Section 2.4.1, we note that the exponential form (2.7) of the weight function yields a density ratio model that is equivalent to a retrospective logistic regression model. In the presence of missing marks, the model duality suggests the possibility to use an alternative approach based on a logistic regression analysis using weighted estimating equations, developed in a general framework by Robins, Rotnitzky, and Zhao (1994) and further discussed in Zhao, Lipsitz, and Lew (1996). Efficiency and modeling robustness of the augmented estimator for parameters of a logistic regression are compared to those of the IPW estimator in Tchetgen (2009).

Alternatively, to analyze mark-specific vaccine efficacy, Sun and Gilbert (2012) proposed the stratified mark-specific proportional hazards model with univariate missing marks. They developed a consistent estimation procedure utilizing the IPW complete-case technique and augmentation of the IPW estimating equation by leveraging auxiliary data predictive of the mark. Their proposed method, however, does not accommodate multivariate marks due to numerical limitations posed by the employed kernel smoothing procedure.

4.2 Notation and assumptions

The basic survival analysis notation is introduced in Section 2.2. Henceforth we consider the mark V defined as the (set of) genetic distance(s) based on a virus isolated in the acute phase of infection, i.e., prior to the generation of HIV-specific antibodies (discussed in Chapter 1). For infected subjects unobserved in the acute phase, the mark V is missing. So, if $\delta = 1$, we define the indicator R of observing the mark V as follows: let $R = 1$ if all components of V are observed and let $R = 0$ otherwise (hence, we consider the 'all-or-none' type of missingness). Let A denote a random vector of auxiliary covariates. It suffices to consider the observation of the covariate vector A in individuals with an observed failure time ($\delta = 1$) because only those observations of A can contribute to predicting (i) the probability of a missing

V, and (ii) the expected profile score vector. Marks $\widetilde{V}$ based on viral isolates from post-acute phase samples and the corresponding sampling times can be included as a subset of A.

Let $(X_i, \delta_i, R_i, V_i, Z_i, A_i)$, $i = 1, \ldots, n$, be i.i.d. replicates of (X, δ, R, V, Z, A). The observed data consist of the observations $(X_i, R_i, R_iV_i, Z_i, A_i)$ for individuals with $\delta_i = 1$ and the observations (X_i, Z_i) for those with $\delta_i = 0$. Denote $W = (Z, A)$. We make the following assumptions about the missing mark mechanism:

$$P(R = 1|\delta = 1, W) = P(R = 1|V, \delta = 1, W) \tag{4.1}$$

and

$$\pi(W) := P(R = 1|\delta = 1, W) \geq \sigma \text{ with probability 1 for some } \sigma > 0. \tag{4.2}$$

Condition (4.1) conveys that the mark V is missing at random (Rubin, 1976), that is, given $\delta = 1$ and W, the probability of a missing V depends only on the observed W, not on the value of V. Condition (4.2) ensures that an $n^{\frac{1}{2}}$-consistent estimator for $(\phi^T, \lambda)^T$ exists (Robins, Rotnitzky, and Zhao, 1994).

The probability of a missing V is largely affected by the frequency of HIV testing (see, e.g., Table 1.1). Viral load may also be associated with missingness as levels below a detection limit may preclude sequencing of the virus. The mark V does not seem to be associated with the drop-out rate which renders missingness at random a plausible assumption in this setting.

4.3 Inverse probability weighted complete-case estimator

The idea of the inverse probability weighted complete-case estimator, originally proposed by Horvitz and Thompson (1952), is based on weighting of the complete cases by the inverse of the probabilities $\pi(W_i)$ or their estimates. We suppose that we have a correctly specified parametric model $\pi(W, \psi)$ for $\pi(W)$, i.e.,

$$\pi(W) = \pi(W, \psi_0) \tag{4.3}$$

where ψ_0 is an unknown parameter vector and $\pi(\cdot, \psi)$ is a known smooth function taking values in $(0, 1]$. Typically we posit a logistic model

$$\operatorname{logit} \pi(W, \psi) = \psi^T h(W)$$

where h is a vector function defined on the support of W. The maximum likelihood estimator $\hat{\psi}$ for ψ can be obtained by solving

$$\sum_{i \in I} S_{\psi,i}(\psi) = 0,$$

with

$$\begin{aligned} S_{\psi,i}(\psi) &= \partial \log \left\{ \pi(W_i, \psi)^{R_i} (1 - \pi(W_i, \psi))^{1-R_i} \right\} / \partial \psi \\ &= (R_i - \pi(W_i, \psi))\, \partial \operatorname{logit} \pi(W_i, \psi) / \partial \psi. \end{aligned}$$

To estimate the parameter of interest ϕ (and λ), we define

$$U_i^{ipw}(\phi, \lambda, \psi) = U_i(\phi, \lambda) \frac{R_i}{\pi(W_i, \psi)}, \quad i \in I,$$

where $U_i(\phi, \lambda) = (U_{\phi,i}^T(\phi, \lambda), U_{\lambda,i}(\phi, \lambda))^T$ is the i-th individual's contribution to the profile score vector introduced in (2.16). Let $(\hat{\phi}_{ipw}^T, \hat{\lambda}_{ipw})^T$ denote the solution to the inverse probability weighted estimating equations

$$\sum_{i \in I} U_i^{ipw}(\phi, \lambda, \hat{\psi}) = 0.$$

In the sequel, we show that $(\hat{\phi}_{ipw}^T, \hat{\lambda}_{ipw})^T$ is a consistent estimator and characterize its asymptotic distribution. Define $x^{\otimes 2} = xx^T$ for $x \in \mathbb{R}^p$. Let $H_i(\omega) = (U_i^{ipw}(\phi, \lambda, \psi)^T, S_{\psi,i}(\psi)^T)^T$ with $\omega = (\phi^T, \lambda, \psi^T)^T$. We assume that the following regularity conditions hold:

Condition A.

(i) $(\phi^T, \lambda)^T$ and ψ lie in the interior of compact sets ϕ_λ^* and ψ^*;

(ii) $\pi(\cdot, \psi) > \varepsilon > 0$ for all $\psi \in \psi^*$ and some ε;

(iii) $EH_i(\omega) = 0 \Leftrightarrow \omega = \omega_0$;

(iv) $\operatorname{var} H_i(\omega_0)$ is finite and positive definite;

(v) $E\left\{\partial H_i(\omega_0)/\partial\omega^T\right\}$ exists and is invertible;

(vi) $E\left\{\sup_{\omega\in\omega^*}\|H_i(\omega)\|\right\} < \infty$, $E\left\{\sup_{\omega\in\omega^*}\|\partial H_i(\omega)/\partial\omega^T\|\right\} < \infty$, and $E\left\{\sup_{\omega\in\omega^*}\|H_i(\omega)^{\otimes 2}\|\right\} < \infty$ where $\|A\| := \left(\sum_{ij} A_{ij}^2\right)^{\frac{1}{2}}$ for any matrix $A = (A_{ij})$ and ω^* is the Cartesian product of ϕ_λ^* and ψ^*.

Theorem 4.1. *If* (4.1), (4.2), (4.3), *and Condition A are true,* $(\phi_{ipw}^T, \lambda_{ipw})^T \xrightarrow{P} (\phi_0^T, \lambda_0)^T$ *as* $m \to \infty$.

Proof. If (4.3) is true, $\hat{\psi} \xrightarrow{P} \psi_0$ as $m \to \infty$. It follows that

$$m^{-1}U^{ipw}(\phi, \lambda, \hat{\psi}) = m^{-1}U^{ipw}(\phi, \lambda, \psi_0) + o_p(1)$$

uniformly in $(\phi, \lambda) \in \phi_\lambda^*$. By the Glivenko-Cantelli theorem, $m^{-1}U^{ipw}(\phi, \lambda, \psi_0) = EU_i^{ipw}(\phi, \lambda, \psi_0) + o_p(1)$, uniformly in $(\phi, \lambda) \in \phi_\lambda^*$, where, using the double expectation formula $E[\cdot] = E\{E[\cdot|V_i, W_i, \delta = 1]\}$ and the missing at random assumption (4.1), the expectation is zero. Therefore,

$$m^{-1}U^{ipw}(\phi, \lambda, \hat{\psi}) = EU_i^{ipw}(\phi, \lambda, \psi_0) + o_p(1) = o_p(1),$$

uniformly in $(\phi, \lambda) \in \phi_\lambda^*$, and, by van der Vaart (1998, Theorem 5.9), $(\phi_{ipw}^T, \lambda_{ipw})^T \xrightarrow{P} (\phi_0^T, \lambda_0)^T$ as $m \to \infty$. □

Theorem 4.2. *If* (4.1), (4.2), (4.3), *and Condition A are true, then (i) with probability approaching 1,* $(\hat{\phi}_{ipw}^T, \hat{\lambda}_{ipw})^T$ *exists and is unique; (ii)* $m^{\frac{1}{2}}\left((\hat{\phi}_{ipw}^T, \hat{\lambda}_{ipw})^T - (\phi_0^T, \lambda_0)^T\right)$ *is asymptotically normal with mean 0 and variance* $J^{-1}D(J^{-1})^T$, *that can be consistently estimated by* $\widehat{J}^{-1}\widehat{D}(\widehat{J}^{-1})^T$, *where*

$$J = E\left\{\partial U_i^{ipw}(\phi_0, \lambda_0, \psi_0)/\partial(\phi^T, \lambda)\right\}, \quad D = E\left\{\operatorname{resid}(U_i^{ipw}, S_{\psi,i})^{\otimes 2}\right\},$$
$$U_i^{ipw} = U_i^{ipw}(\phi_0, \lambda_0, \psi_0), \quad S_{\psi,i} = S_{\psi,i}(\psi_0),$$

with $\text{resid}(A,B) = A - E(AB^T)(E(BB^T))^{-1}B$ *the residual vector from the population least squares regressions of the components of* A *on* B*,*

$$\widehat{J} = m^{-1}\sum_{i\in I}\partial U_i^{ipw}(\hat{\phi}_{ipw},\hat{\lambda}_{ipw},\hat{\psi})/\partial(\phi^T,\lambda), \quad \widehat{D} = m^{-1}\sum_{i\in I}\text{Resid}(U_i^{ipw},S_{\psi,i})^{\otimes 2},$$

with $\text{Resid}(U_i^{ipw},S_{\psi,i})$ *the residual vector for subject* i *from the least squares regressions of the components of* $U_i^{ipw}(\hat{\phi}_{ipw},\hat{\lambda}_{ipw},\hat{\psi})$ *on* $S_{\psi,i}(\hat{\psi})$*,* $i \in I$*.*

Proof. The proof is constructed in a fashion analogous to the proof of Proposition 6.1 of Robins, Rotnitzky, and Zhao (1994). Under Condition A, Theorems 2.6 and 3.4 of Newey and McFadden (1993) imply that, with probability approaching 1, there exists a unique solution to $\sum_{i\in I} H_i(\omega) = 0$.

Condition A ensures the validity of the standard Taylor expansions

$$m^{\frac{1}{2}}(\hat{\psi}-\psi_0) = -m^{-\frac{1}{2}}\left(E\,\frac{\partial S_{\psi,i}}{\partial\psi^T}\right)^{-1}\sum_{i\in I}S_{\psi,i} + o_p(1) \tag{4.4}$$

and

$$\begin{aligned} &m^{\frac{1}{2}}\left((\hat{\phi}_{ipw}^T,\hat{\lambda}_{ipw})^T - (\phi_0^T,\lambda_0)^T\right) = \\ &-m^{-\frac{1}{2}}\left(E\,\frac{\partial U_i^{ipw}}{\partial(\phi^T,\lambda)}\right)^{-1}\left\{\sum_{i\in I}U_i^{ipw} - E\,\frac{\partial U_i^{ipw}}{\partial\psi^T}\left(E\,\frac{\partial S_{\psi,i}}{\partial\psi^T}\right)^{-1}\sum_{i\in I}S_{\psi,i}\right\} + o_p(1). \end{aligned} \tag{4.5}$$

Pierce (1982) showed that $E\,\frac{\partial U_i^{ipw}}{\partial\psi^T} = -E(U_i^{ipw}S_{\psi,i}^T)$ and $-E\,\frac{\partial S_{\psi,i}}{\partial\psi^T} = \text{var}\,S_{\psi,i}$. Applying the identities in (4.5), we obtain

$$m^{\frac{1}{2}}\left((\hat{\phi}_{ipw}^T,\hat{\lambda}_{ipw})^T - (\phi_0^T,\lambda_0)^T\right) = -J^{-1}m^{-\frac{1}{2}}\sum_{i\in I}\text{resid}(U_i^{ipw},S_{\psi,i}) + o_p(1). \tag{4.6}$$

The asymptotic distribution of $m^{\frac{1}{2}}\left((\hat{\phi}_{ipw}^T,\hat{\lambda}_{ipw})^T - (\phi_0^T,\lambda_0)^T\right)$ follows from (4.6) and the Central Limit Theorem. The consistency of $\widehat{J}$ and $\widehat{D}$ is implied by the Law of Large Numbers and the consistency of $(\phi_{ipw}^T,\lambda_{ipw})^T$. □

4.4 Augmented inverse probability weighted complete-case estimator

The IPW estimator fails to make optimal use of the available data and may be inefficient even if the model posited for the probabilities $\pi(w)$ is correct. Theorem 4.1 shows that the estimator $(\hat{\phi}_{ipw}^T, \hat{\lambda}_{ipw})^T$ is consistent if the probabilities $\pi(w)$ are modeled correctly; otherwise it may be biased. To improve efficiency and robustness to mis-specification of the missingness model, Robins, Rotnitzky, and Zhao (1994) proposed the idea of adding an augmented term to the IPW estimating function in a general framework. In the model (2.6), a more efficient estimator for $(\phi^T, \lambda)^T$ can be obtained by adding information about the conditional expectation $E[U(\phi, \lambda)|\delta = 1, W]$ into the estimation procedure. Assumption (4.1) implies that $E[U(\phi, \lambda)|\delta = 1, W] = E[U(\phi, \lambda)|R = 1, \delta = 1, W]$. We suppose that we have a correctly specified parametric model $q(W, \phi, \lambda, \nu)$ for $E[U(\phi, \lambda)|\delta = 1, W]$, i.e., for $(\phi^T, \lambda)^T \in \phi_\lambda^*$,

$$E[U(\phi, \lambda)|\delta = 1, W] = q(W, \phi, \lambda, \nu_0) \tag{4.7}$$

where $\nu_0 = \nu_0(\phi, \lambda)$ is an unknown parameter vector and $q(\cdot, \phi, \lambda, \nu)$ is a known smooth function in $(\phi^T, \lambda)^T$. Denote $\hat{\nu} = \hat{\nu}(\phi, \lambda)$ a uniformly consistent estimator for $\nu_0(\phi, \lambda)$ in $(\phi, \lambda) \in \phi_\lambda^*$. We define

$$U_i^{aug}(\phi, \lambda, \psi, \nu) = U_i(\phi, \lambda)\frac{R_i}{\pi(W_i, \psi)} + q(W_i, \phi, \lambda, \nu)\left(1 - \frac{R_i}{\pi(W_i, \psi)}\right), \quad i \in I.$$

The estimator $(\hat{\phi}_{aug}^T, \hat{\lambda}_{aug})^T$ denotes the solution to the augmented IPW estimating equations

$$\sum_{i \in I} U_i^{aug}(\phi, \lambda, \hat{\psi}, \hat{\nu}) = 0. \tag{4.8}$$

Next we show that $(\hat{\phi}_{aug}^T, \hat{\lambda}_{aug})^T$ is a consistent estimator and characterize its asymptotic distribution. In Condition A, we now consider $H_i(\omega) = (U_i^{aug}(\phi, \lambda, \psi, \nu)^T, S_{\psi,i}(\psi)^T)^T$ with $\omega = (\phi^T, \lambda, \psi^T)^T$.

Theorem 4.3. *Under validity of* (4.1), (4.2), *and Condition A,* $(\phi_{aug}^T, \lambda_{aug})^T \xrightarrow{P} (\phi_0^T, \lambda_0)^T$ *as* $m \to \infty$ *if either* $\pi(w, \psi)$ *or* $q(w, \phi, \lambda, \nu)$ *is correctly specified.*

Proof. Suppose $q(w, \phi, \lambda, \hat{\nu}) \xrightarrow{P} q(w, \phi, \lambda, \nu^*)$. It follows that

$$m^{-1}U^{aug}(\phi, \lambda, \hat{\psi}, \hat{\nu}) = m^{-1}U^{aug}(\phi, \lambda, \psi_0, \nu^*) + o_p(1),$$

uniformly in $(\phi, \lambda) \in \phi^*_\lambda$. By the Glivenko-Cantelli theorem, $m^{-1}U^{aug}(\phi, \lambda, \psi_0, \nu^*) = EU_i^{aug}(\phi, \lambda, \psi_0, \nu^*) + o_p(1)$, uniformly in $(\phi, \lambda) \in \phi^*_\lambda$, where, using the double expectation formula $E[\cdot] = E\{E[\cdot|V_i, W_i, \delta = 1]\}$ and the missing at random assumption (4.1), the expectation is zero if $\pi(w, \psi)$ or $q(w, \phi, \lambda, \nu)$ is correctly specified. Therefore,

$$m^{-1}U^{aug}(\phi, \lambda, \hat{\psi}, \hat{\nu}) = EU_i^{aug}(\phi, \lambda, \psi_0, \nu_0) + o_p(1) = o_p(1),$$

uniformly in $(\phi, \lambda) \in \phi^*_\lambda$, and, by van der Vaart (1998, Theorem 5.9), $(\phi^T_{aug}, \lambda_{aug})^T \xrightarrow{P} (\phi^T_0, \lambda_0)^T$ as $m \to \infty$. □

Theorem 4.3 demonstrates that the augmented IPW estimator is partially protected against model mis-specification. The estimator $(\hat{\phi}^T_{aug}, \hat{\lambda}_{aug})^T$ remains consistent for $(\phi^T, \lambda)^T$ if (i) the missingness model for $\pi(w)$ is mis-specified provided that the conditional expectation $E[U(\phi, \lambda)|\delta = 1, W]$ is correctly modeled, and (ii) if the model for the weights $\pi(w)$ is correct regardless of the correctness of the model for $E[U(\phi, \lambda)|\delta = 1, W]$. This is the so-called double robustness property which is appealing because it provides the analyst with two separate modeling opportunities to achieve a consistent estimator for $(\phi^T, \lambda)^T$ as opposed to only a single such opportunity in case of the IPW estimator.

Theorem 4.4. *If* (4.1), (4.2), (4.3), (4.7), *and Condition A are true, then* $m^{\frac{1}{2}}\left((\hat{\phi}^T_{aug}, \hat{\lambda}_{aug})^T - (\phi^T_0, \lambda_0)^T\right)$ *is asymptotically normal with mean 0 and variance* $J_*^{-1}D_*(J_*^{-1})^T$, *that can be consistently estimated by* $\widehat{J}_*^{-1}\widehat{D}_*(\widehat{J}_*^{-1})^T$, *where*

$$J_* = E\left\{\partial U_i^{aug}(\phi_0, \lambda_0, \psi_0, \nu_0)/\partial(\phi^T, \lambda)\right\}, \quad D_* = E\left\{\text{resid}(U_i^{aug}, S_{\psi,i})^{\otimes 2}\right\},$$
$$U_i^{aug} = U_i^{aug}(\phi_0, \lambda_0, \psi_0, \nu_0), \quad S_{\psi,i} = S_{\psi,i}(\psi_0),$$

with $\text{resid}(A, B) = A - E(AB^T)(E(BB^T))^{-1}B$ *the residual vector from the population least squares regressions of the components of* A *on* B*,*

$$\widehat{J}_* = m^{-1}\sum_{i\in I}\partial U_i^{aug}(\hat{\phi}_{aug}, \hat{\lambda}_{aug}, \hat{\psi}, \hat{\nu})/\partial(\phi^T, \lambda), \quad \widehat{D}_* = m^{-1}\sum_{i\in I}\text{Resid}(U_i^{aug}, S_{\psi,i})^{\otimes 2},$$

with $\text{Resid}(U_i^{aug}, S_{\psi,i})$ *the residual vector for subject* i *from the least squares regressions of the components of* $U_i^{aug}(\hat{\phi}_{aug}, \hat{\lambda}_{aug}, \hat{\psi}, \hat{\nu})$ *on* $S_{\psi,i}(\hat{\psi})$*,* $i \in I$*.*

Proof. By the Taylor expansion of $U^{aug}(\hat{\phi}_{aug}, \hat{\lambda}_{aug}, \hat{\psi}, \hat{\nu})$ around $(\phi_0^T, \lambda_0)^T$, and consistency of $(\hat{\phi}_{aug}^T, \hat{\lambda}_{aug})^T$,

$$m^{\frac{1}{2}}\left((\hat{\phi}_{aug}^T, \hat{\lambda}_{aug})^T - (\phi_0^T, \lambda_0)^T\right) = \\ -\left(E\frac{\partial U_i^{aug}}{\partial(\phi^T, \lambda)}\right)^{-1} m^{-\frac{1}{2}}\, U^{aug}(\phi_0, \lambda_0, \hat{\psi}, \hat{\nu}) + o_p(1). \quad (4.9)$$

Next we study the asymptotic behavior of $m^{-\frac{1}{2}}\, U^{aug}(\phi_0, \lambda_0, \hat{\psi}, \hat{\nu})$. To this end, let

$$d_m(W_i, \phi, \lambda) = q(W_i, \phi, \lambda, \hat{\nu}) - q(W_i, \phi, \lambda, \nu_0),$$
$$C = m^{-\frac{1}{2}}\sum_{i\in I} d_m(W_i, \phi_0, \lambda_0)\left(1 - \frac{R_i}{\pi(W_i, \hat{\psi})}\right).$$

It follows that $d_m(W_i, \phi_0, \lambda_0) \xrightarrow{P} 0$ uniformly in W_i. The application of the Taylor expansion of $1/\pi(W_i, \hat{\psi})$ around ψ_0 yields

$$C = m^{-\frac{1}{2}}\sum_{i\in I} d_m(W_i, \phi_0, \lambda_0)\left(1 - \frac{R_i}{\pi(W_i, \psi_0)}\right) \\ + m^{-\frac{1}{2}}\sum_{i\in I} d_m(W_i, \phi_0, \lambda_0)\frac{R_i}{(\pi(W_i, \psi_0))^2}\frac{\partial\pi(W_i, \psi_0)}{\partial\psi^T}(\hat{\psi} - \psi_0) + o_p(1).$$

The first summand of C is $o_p(1)$ by Sun and Gilbert (2011, Lemma 3). Since $m^{1/2}(\hat{\psi} - \psi_0) = O_p(1)$ and $d_m(W_i, \phi_0, \lambda_0) = o_p(1)$ uniformly in W_i, the second summand of C is $o_p(1)$. It follows that $C = o_p(1)$.

Subsequently, using the Taylor expansion of $U^{aug}(\phi_0, \lambda_0, \hat{\psi}, \nu_0)$ around ψ_0 and (4.4)

yield

$$
\begin{aligned}
m^{-\frac{1}{2}}\, U^{aug}(\phi_0,&\lambda_0, \hat{\psi}, \hat{\nu}) \\
&= m^{-\frac{1}{2}}\, U^{aug}(\phi_0, \lambda_0, \hat{\psi}, \nu_0) + C \\
&= m^{-\frac{1}{2}} \left\{ U^{aug}(\phi_0, \lambda_0, \psi_0, \nu_0) + mE\frac{U_i^{aug}}{\partial \psi^T}(\hat{\psi} - \psi_0) \right\} + o_p(1) \\
&= m^{-\frac{1}{2}} \left\{ U^{aug}(\phi_0, \lambda_0, \psi_0, \nu_0) - E\frac{U_i^{aug}}{\partial \psi^T} \left(E\frac{\partial S_{\psi,i}}{\partial \psi^T} \right)^{-1} S_\psi(\psi_0) \right\} + o_p(1)
\end{aligned}
\tag{4.10}
$$

Pierce (1982) showed that $E\,\frac{\partial U_i^{aug}}{\partial \psi^T} = -E(U_i^{aug} S_{\psi,i}^T)$ and $-E\,\frac{\partial S_{\psi,i}}{\partial \psi^T} = \operatorname{var} S_{\psi,i}$. Applying the identities in (4.10) and combining the result with (4.9), we obtain

$$
m^{\frac{1}{2}} \left((\hat{\phi}_{aug}^T, \hat{\lambda}_{aug})^T - (\phi_0^T, \lambda_0)^T \right) = -J_*^{-1} m^{-\frac{1}{2}} \sum_{i \in I} \operatorname{resid}(U_i^{aug}, S_{\psi,i}) + o_p(1). \tag{4.11}
$$

The asymptotic distribution of $m^{\frac{1}{2}} \left((\hat{\phi}_{aug}^T, \hat{\lambda}_{aug})^T - (\phi_0^T, \lambda_0)^T \right)$ follows from (4.11) and the Central Limit Theorem. The consistency of $\widehat{J}_*$ and $\widehat{D}_*$ is implied by the Law of Large Numbers and the consistency of $(\hat{\phi}_{aug}^T, \hat{\lambda}_{aug})^T$. □

Estimation of $(\phi^T, \lambda)^T$ assuming (4.7) is not feasible because it depends on the unknown population quantity $E[U(\phi, \lambda)|\delta = 1, W]$. Therefore, in practice, we specify good-fitting linear regression models $\widetilde{q}(W, \phi, \lambda, \nu)$ for the components of $E[U(\phi, \lambda)|\delta = 1, W]$. First, we estimate ν by the ordinary least squares method using observations with $R = 1$. Based on the fitted model, we compute a predicted value $\hat{E}[U_i(\phi, \lambda)|\delta_i = 1, W_i] = \widetilde{q}(W_i, \phi, \lambda, \hat{\nu})$ for each subject $i \in I$. Second, the augmented IPW estimating equations in (4.8) are solved upon replacing the unknown quantities $q(W_i, \phi, \lambda, \hat{\nu})$ by the predicted values $\widetilde{q}(W_i, \phi, \lambda, \hat{\nu})$.

4.5 Hypothesis testing

In Section 2.6, we present three Wald-based testing procedures for the null hypotheses H_0^0 in (2.20) and H_0 in (2.21). In the presence of missing marks, we consider

versions of the Wald test statistics induced by the estimators $\hat{\beta}_{ipw}$ and $\hat{\beta}_{aug}$; for H_0^0 in conjunction with the marginal log hazard ratio estimator $\hat{\gamma}$ which is not impacted by the incompleteness of mark data.

4.6 Discussion

In HIV-1 vaccine trials, the mark based on an 'early' virus is commonly subject to missingness. In this chapter, we discuss two estimation methods for the parameters of the density ratio model that accommodate marks missing at random. The first approach is based on the IPW technique which provides a re-weighted complete-case estimator. As applied to all IPW-based procedures, if $\pi(W_i, \hat{\psi})$ is close to zero, the i-th observation has a large influence on the IPW estimator which may result in an unstable estimator in small to moderate sample sizes. Therefore, covariates modeling the probability of a missing mark should be chosen with caution to prevent poor performance of the IPW estimator.

The second approach augments the IPW estimating equation by utilizing the potential correlation between the mark and auxiliary data to predict the expected profile scores for failures with both complete and incomplete marks. The augmented IPW estimator exhibits the attractive double robustness property and, as the simulation study in Chapter 5 indicates, is considerably more efficient than the IPW estimator.

Chapter 5

SIMULATION STUDY OF THE INVERSE PROBABILITY WEIGHTED COMPLETE-CASE AND THE AUGMENTED INVERSE PROBABILITY WEIGHTED ESTIMATORS

5.1 Introduction

We conduct a simulation study to investigate the finite-sample performance of the proposed estimation and testing procedures in the presence of missing marks. The augmented inverse probability weighted estimator (AUG) for $VE(v) = 1 - \mathrm{e}^{\alpha+\beta v+\gamma}$, $v \in [0, 1]$, is compared to the complete-case estimator (CC) which ignores information about failures with missing marks, and the inverse probability weighted complete-case estimator (IPW). We additionally compare the aforementioned estimators to the full data likelihood estimator (Full) which uses the complete set of marks before a fraction of them is deleted.

The failure times T, censoring times C, and marks V are generated as described in Chapter 3 considering two values of the exponential failure rate, $\lambda_1 = \log(0.85)/(-3)$ and $\lambda_2 = \log(0.7)/(-3)$, so that, respectively, 85% and 70% of failure times in the placebo group are administratively censored at year 3. We consider simulation scenarios (M1)–(M5), involving a univariate mark, characterized in Section 3.2.2. The sample size $N = 1481$ per arm results in $N_{pI} = 200$ for the failure rate λ_1 and $N_{pI} = 400$ for λ_2. In the vaccine group, the expected number of observed failures N_{vI} additionally depends on the marginal log hazard ratio γ, and thus, for scenarios (M1)–(M5), $N_{vI} = 200$, 152, 94, 167, and 58 for λ_1 and $N_{vI} = 400$, 304, 188, 334, and 116 for λ_2.

5.2 Assessment of the IPW and AUG estimation procedures under correctly specified missing mark models

We generate complete-case indicators R with conditional probabilities $\pi(W) = P(R = 1|\delta = 1, W)$ satisfying the models

$$\text{(L1): } \operatorname{logit} \pi(W, \psi) = \psi_0 + \psi_1 Z + \psi_2 A + \psi_3 ZA,$$

$$\text{(L2): } \operatorname{logit} \pi(W, \psi) = \psi_0 + \psi_1 Z + \psi_2 A^* + \psi_3 ZA^*,$$

$$\text{(L3): } \operatorname{logit} \pi(W, \psi) = \psi_0 + \psi_1 Z.$$

We assume a continuous auxiliary variable A that, conditional on (V, Z), follows the model

$$A = (1 + \kappa)^{-1}(V + \kappa U), \quad \kappa > 0, \tag{5.1}$$

where $U \sim \mathsf{Uniform}(0, 1)$, independent of V. The parameter κ governs the level of association between V and A. For each of (L1)–(L3), we evaluate three AUG estimators: AUG-1 for $\kappa = 0.2$ corresponding to $\rho \approx 0.98$, AUG-2 for $\kappa = 0.4$ with $\rho \approx 0.92$, and AUG-3 for $\kappa = 0.8$ with $\rho \approx 0.76$ where ρ denotes the correlation coefficient between V and A. For model (L1), we study three IPW estimators: IPW-1 for $\rho \approx 0.98$, IPW-2 for $\rho \approx 0.92$, and IPW-3 for $\rho \approx 0.76$, whereas each of models (L2) and (L3) evaluates a single IPW estimator. In model (L2), we consider a dichotomous auxiliary covariate A^* that is conditionally independent of A given V and generated in two steps: first, generate $\widetilde{A}$ following (5.1) with $\kappa = 0.4$, and second, generate A^* from $\mathsf{Bernoulli}(\widetilde{A})$.

We investigate settings with relatively high correlations because of their feasibility in real data sets as between-subject HIV sequence diversity is considerably larger than within-subject HIV sequence diversity (Keele et al., 2008). Correlations between sequence distances based on an early and later virus have been found as high as 0.98.

In (L1)–(L3), we consider the following values of ψ with respective missing mark rates in the placebo and vaccine group:

(L1): $\psi = (-2, 0.4, 0.5, 0.8)$ resulting in $\approx 87\%$ and 73% of the marks missing;

(L2): $\psi = (-2, 0.4, 0.5, 0.8)$ resulting in $\approx 86\%$ and 70% of the marks missing;

(L3): $\psi = (-0.8, 0.5)$ resulting in $\approx 69\%$ and 58% of the marks missing.

For the AUG estimator, we assume a linear regression model $E[U(\phi, \lambda)|W] = h(W, \nu)$ of the form

$$E[U(\phi, \lambda)|W] = \nu_0 + \nu_1 Z + \nu_2 A + \nu_3 A^2 + \nu_4 ZA; \tag{5.2}$$

fitted for subjects with $R_i = 1$ by the ordinary least squares method. Rotnitzky and Robins (1995) advocate the use of highly parameterized regression models in this setting. Predicted values $\hat{E}[U_i(\phi, \lambda)|W_i]$ used to construct the AUG estimator are specified for subjects with $\delta_i = 1$ as $\hat{E}[U_i(\phi, \lambda)|W_i] = h(W_i, \hat{\nu})$. In each simulation scenario, the model (5.2) is flexible but likely to be mis-specified. The results are based on 1000 replicated data sets.

5.2.1 Simulation results

Figures 5.1–5.9 and Tables 5.1–5.4 summarize the simulation results. Finite-sample bias, asymptotic standard errors of $\widehat{VE}(v)$, and coverage probabilities of Wald-based pointwise confidence intervals for $VE(v)$ using four estimation procedures - Full, AUG, IPW and CC - are investigated in Figures 5.1–5.3 for model (L1), Figures 5.4–5.6 for model (L2), and Figures 5.7–5.9 for model (L3). In addition to the study of $VE(v)$, Tables 5.1–5.3 characterize finite-sample performance of the estimators for the mark coefficient β in the density ratio model (2.6). Size of the Wald tests of H_0^0 and H_0 is examined in Table 5.4.

Estimation of mark-specific vaccine efficacy $VE(v)$

Finite-sample bias of the AUG and IPW estimators for $VE(v)$ is small under each of (L1)–(L3). It tends to be smaller for the AUG than the IPW estimator, sometimes markedly so (Figures 5.4 and 5.7). Bias of the AUG-1 estimator tends to be close to that of the Full estimator representing the unachievable benchmark. Under (L1)

and (L2), the CC estimator is substantially biased illustrating the inappropriateness of its use when marks are missing at random (Figures 5.1 and 5.4). The change in bias along the mark support is similar for each estimation procedure. It is minimal for v in the central region and tends to increase in magnitude in the right tail of the mark distribution.

The evaluation of median asymptotic standard error estimates of $\widehat{VE}(v)$ indicates a substantial efficiency gain of the AUG estimator relative to the IPW estimator. The most pronounced efficiency gains are observed under models (L2) and (L3) (Figures 5.5 and 5.8), with the AUG-1 estimator nearing efficiency of the Full estimator. Under (L1), AUG-k proves to be more efficient than IPW-k, $k = 1, 2, 3$, for each of the density ratio models (M1)–(M5). In each setting, the CC estimator is highly inefficient.

Coverage probabilities of 95% pointwise confidence intervals are near the nominal confidence level for scenarios (M1)–(M3). Under missingness models (L2) and (L3), coverage probabilities in scenarios (M4) and (M5) fluctuate moderately. We verified that, in these scenarios, adequate coverage probabilities were attained for lower missing mark rates.

Estimation of mark coefficient β in the density ratio model

Although finite-sample bias of the AUG and IPW estimators for β in the density ratio model (2.6) is minimal, it tends to be smaller in magnitude for the AUG estimator (Table 5.1). The CC estimator is substantially biased rendering it inappropriate in this setting. Under (L2) and (L3), all AUG estimators are more efficient[1] than the IPW estimator (Table 5.2). Under (L1), AUG-k is more efficient than IPW-k, $k = 1, 2, 3$, for each of (M1)–(M5). Under all missing mark models, the AUG-1 estimator is nearly as efficient as the Full estimator whereas the CC estimator is highly inefficient. Coverage probabilities of 95% confidence intervals for β are near

[1]Relative efficiency is defined as the ratio of median asymptotic variance estimates of the given $\hat{\beta}$ and $\hat{\beta}_{\text{Full}}$.

the nominal confidence level for all AUG and IPW estimators (Table 5.3). For the CC estimator, the coverage probabilities tend to be too low.

Wald tests of the null hypotheses H_0^0 and H_0

Table 5.4 compares size of three Wald tests, each based on four estimation procedures - Full, AUG, IPW and CC: two-sided Wald test of H_0^0, one-sided weighted Wald-type test of H_0^0, and two-sided Wald test of H_0. Size of the Wald tests induced by the AUG and IPW estimators is in accordance with the nominal significance level under each of the missing mark models (L1)–(L3). In contrast, for the CC estimator, size of each Wald test tends to be inadequately large.

5.3 Robustness analysis of the IPW and AUG estimation procedures under mis-specified missing mark models

This section investigates robustness of the proposed estimators to mis-specification of $\pi(w, \psi)$ and to violation of the missing at random assumption. To study robustness to mis-specification of $\pi(w, \psi)$, we assume model (L3) while the complete-case indicator is generated with the conditional probability $\pi(W, \psi)$ satisfying

$$\text{(L4): } \operatorname{logit} \pi(W, \psi) = \psi_0 + \psi_1 Z + \psi_2 X.$$

Considering $\psi = (-0.8, 0.5, -0.5)$, $\approx 82\%$ and 71% of the marks are missing in the placebo and vaccine group, respectively.

To examine robustness to violation of the missing at random assumption, we assume model (L3) while the complete-case indicator R depends on V conditionally on W using the model

$$\text{(L5): } \operatorname{logit} \pi(W, \psi) = \psi_0 + \psi_1 Z + \psi_2 V.$$

Here we consider $\psi = (-0.1, 0.5, -2)$ which results in $\approx 71\%$ and 63% of the marks missing in the placebo and vaccine group, respectively.

5.3.1 Simulation results

Figures 5.10–5.13 and Tables 5.5–5.7 summarize the robustness analysis results. Finite-sample bias and asymptotic standard errors of $\widehat{VE}(v)$ using four estimation procedures - Full, AUG, IPW and CC - are investigated in Figures 5.10–5.11 for mis-specified model (L4) and Figures 5.12–5.13 for model (L5) violating the MAR assumption. Tables 5.5–5.6 characterize robustness of the estimators for the mark coefficient β in the density ratio model (2.6). Robustness of the Wald tests of H_0^0 and H_0 to model mis-specification is examined in Table 5.7.

Estimation of mark-specific vaccine efficacy $VE(v)$

The IPW estimator is biased when the missing mark model is mis-specified while the AUG estimator exhibits bias that is noticeably smaller and approaches that of the Full estimator with an increase in correlation between A and V (Figure 5.10). Violation of the MAR assumption induces severe bias of the IPW estimator (Figure 5.12). On the contrary, bias of the AUG estimator is markedly smaller and, for large levels of correlation between A and V, it approaches that of the Full estimator. Under both (L4) and (L5), bias of the CC estimator is severe.

Under (L4), the IPW estimator is visibly less efficient than the AUG estimator, with the AUG-1 estimator retaining efficiency close to that of the Full estimator (Figure 5.11). Violation of the MAR assumption renders the IPW estimator severely inefficient (Figure 5.13). Surprisingly, the AUG estimator appears to be robust to violation of the MAR assumption, with the AUG-1 estimator exhibiting minimal efficiency loss compared to the Full estimator.

Estimation of mark coefficient β in the density ratio model

Finite-sample bias of the IPW estimator is more impacted by departure from the MAR assumption than mis-specification of the missing mark model (Table 5.5). In contrast,

bias of the AUG estimator appears to be robust under both (L4) and (L5). The CC estimator is biased, more severely under (L5). The AUG estimator is markedly more efficient than the IPW estimator, with high correlation levels between A and V resulting in minimal efficiency loss relative to the Full estimator (Table 5.6). Under both (L4) and (L5), the CC estimator is highly inefficient.

Wald tests of the null hypotheses H_0^0 and H_0

Size of the Wald tests appears to be robust to model mis-specification (Table 5.7). Under violation of the MAR assumption, size of each Wald test tends to slightly exceed the nominal significance level.

5.4 Conclusions

The simulation study demonstrates that the proposed estimation and hypothesis testing methods for missing marks perform well in finite samples. The AUG estimator tends to clearly outperform the IPW estimator in terms of bias, relative efficiency, and model robustness. The complete-case analysis is shown to give a misleading inference. The Wald tests induced by the AUG and IPW estimators attain the nominal significance level under correctly and mis-specified missingness models and are slightly liberal under violation of the MAR assumption.

5.5 *Tables and figures*

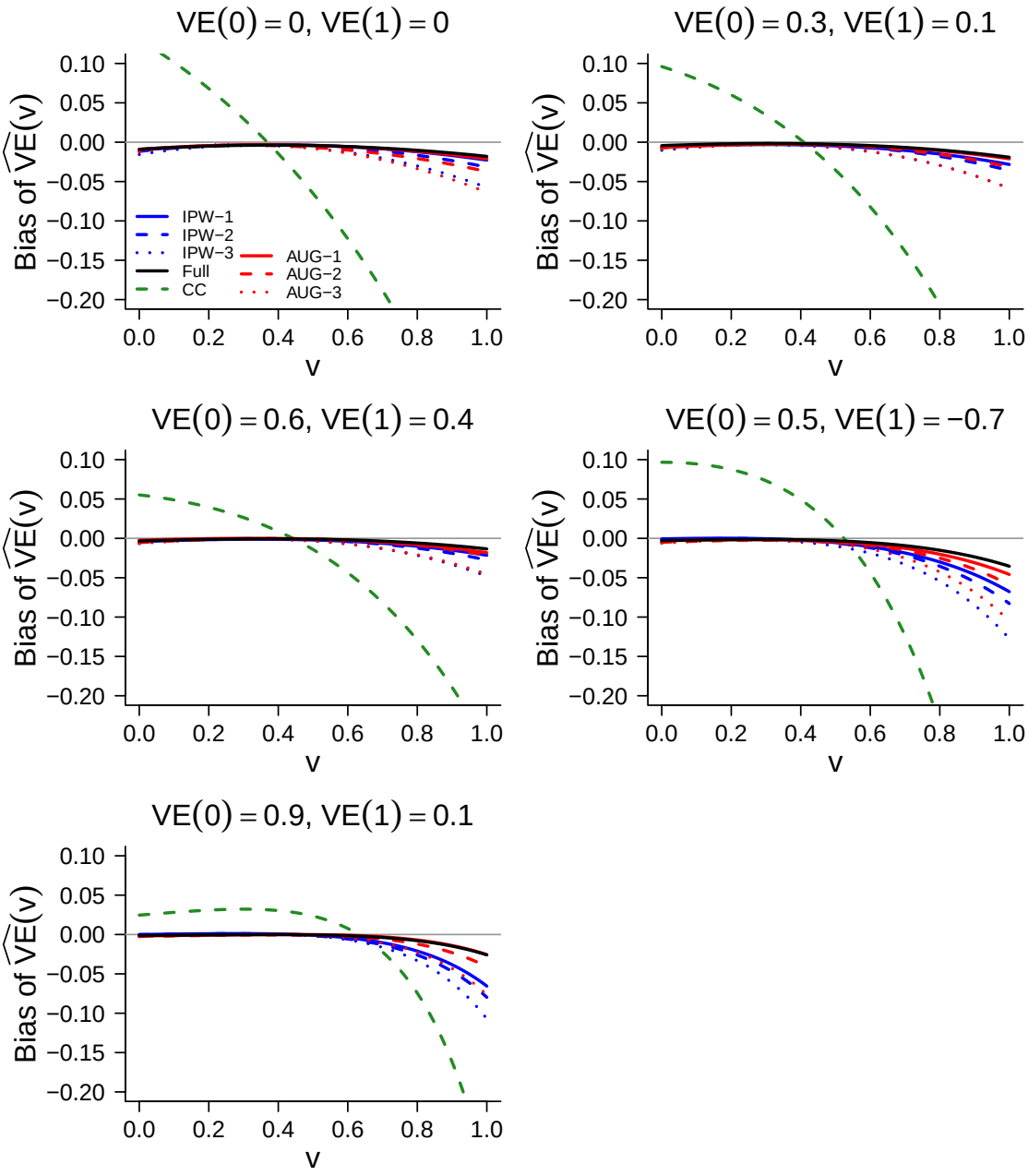


Figure 5.1: Bias of estimation for $VE(v) = 1 - e^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L1) and the expected number of 400 failures in the placebo group. Three AUG and IPW estimators are evaluated: AUG-1 and IPW-1 corresponding to $\rho \approx 0.98$, AUG-2 and IPW-2 for $\rho \approx 0.92$, and AUG-3 and IPW-3 for $\rho \approx 0.76$.

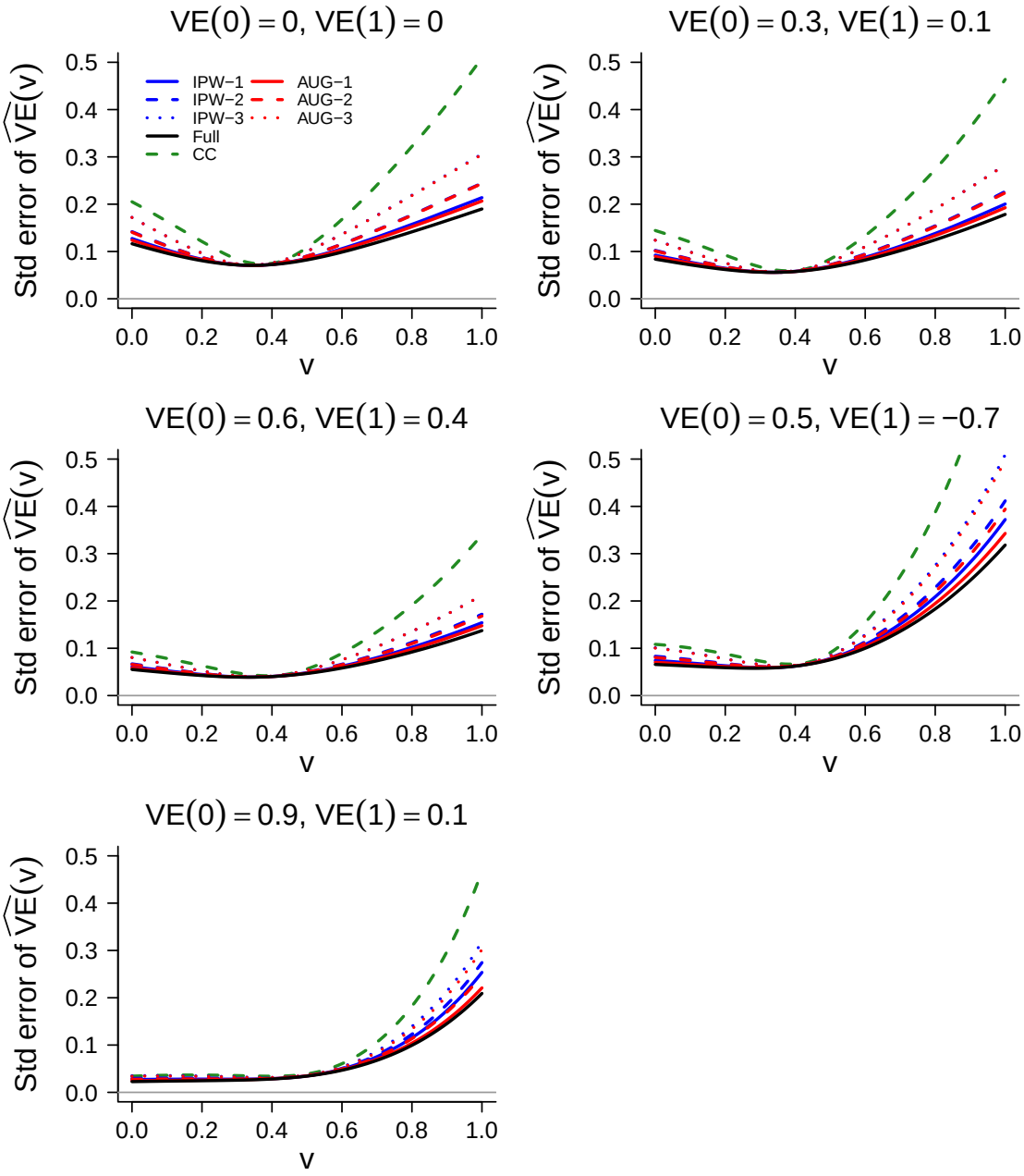


Figure 5.2: Median asymptotic standard error estimates of $\widehat{VE}(v)$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L1) and the expected number of 400 failures in the placebo group. Three AUG and IPW estimators are evaluated: AUG-1 and IPW-1 corresponding to $\rho \approx 0.98$, AUG-2 and IPW-2 for $\rho \approx 0.92$, and AUG-3 and IPW-3 for $\rho \approx 0.76$.

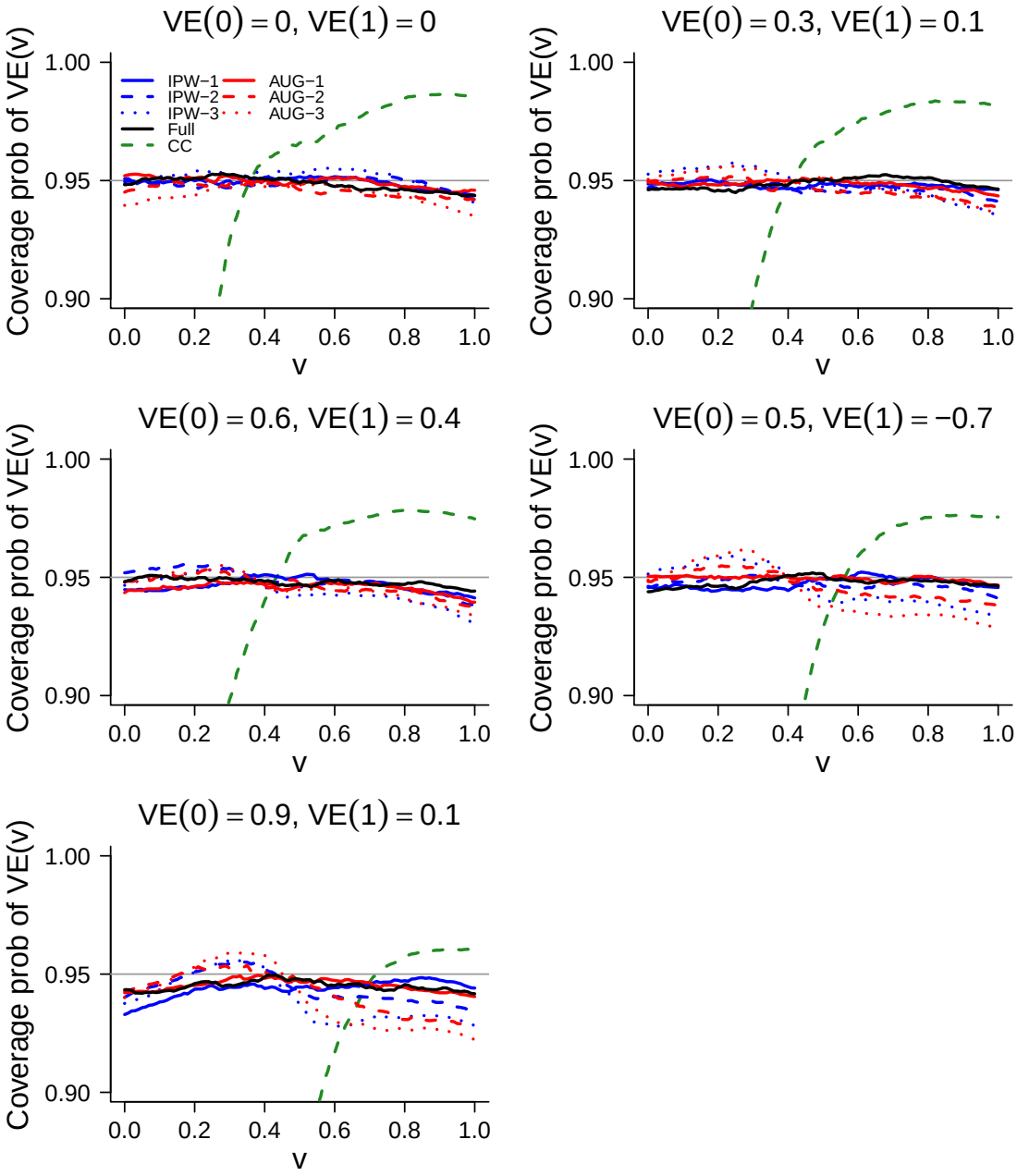


Figure 5.3: Coverage probabilities of 95% pointwise confidence intervals for $VE(v) = 1-e^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L1) and the expected number of 400 failures in the placebo group. Three AUG and IPW estimators are evaluated: AUG-1 and IPW-1 corresponding to $\rho \approx 0.98$, AUG-2 and IPW-2 for $\rho \approx 0.92$, and AUG-3 and IPW-3 for $\rho \approx 0.76$.

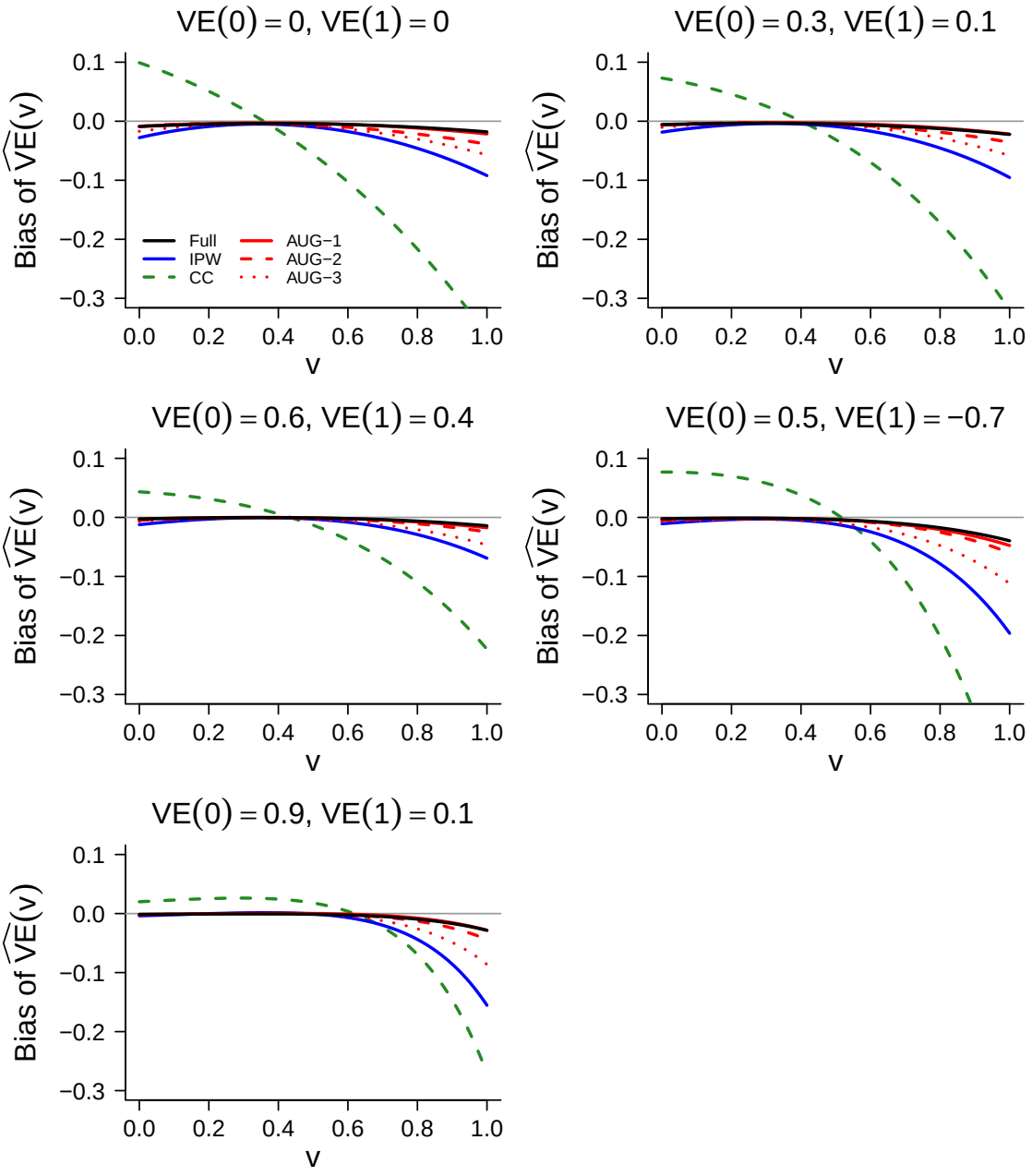


Figure 5.4: Bias of estimation for $VE(v) = 1 - e^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L2) and the expected number of 400 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

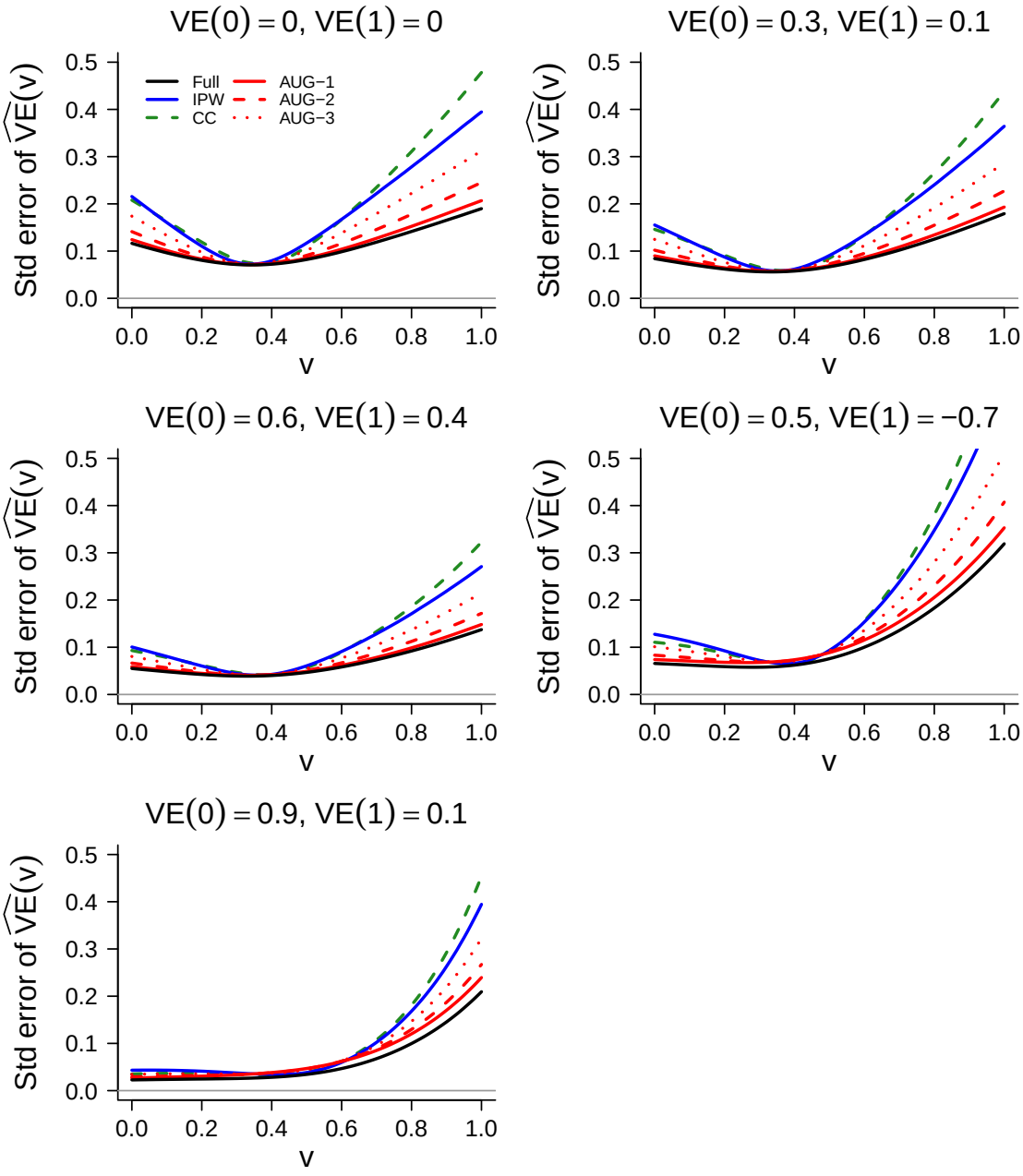


Figure 5.5: Median asymptotic standard error estimates of $\widehat{VE}(v)$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L2) and the expected number of 400 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

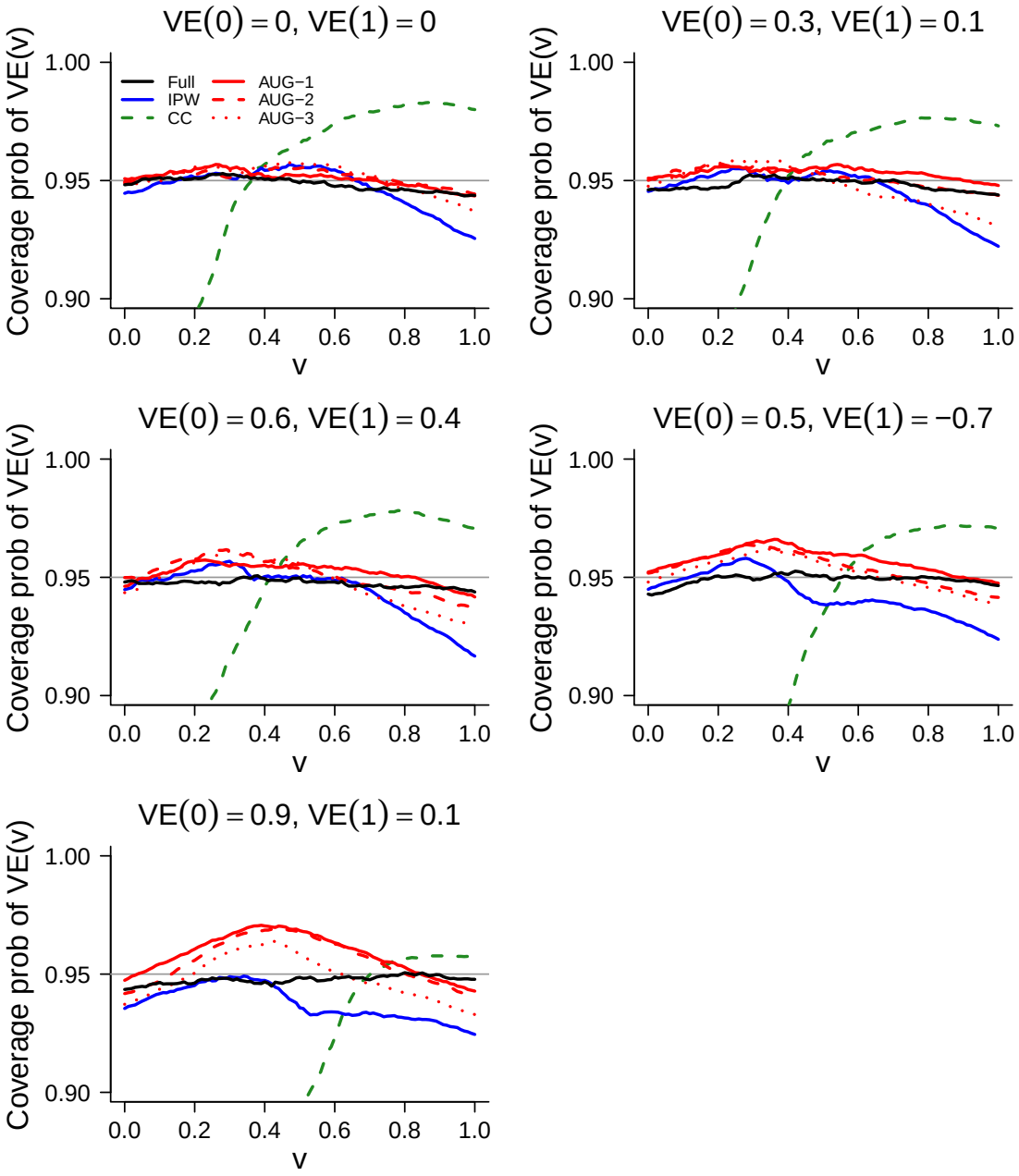


Figure 5.6: Coverage probabilities of 95% pointwise confidence intervals for $VE(v) = 1-\mathrm{e}^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L2) and the expected number of 400 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

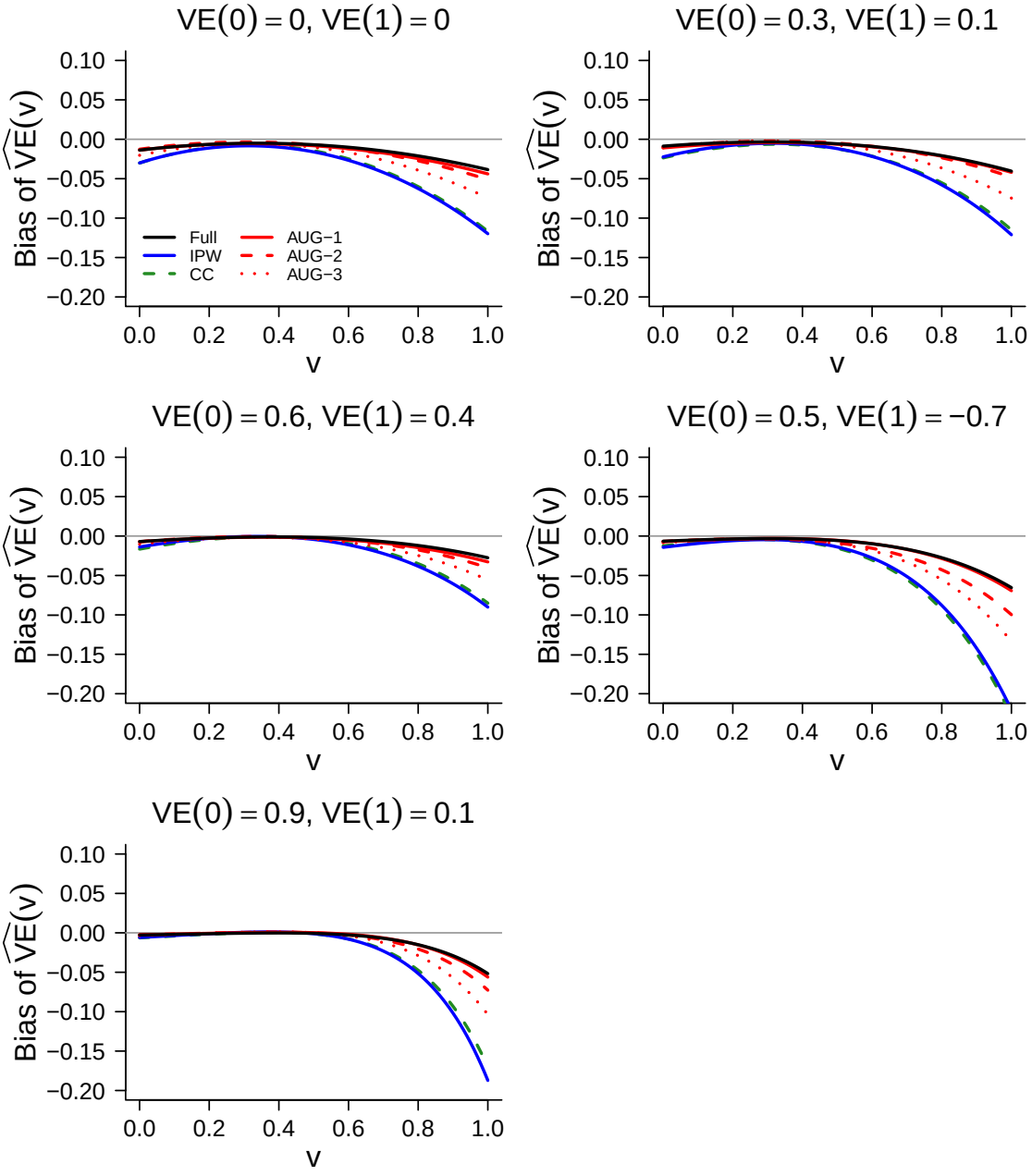


Figure 5.7: Bias of estimation for $VE(v) = 1 - e^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L3) and the expected number of 200 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

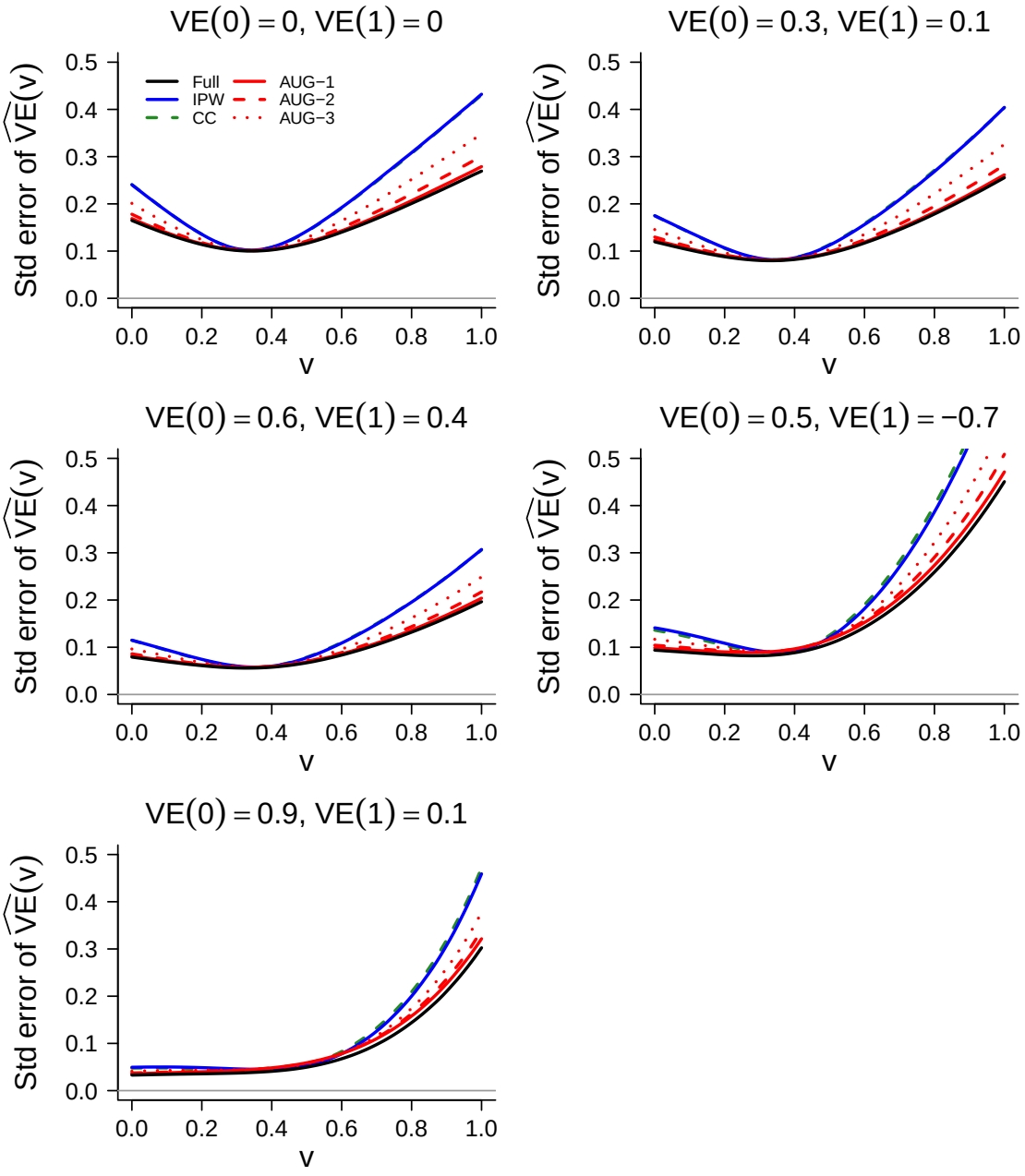


Figure 5.8: Median asymptotic standard error estimates of $\widehat{VE}(v)$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L3) and the expected number of 200 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

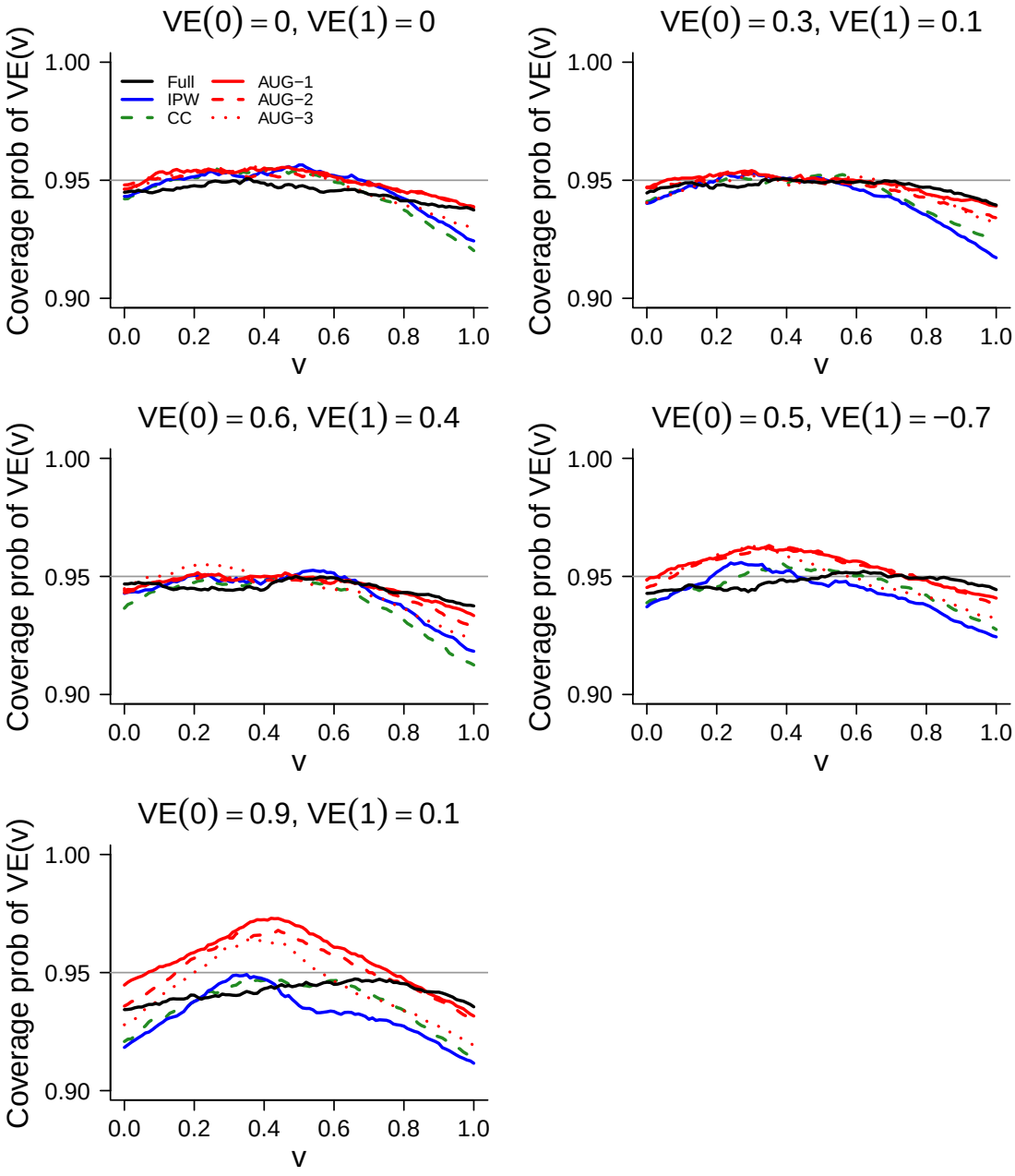


Figure 5.9: Coverage probabilities of 95% pointwise confidence intervals for $VE(v) = 1-\mathrm{e}^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for the missingness model (L3) and the expected number of 200 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

Table 5.1: Bias of estimation for β in the density ratio model (2.6) using four procedures - Full, IPW, CC and AUG - under correctly specified models (L1), (L2), and (L3) for the probability of a missing mark. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$. For model (L1), the IPW estimator considers the case $\rho \approx 0.98$.

Missingness	DR	Bias					
model	model	Full	AUG-1	AUG-2	AUG-3	IPW	CC
(L1)	(M1)	-0.023	-0.005	0.010	0.017	0.005	0.456
	(M2)	-0.002	-0.011	0.013	0.042	-0.008	0.416
	(M3)	-0.017	-0.017	-0.009	-0.012	0.003	0.488
	(M4)	0.011	0.009	0.016	0.023	0.009	0.454
	(M5)	-0.015	0.020	0.018	0.059	0.053	0.516
(L2)	(M1)	-0.023	0.010	0.004	0.010	0.030	0.397
	(M2)	-0.007	0.009	0.006	0.037	-0.002	0.377
	(M3)	0.005	-0.010	0.002	-0.003	-0.027	0.387
	(M4)	0.022	-0.015	0.016	0.034	0.066	0.410
	(M5)	0.004	-0.009	0.059	0.040	0.066	0.446
(L3)	(M1)	-0.023	-0.004	0.015	-0.007	0.019	0.016
	(M2)	0.012	-0.017	0.000	0.018	0.026	0.009
	(M3)	0.034	0.007	0.018	0.005	-0.012	-0.004
	(M4)	0.027	0.012	0.022	0.003	0.028	0.011
	(M5)	0.034	0.027	0.001	0.019	0.079	0.056

Table 5.2: Relative efficiency of estimation for β in the density ratio model (2.6) using four procedures - Full, IPW, CC and AUG - under correctly specified models (L1), (L2), and (L3) for the probability of a missing mark. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$. For model (L1), the IPW estimator considers the case $\rho \approx 0.98$.

Missingness	DR	Relative efficiency					
model	model	Full	AUG-1	AUG-2	AUG-3	IPW	CC
(L1)	(M1)	1.000	1.213	1.741	2.943	1.326	5.247
	(M2)	1.000	1.194	1.695	2.855	1.306	5.140
	(M3)	1.000	1.173	1.634	2.668	1.317	4.922
	(M4)	1.000	1.192	1.672	2.823	1.402	5.302
	(M5)	1.000	1.167	1.532	2.482	1.563	4.824
(L2)	(M1)	1.000	1.223	1.772	2.975	4.972	5.170
	(M2)	1.000	1.202	1.724	2.884	4.897	4.991
	(M3)	1.000	1.181	1.654	2.732	4.658	4.741
	(M4)	1.000	1.181	1.685	2.849	4.924	5.163
	(M5)	1.000	1.160	1.570	2.497	4.401	4.591
(L3)	(M1)	1.000	1.079	1.283	1.794	2.868	2.830
	(M2)	1.000	1.072	1.267	1.777	2.801	2.770
	(M3)	1.000	1.071	1.251	1.720	2.743	2.720
	(M4)	1.000	1.072	1.267	1.734	2.832	2.844
	(M5)	1.000	1.059	1.222	1.632	2.682	2.671

Table 5.3: Coverage probabilities of 95% confidence intervals for β in the density ratio model (2.6) using four procedures - Full, IPW, CC and AUG - under correctly specified models (L1), (L2), and (L3) for the probability of a missing mark. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$. For model (L1), the IPW estimator considers the case $\rho \approx 0.98$.

Missingness	DR	Coverage probability of 95% CI					
model	model	Full	AUG-1	AUG-2	AUG-3	IPW	CC
(L1)	(M1)	0.952	0.945	0.947	0.959	0.956	0.896
	(M2)	0.953	0.953	0.948	0.952	0.962	0.916
	(M3)	0.955	0.946	0.951	0.951	0.955	0.912
	(M4)	0.941	0.940	0.944	0.958	0.957	0.913
	(M5)	0.948	0.961	0.957	0.939	0.938	0.926
(L2)	(M1)	0.952	0.952	0.940	0.965	0.948	0.901
	(M2)	0.949	0.965	0.948	0.934	0.952	0.923
	(M3)	0.946	0.965	0.930	0.951	0.949	0.926
	(M4)	0.936	0.963	0.944	0.939	0.947	0.920
	(M5)	0.955	0.939	0.942	0.934	0.932	0.929
(L3)	(M1)	0.947	0.955	0.953	0.946	0.967	0.951
	(M2)	0.942	0.961	0.940	0.952	0.953	0.959
	(M3)	0.954	0.948	0.956	0.953	0.948	0.944
	(M4)	0.963	0.948	0.938	0.944	0.946	0.945
	(M5)	0.946	0.958	0.946	0.948	0.947	0.945

Table 5.4: Size of Wald tests of H_0^0 and H_0 using four procedures - Full, IPW, CC and AUG - under correctly specified models (L1), (L2), and (L3) for the probability of a missing mark. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$. For model (L1), the IPW estimator considers the case $\rho \approx 0.98$.

Missingness	Estimator					
model	Full	AUG-1	AUG-2	AUG-3	IPW	CC
	Size of Wald test of H_0^0 *(*$\alpha = 0.05$*)*					
(L1)	0.042	0.059	0.053	0.046	0.035	0.086
(L2)	0.042	0.048	0.046	0.041	0.057	0.082
(L3)	0.042	0.058	0.053	0.048	0.038	0.042
	Size of weighted Wald-type test of H_0^0 *(*$\alpha = 0.025$*)*					
(L1)	0.017	0.026	0.025	0.022	0.024	0.040
(L2)	0.017	0.028	0.016	0.019	0.020	0.034
(L3)	0.017	0.028	0.026	0.021	0.024	0.020
	Size of Wald test of H_0 *(*$\alpha = 0.05$*)*					
(L1)	0.048	0.055	0.053	0.041	0.041	0.087
(L2)	0.048	0.048	0.060	0.035	0.051	0.088
(L3)	0.048	0.050	0.063	0.046	0.046	0.045

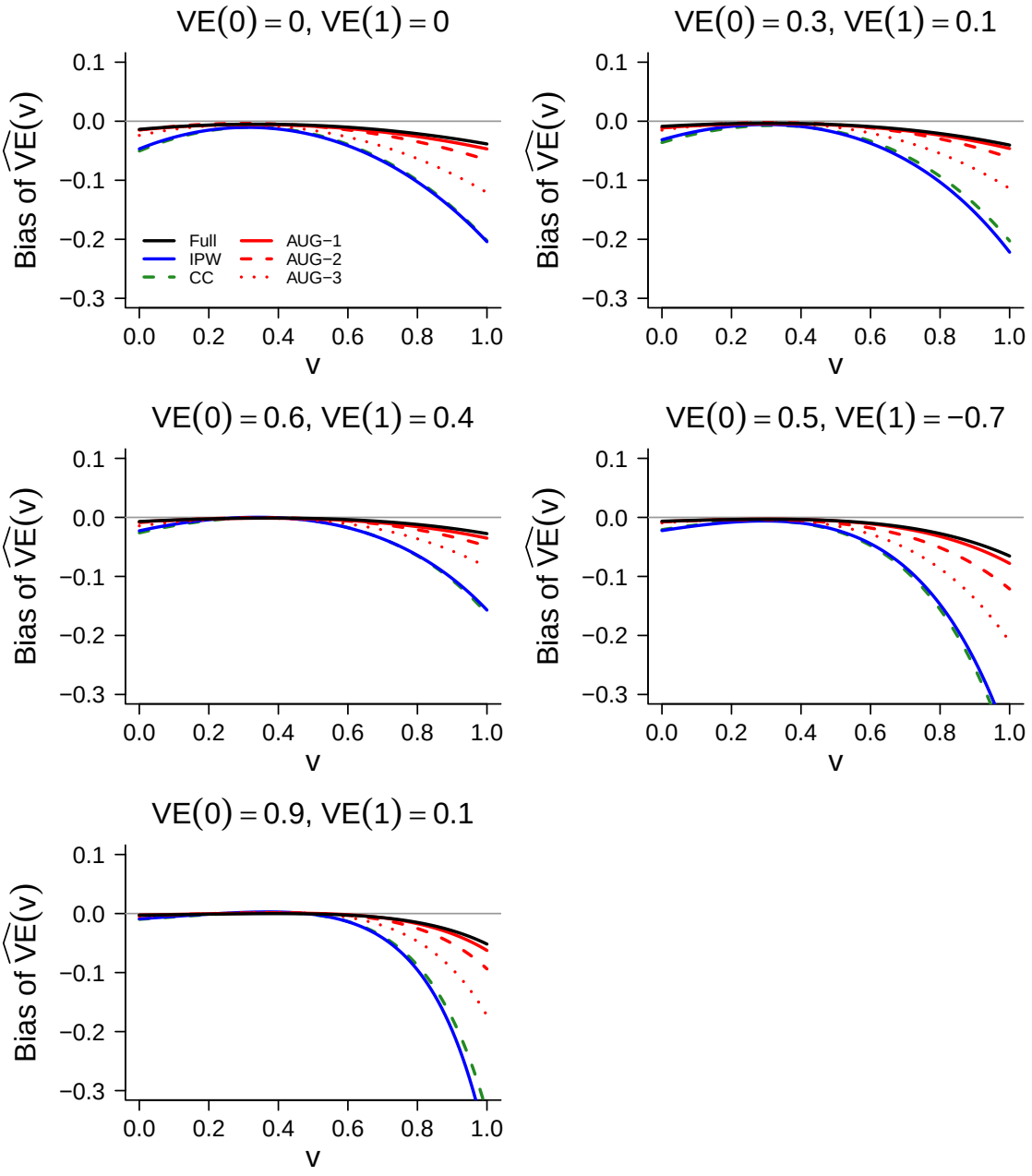


Figure 5.10: Bias of estimation for $VE(v) = 1 - e^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for a mis-specified missingness model following (L4) and the expected number of 200 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

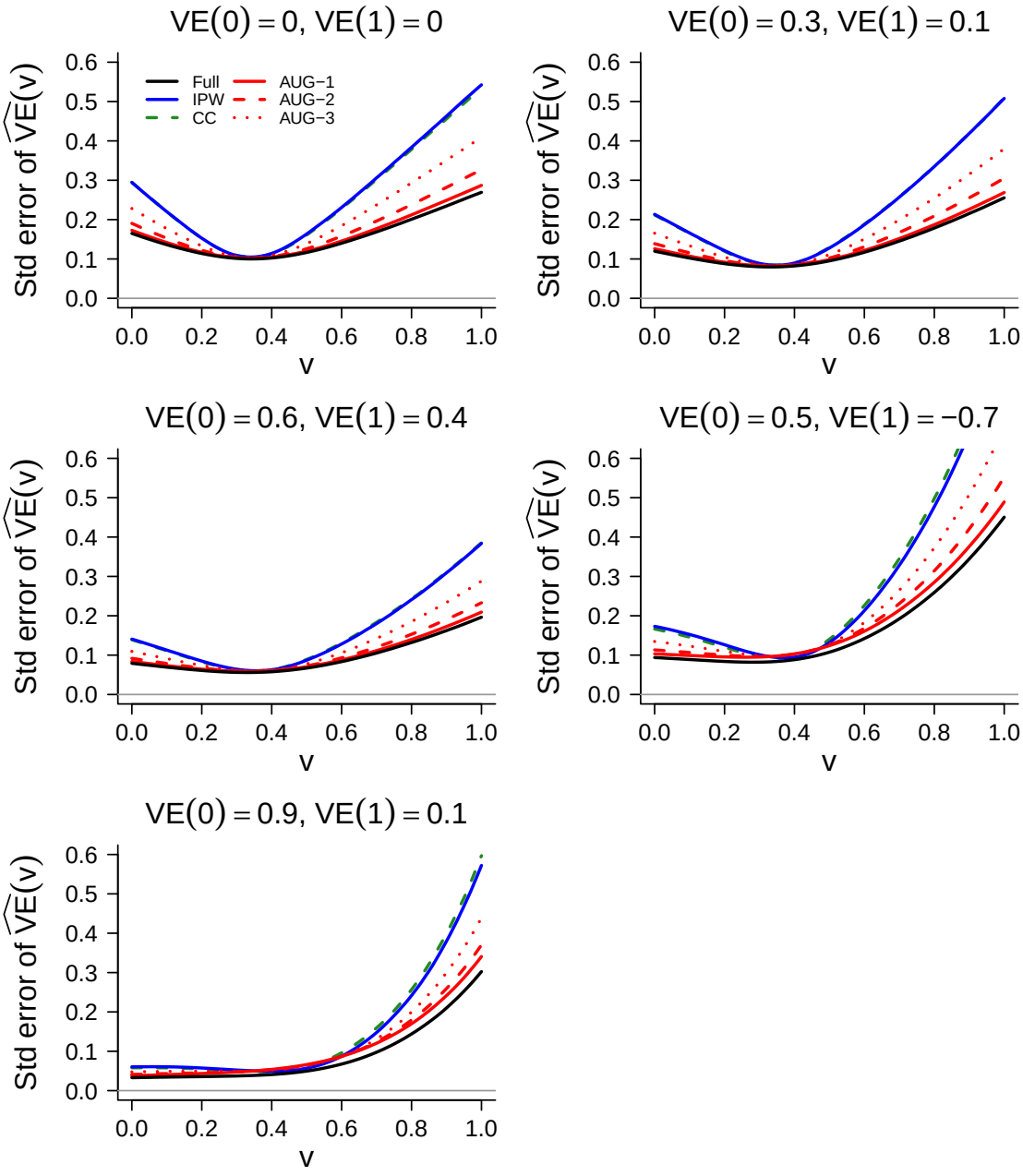


Figure 5.11: Median asymptotic standard error estimates of $\widehat{VE}(v)$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - for a mis-specified missingness model following (L4) and the expected number of 200 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

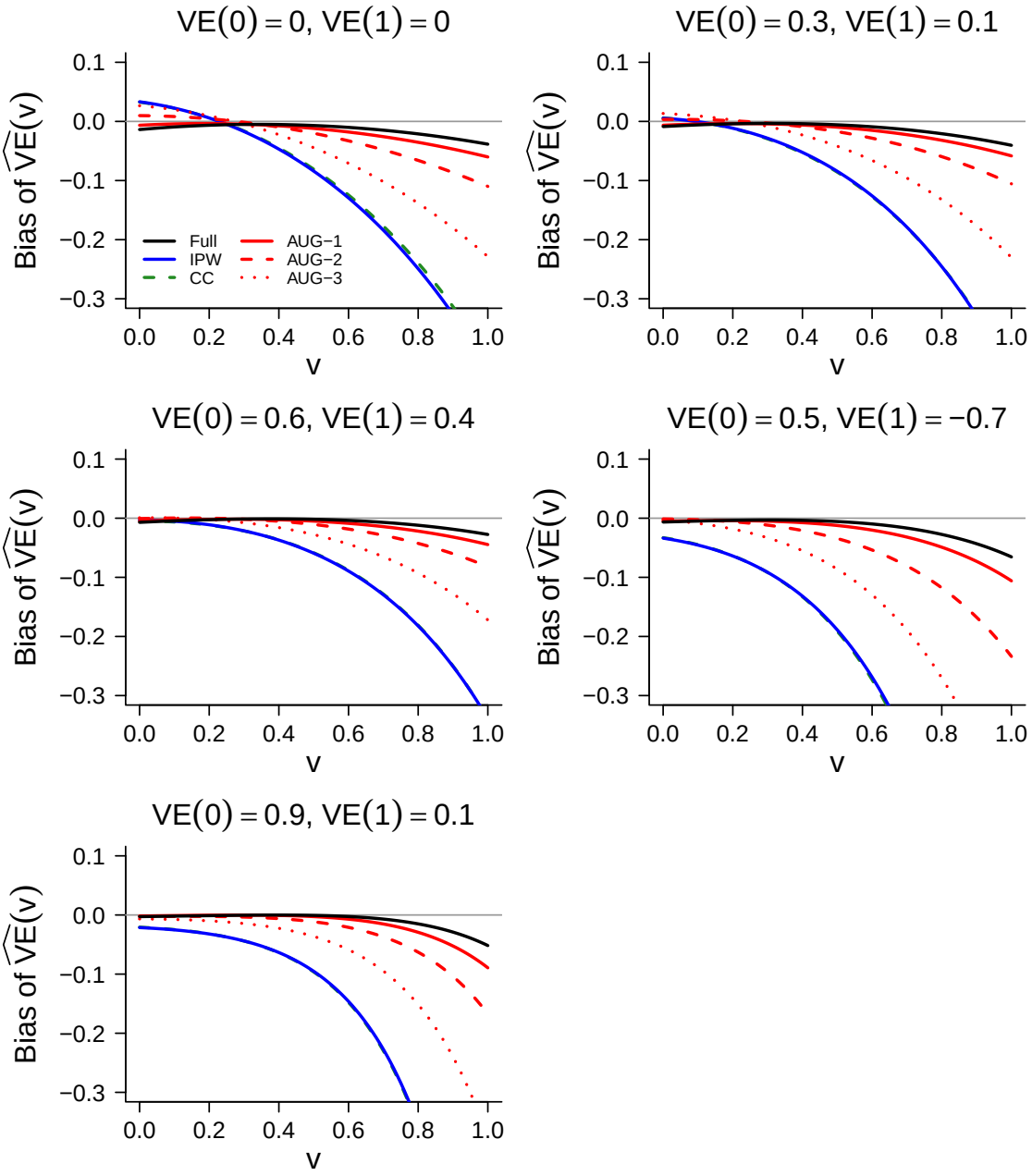


Figure 5.12: Bias of estimation for $VE(v) = 1 - e^{\alpha+\beta v+\gamma}$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - when $\pi(W, \psi)$ depends on V following (L5). The expected number of failures in the placebo group is 200. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

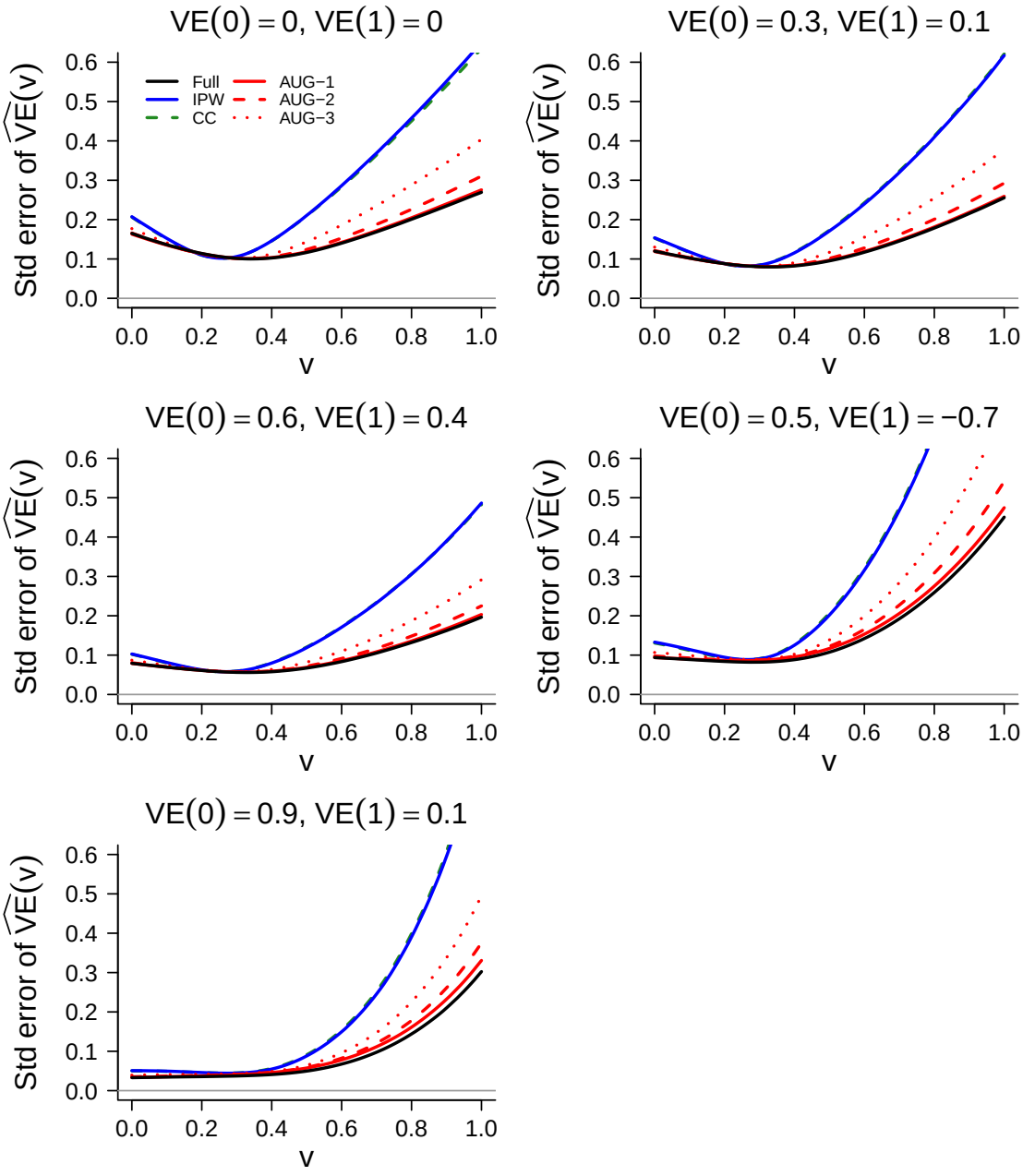


Figure 5.13: Median asymptotic standard error estimates of $\widehat{VE}(v)$ in models (M1)–(M5) using four procedures - Full, IPW, CC and AUG - when $\pi(W, \psi)$ depends on V following (L5). The expected number of failures in the placebo group is 200. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

Table 5.5: Bias of estimation for β in the density ratio model (2.6) using four procedures - Full, IPW, CC and AUG - for mis-specified missingness models following (L4) and (L5), and the expected number of 200 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

Missingness	DR	Bias					
model	model	Full	AUG-1	AUG-2	AUG-3	IPW	CC
(L4)	(M1)	-0.023	0.000	0.009	0.006	-0.044	0.011
	(M2)	-0.004	-0.005	-0.002	0.039	0.004	0.022
	(M3)	0.013	-0.014	0.011	-0.006	0.020	-0.020
	(M4)	-0.011	-0.005	0.019	0.054	0.078	0.073
	(M5)	-0.001	0.036	0.018	0.036	0.052	0.163
(L5)	(M1)	-0.023	0.016	0.086	0.155	0.249	0.224
	(M2)	0.012	0.010	0.067	0.162	0.262	0.250
	(M3)	0.034	0.026	0.088	0.140	0.205	0.184
	(M4)	0.027	0.037	0.090	0.157	0.249	0.235
	(M5)	0.034	0.067	0.088	0.212	0.334	0.259

Table 5.6: Relative efficiency of estimation for β in the density ratio model (2.6) using four procedures - Full, IPW, CC and AUG - for mis-specified missingness models following (L4) and (L5), and the expected number of 200 failures in the placebo group. Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

Missingness	DR	Relative efficiency					
model	model	Full	AUG-1	AUG-2	AUG-3	IPW	CC
(L4)	(M1)	1.000	1.154	1.550	2.516	4.589	4.595
	(M2)	1.000	1.145	1.519	2.467	4.535	4.495
	(M3)	1.000	1.118	1.485	2.406	4.404	4.437
	(M4)	1.000	1.135	1.514	2.438	4.593	4.560
	(M5)	1.000	1.116	1.431	2.270	4.338	4.402
(L5)	(M1)	1.000	1.013	1.206	1.768	3.820	3.720
	(M2)	1.000	1.007	1.183	1.760	3.742	3.659
	(M3)	1.000	1.013	1.181	1.708	3.645	3.618
	(M4)	1.000	1.013	1.204	1.764	3.758	3.797
	(M5)	1.000	1.034	1.204	1.727	3.614	3.579

Table 5.7: Size of Wald tests of H_0^0 and H_0 using four procedures - Full, AUG, IPW and CC - for mis-specified missingness models following (L4) and (L5). Three AUG estimators are evaluated: AUG-1 corresponding to $\rho \approx 0.98$, AUG-2 for $\rho \approx 0.92$, and AUG-3 for $\rho \approx 0.76$.

Missingness model	Estimator					
	Full	AUG-1	AUG-2	AUG-3	IPW	CC
	Size of Wald test of H_0^0 *(*$\alpha = 0.05$*)*					
(L4)	0.042	0.062	0.055	0.051	0.043	0.038
(L5)	0.042	0.067	0.071	0.087	0.059	0.053
	Size of weighted Wald-type test of H_0^0 *(*$\alpha = 0.025$*)*					
(L4)	0.017	0.026	0.022	0.023	0.022	0.024
(L5)	0.017	0.028	0.027	0.029	0.028	0.029
	Size of Wald test of H_0 *(*$\alpha = 0.05$*)*					
(L4)	0.048	0.048	0.064	0.037	0.047	0.035
(L5)	0.048	0.063	0.080	0.101	0.072	0.065

Chapter 6

RV144 V1/V2-FOCUSED SIEVE ANALYSIS

6.1 Introduction

The RV144 Phase III HIV-1 vaccine trial of the ALVAC-HIV (vCP1521) and AIDSVAX B/E prime-boost vaccine regimen, introduced in Section 1.2, showed the estimated marginal vaccine efficacy of 31.2% (95% CI, 1.1 to 52.1; $p = 0.04$) to prevent HIV-1 infection within 42 months following the first vaccination (Rerks-Ngarm et al., 2009). A subsequent case-control study (Haynes et al., 2012), conducted to identify immune correlates of HIV-1 infection risk in RV144 vaccine recipients, showed that IgG antibodies binding to the first and second variable loops (V1/V2) in the gp120 region of the Env protein scaffolded to the murine leukemia virus gp70 (Wang et al., 1995) associated with decreased infection rate. Antibody response levels, however, are not randomized among vaccine recipients which raises the possibility that the V1/V2-targeted antibody response is merely a correlate of infection risk that is not predictive of vaccine protection: this phenomenon could occur if, for instance, less exposed vaccine recipients were likely to have higher anti-V1/V2 antibody levels. Therefore, an independent study is necessary to evaluate the hypothesis that the vaccine-induced anti-V1/V2 antibody response provides a mechanism of vaccine protection. The correlates analysis compared infected vaccine recipients to uninfected vaccine recipients. Placebo recipients were not included because they had no vaccine-induced immune responses. To demonstrate protection, a comparison of vaccine recipients to placebo recipients is needed which renders sieve analysis an integral part in the assessment of immune correlates. Edlefsen et al. (2011) conducted a site-specific RV144 sieve analysis and identified two amino acid sites in and near the crown of the V2 loop that

significantly differ between vaccine and placebo sequences. In this chapter, we present a V1/V2-focused genetic distance-based RV144 sieve analysis to help discern the potential role of V1/V2 antibody epitopes as an immunologic cause of the observed partial vaccine efficacy.

6.2 Genetic distance definition

Full length HIV-1 genomes were measured from 121 of the 125 infected subjects; 3 are missing data because their viral load was too low for the Sanger sequencing technology to work, and 1 dropped out. The 4 subjects with missing data are excluded from the analysis. Three gp120 sequences were included in the vaccine: 92TH023 in the ALVAC canarypox prime, and A244, MN in the AIDSVAX gp120 boost. 92TH023 and A244 are subtype E sequences whereas MN is subtype B, and 110 (91%) of the 121 subjects were infected with subtype E viruses (44 vaccine recipients and 66 placebo recipients). The subtype E vaccine insert sequences are closer genetically to the infecting sequences than MN, and thus are more likely to stimulate protective immune responses. Accordingly, the analysis of mark-specific vaccine efficacy presented here focuses on the 92TH023 and A244 insert sequences, and excludes the 11 subjects with non-E infecting sequences.

To maximize biological relevance and statistical power of the analysis, it is important to focus the V1/V2 genetic distances by pre-screening amino acid sites based on external or treatment-blinded criteria for alignability, variability, and immunogenicity (the V1/V2 region spans 85 amino acid (AA) residues of HXB2 [AA120 to 204]). After the nucleotide sequences isolated from the plasma of the 110 subtype-E-infected subjects were codon-aligned and translated, there remained uncertainty about accurate alignment at some positions which therefore were excluded from the distance definition. Additionally, positions at which the number of vaccine insert amino acids among isolated sequences was too low (three or fewer) or too high (all except three or more) were excluded because such positions are unlikely to be involved

in observing potential sieve effects. Of the remaining amino acid sites, we focused the V1/V2 distances on sites that were potentially in antibody epitopes, and, moreover, were in antibody epitopes to which vaccine recipients were able to mount antibody responses. Accordingly, we considered the published set of V1/V2 sites that were contact residues for monoclonal antibodies (data provided by Ivelin Georgiev, Peter Kwong, Robin Stanfield and Ian Wilson). Moreover, we considered the subset of published V1/V2 sites that potentially could be in antibody epitopes as indicated by neutralization sensitivity assays (Wei et al., 2003; Moore et al., 2009; Tomaras et al., 2011). Then, we restricted attention to the subset of V1/V2 sites that were in one or the other of these two "antibody epitope relevant" sets. Next, so as to further restrict attention to V1/V2 sites in epitopes to which vaccine recipients could mount antibody responses, a linear peptide binding microarray analysis was conducted (Haynes et al., 2012). Based on baseline and post-immunization RV144 samples from 80 uninfected vaccine recipients and 20 uninfected placebo recipients, antibody binding reactivity was measured to each of 1453 15-mer gp120 peptides, which were tiled across gp120 overlapping by 11 amino acids. This yielded a set of 33 sites in the V1/V2 region where there was significant binding reactivity (based on a testing procedure using false discovery rate adjustment). Finally, the set of V1/V2 sites for analysis was defined as the intersection of this set with the first set described above, which has $q = 9$ sites with HXB2 coordinates 120, 124, 125, 165, 166, 168, 169, 171, and 181.

For each of the two subtype E insert reference sequences and each sequence from a subject, the genetic distance was computed using the Blosum90 scoring matrix, adjusted such that the addition or subtraction of an N-linked glycosylation site is scored 3-fold or 1.5-fold higher than the largest Blosum90 score depending on whether the mismatch is in position 2 of the amino acid triple comparing the N-linked glycosylation site. First, a distance was computed as

$$d(seq_{\text{subject}}, seq_{\text{reference}}) = \sum_{i=1}^{q} \frac{2s(a_i^s, a_i^r)}{s(a_i^s, a_i^s) + s(a_i^r, a_i^r)},$$

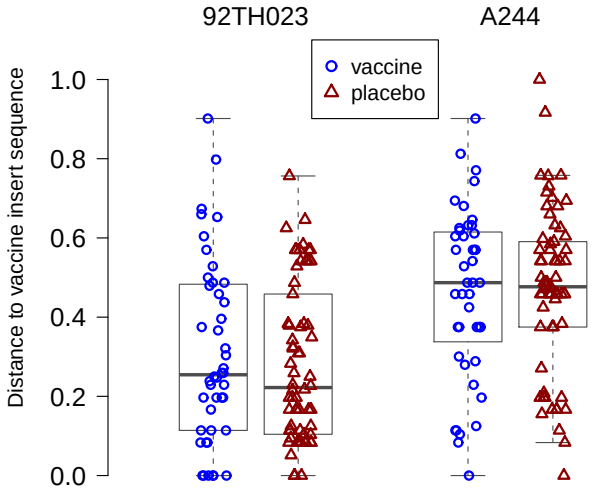
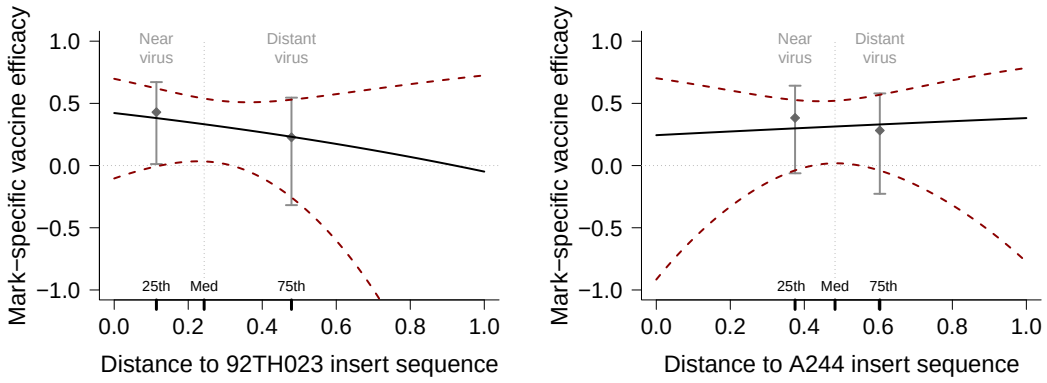


Figure 6.1: RV144 trial: distribution of the V1/V2 distances to the 92TH023 and A244 vaccine insert sequence by vaccine/placebo group.

where a_i^s is the amino acid at site i in the subject's sequence, a_i^r is the amino acid at site i in the reference sequence, and $s(x, y)$ is the score from the modified Blosum90 matrix. Second, the distances were re-scaled to values between 0 and 1 (2 identical sequences have distance 0). Between 2 and 13 sequences (total 1030 sequences) were measured per infected subject, and the final distance was defined as the subject-specific median distance. This yields a data set of a bivariate mark (V_1, V_2) measured from 110 infected subjects, with 49 and 52 unique mark values in V_1 and V_2, respectively. Figure 6.1 shows the distributions of V_1 and V_2 by vaccine/placebo group, and shows a scatterplot of the two distances.

6.3 Inference about mark-specific vaccine efficacy

The failure time variable specified in the presented analysis is the estimated time to HIV-1 infection, a co-primary endpoint in RV144, defined as the midpoint between the last negative and the first positive PCR-confirmed test result. We consider it



(a) (b)

Figure 6.2: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for the univariate distance to (a) the 92TH023 insert sequence, and (b) the A244 insert sequence. The dichotomous-mark estimates following Prentice et al. (1978) are included for comparison.

implausible that RV144 participants might have affected each other's time to HIV-1 infection.

6.3.1 Complete mark data

Initially, we consider a complete set of marks, computed for each subject based on the available sequences, disregarding the post-infection time-points at which the samples were obtained. Using this mark definition, the estimates of $VE(v)$ with 95% pointwise confidence bands for univariate distances to the 92TH023 and A244 vaccine insert sequence are plotted in Figure 6.2. In addition, for the comparison purpose with the Prentice et al. (1978) approach, we dichotomize the continuous marks by discriminating infections with 'near' $[V \leq \text{med}(V)]$ versus 'distant' $[V > \text{med}(V)]$ viruses. Figure 6.2 includes estimates with 95% confidence intervals for vaccine efficacy to prevent near- and distant-virus infections (see Table 6.1 for complete inference).

Table 6.1: Dichotomized mark: inference for RV144 virus type-specific vaccine efficacy via the Prentice et al. (1978) approach

Vaccine insert	Virus type	$VE^{\sharp}_{\text{type}}$	95% CI$^{\sharp}$	p-value	
				$VE_{\text{type}}=0^{\dagger}$	$VE_{\text{near}}=VE^{\ddagger}_{\text{distant}}$
92TH023	near	42.9	11.6 to 74.3	0.043	0.568
	distant	22.7	−18.5 to 63.9	0.343	
A244	near	38.3	4.8 to 71.9	0.079	0.837
	distant	28.2	−10.3 to 66.7	0.223	

$^{\sharp}$ in % † score test ‡ Lunn and McNeil (1995) test

The weighted Wald test of $H_0^0 : VE(v) \equiv 0$ yields p-values of 0.016 and 0.024 for the 92TH023 and A244 distances, respectively. The likelihood ratio test of $H_0 : VE(v) = VE$ yields $p = 0.52$ for the 92TH023 and $p = 0.83$ for the A244 distance. These p-values are consistent with the broad confidence bands in Figures 6.2a and 6.2b, and they agree with the p-values using the Lunn and McNeil (1995) test for the dichotomous mark (Table 6.1). The RV144 trial lacked sufficient power to detect a change in $VE(v)$ with viral divergence. The assumption of conditional independence between T and V given Z was verified using the Kolmogorov–Smirnov-type test, described in Section 2.6.1, with the p-value of 0.88 in the placebo and 0.98 in the vaccine group based on 1000 bootstrap iterations. The Qin and Zhang (1997) goodness-of-fit test does not reject the validity of the mark density ratio model for either the 92TH023 ($p = 0.49$) or the A244 distance ($p = 0.51$). Moreover, a test of proportional hazards assumed in (2.17) (Grambsch and Therneau, 1994) does not reject ($p = 0.21$).

Figure 6.3 displays $\widehat{VE}(v_1, v_2)$ for the bivariate mark (V_1, V_2). The weighted Wald test of H_0^0 yields $p = 0.019$. The likelihood ratio test of H_0 yields $p = 0.81$. The p-value for the test of conditional independence between T and (V_1, V_2) given Z, using the

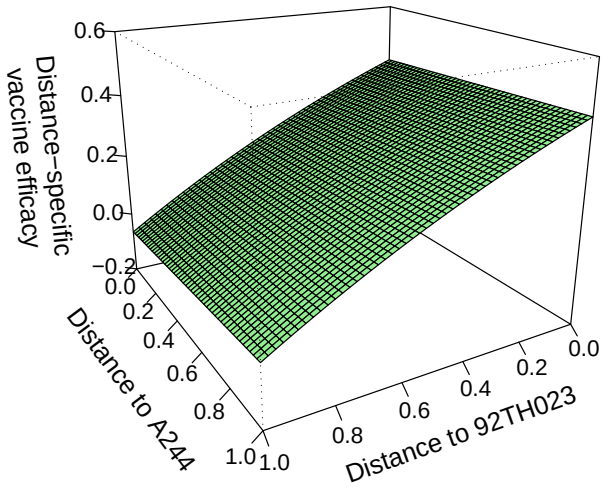


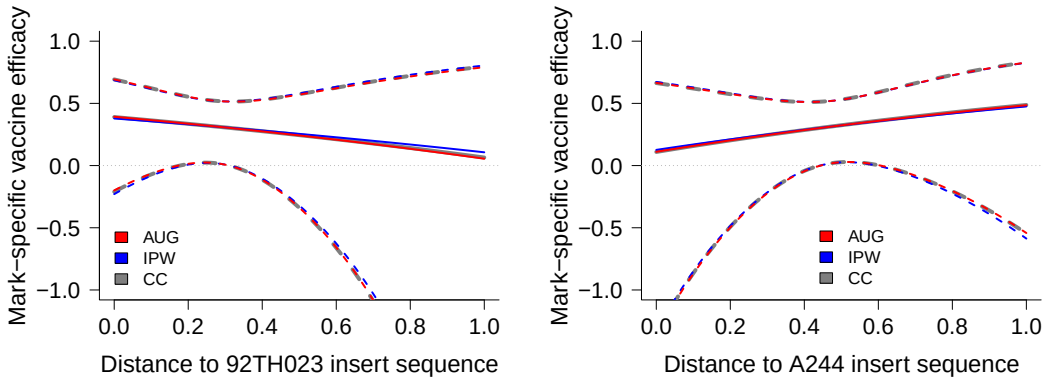
Figure 6.3: The estimated mark-specific vaccine efficacy for the bivariate mark consisting of the distance to the 92TH023 and A244 insert sequence.

Kolmogorov–Smirnov-type test, is 0.64 in the placebo and 0.89 in the vaccine group, again supporting the appropriateness of the method. The goodness-of-fit test does not reject the validity of the bivariate-mark density ratio model ($p = 0.46$).

6.3.2 Incomplete mark data

Henceforth we consider the marks V_1 and V_2 defined based on sequences measured from plasma samples drawn at the time of HIV-1 diagnosis. Twelve (11%) of the 110 subtype-E-infected subjects have sequences isolated from samples collected at post-diagnosis visits, and thus the marks V_1 and V_2 are considered to be missing for these subjects. This mark definition 'purifies' the analysis of vaccine efficacy against HIV-1 acquisition by filtering out potential post-diagnosis effects of the vaccine on the viral evolution, and thus is of greater scientific relevance.

To predict the probability of sequencing the virus at diagnosis, we used a logistic regression model adjusted for viral load measured at the diagnosis visit and the



(a) (b)

Figure 6.4: Estimated mark-specific vaccine efficacy using three procedures – AUG, IPW and CC – with 95% pointwise confidence bands for the univariate distance to (a) the 92TH023 insert sequence, and (b) the A244 insert sequence.

calendar year of diagnosis. To use the augmented IPW estimation method, we additionally specified a linear regression model for the expected profile scores adjusted for behavioral risk (low, medium, high), and age (≤ 20, 21–25, ≥ 26). The auxiliary covariates were selected from available RV144 measurements using the all-subsets model selection procedure carried out by the R `regsubsets` function.

Figure 6.4 shows point and interval estimates of $VE(v)$ using three estimation procedures – AUG, IPW and CC – for 92TH023 and A244 marks based on sequences measured from diagnosis-visit samples. The AUG (IPW) weighted Wald tests of $H_0^0 : VE(v) \equiv 0$ yield the p-values of 0.019 (0.019) for the 92TH023 and 0.027 (0.027) for the A244 distance. The AUG (IPW) Wald tests of $H_0 : VE(v) = VE$ yield $p = 0.75$ (0.73) for the 92TH023 and $p = 0.61$ (0.61) for the A244 distance. The missing-mark inference is comparable with that using the complete-case analysis which may be ascribed to the fact that the measured auxiliary variables are weak predictors of the probability of missing marks and the expected profile scores. This

observation emphasizes the importance of collecting longitudinal sequence data in the next conducted HIV-1 vaccine trial, which, together with the associated measurement times, would allow for improvement of the prediction models. In addition, the low missing mark rate may also be a contributing factor to the comparable performance of the complete-case analysis.

6.4 Exploratory sieve analysis using other V1/V2 and gp120 distances

The specification of RV144 genetic distances is the result of an extensive collaborative effort. Despite the existing scientific guidance in this process, we lack a way to prioritize which biological framework for determining the distance definition should be primary and which ones should be relegated to an exploratory status. Given the insufficiency of scientific rationale for prioritizing the different biologically-relevant V1/V2 and Env gp120 distances, we analyze them all. To weight amino acid mismatches between a breakthrough and vaccine insert sequence, we consider both Blosum90 and HIV-specific PAM25 matrices. The following four sets of gp120 sites are used (individually and in combinations) to define the distances:

(i) published monoclonal antibody contact sites,

(ii) published antibody contact sites identified by neutralization sensitivity assays,

(iii) hotspots in the linear peptide binding microarray analysis, and

(iv) sites with the highest frequency of occurrence in predicted antibody epitopes based on structural biology.

Approaches (i)–(iii) are described in Section 6.2. Approach (iv) is based on structural predictions of antibody epitopes: a new method is used to predict thousands of antibody epitopes centered on gp120 exposed surface residues for the three Env proteins included in the vaccine. These predicted 'patches' of amino acids are used to rank the 85 amino acid sites (that constitute the V1/V2 region) by their likelihood of being vaccine-induced antibody targets, computed as the sum over the prime and the

maximum of the two boost inserts of the percentage of patches including the residue in the insert. With this approach that captures conformational antibody targets, top-ranking scores are found for 22 V1/V2 sites that are used for analysis (data provided by William Schief, Chris Carrico, and Sergey Menis).

The exploratory sieve analysis results are presented in Appendix B. All analyses show that H_0^0 is rejected with p-value < 0.05 using the weighted Wald test, consistent with the original RV144 primary analysis on overall vaccine efficacy. In addition, for the 92TH023 and A244 insert sequences, the results tend to show that the estimate of $VE(v)$ decreases with the genetic distance v, consistent with greater protection against viruses closer to the insert antigens and waning protection with the degree of mismatch. However, the tests of H_0 (i.e., for a sieve effect) are never significant at the $p < 0.05$ level using the likelihood ratio test, although the results for whole-gp120 and set (i) of amino acid sites with Blosum90 weights trend toward significance (Figure B.14). The estimated $VE(v)$ curves change sharply enough in the genetic distance v that it seems feasible that there is an effect that is non-significant due to lack of a sufficiently large number of infections. In contrast to the 92TH023 and A244 insert sequences, the estimated $VE(v)$ curves tend to be constant or increase slightly for the MN insert sequence.

6.5 Discussion

Within-host HIV-1 evolution poses a challenge to the study of mark-specific vaccine effects on HIV-1 acquisition. To maximize information accuracy about the breakthrough viruses, it is desirable to collect sequence data as early as possible after the acquisition time and analyze marks based on such sequence data. In RV144, the fraction of sequences measured in the acute (PCR+/Ab−) phase of infection was too small to warrant a sieve analysis using acute-phase marks. The marks as defined in Section 6.3.2 absorb statistical noise introduced by viral evolution occurring between the acquisition and diagnosis time points. Consequently, the analysis presented above

may reflect a mixture of acquisition and early post-infection evolutionary effects of the vaccine.

In the distance definition in Section 6.2, we used the modified Blosum90 matrix which up-weights amino acid mismatches, relative to a vaccine insert sequence, which are relevant to antibody binding. The selective filtering of breakthrough viruses in vaccinated subjects is likely related to 'antibody-relevant' amino acid substitutions. For this reason, the Blosum90 matrix was chosen to maximize power for a sieve effect wherein the vaccine blocks infection with variants that match the insert in the antibody binding regions (in Appendix B, we compare the use of Blosum90 and PAM25 matrices).

Chapter 7

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This dissertation develops inferential procedures for assessing differential vaccine protection in HIV-1 vaccine efficacy trials wherein viral characteristics are summarized by a multivariate continuous mark variable. We show that the mark-specific hazard ratio factors as the mark density ratio and the ordinary marginal hazard ratio. This factorization is appealing because it enables the two components to be estimated separately. The semiparametric method of maximum profile likelihood estimation in the mark density ratio model yields improved efficiency of mark-specific hazard ratio estimation relative to current alternative approaches. We derive the joint asymptotic distribution of the estimators for the mark density ratio and the marginal hazard ratio. This result allows estimation of mark-specific vaccine efficacy ($VE(v)$) with pointwise confidence intervals and testing of hypotheses of interest about $VE(v)$. The proposed weighted Wald-type test is recommended for testing $VE(v) \equiv 0$ for all v, and the likelihood ratio test for testing $VE(v)$ invariant in v.

In practice, marks of greatest scientific relevance are commonly subject to missingness. We, therefore, extend the proposed inferential procedures for $VE(v)$ to accommodate missing, possibly multivariate, marks. To this end, we consider two approaches to parameter estimation in the mark density ratio model: (i) inverse probability weighting of the complete cases, and (ii) augmenting of the inverse probability weighted estimating functions by leveraging auxiliary data predictive of the mark. The augmented estimator possesses the attractive double robustness property.

Due to growing complexity in HIV-1 vaccine design, it becomes increasingly de-

sirable to evaluate $VE(v)$ as a function of a multivariate mark representing a set of genetic distances. Our approach allows flexible semiparametric modeling of a multivariate mark variable, and the estimation methods do not impose practical constraints on the mark dimension.

Although the methods are motivated by a specific scientific application (evaluation of the impact of HIV-1 genetic diversity on vaccine efficacy), they provide a general approach to the analysis of survival data in the presence of a continuous, possibly multivariate, mark variable. This dissertation also addresses the general problem of parameter estimation in density ratio models with missing data. The R code implementing the proposed methods is available upon request.

7.1.1 Other applications of sieve analysis in HIV vaccine research

Genetic distance selection and vaccine design

The importance of biological relevance in the selection of genetic distance measures used for sieve analysis is addressed in Chapter 1. If the mechanism of vaccine protection is only partially understood, a large number of meaningful distance measures may be proposed with the goal to identify distances that yield the steepest decline in $VE(v)$. Such distances motivate subsequent exploration to identify (i) aspects of the vaccine design that may account (at least partially) for protective effects against HIV strains with small distances, and (ii) deficiencies in the vaccine design that may account (at least partially) for the lack of protective effects against strains with large distances. These findings can be integrated in the design of a new vaccine candidate that, when evaluated in the identical population, may yield a mark distribution shifted downwards compared to that for the original vaccine. It is hoped that this design improvement strategy results in a greater overall vaccine efficacy.

Evaluation of immunological measurements as potential surrogates of protection

Sieve analysis may serve to evaluate the degree to which an immunological measurement is a valid surrogate of protection (Gilbert et al., 2011). To illustrate, suppose that elevated titers of neutralizing antibodies binding to a specific envelope (Env) region are shown to be associated with a decreased infection risk in vaccine recipients. If the antibody titer has surrogate value to predict vaccine efficacy, we would expect $VE(v)$ (with v focused on the target region in Env) to be the greatest against HIV strains that the antibodies can easily neutralize, i.e., strains that are genetically similar (in terms of v) to vaccine antigens inducing the generation of these antibodies. In addition, we would anticipate reduced $VE(v)$ against dissimilar strains in view of the fact that genetic mutations in antibody epitopes impede neutralization responses. Sieve analysis, therefore, is a useful tool for assessing the surrogate value of immunological measurements.

7.2 Future work

7.2.1 Elimination of the $T \perp\!\!\!\perp V|Z$ assumption: an alternative mark-specific vaccine efficacy model

An alternative factorization to (2.4) of the mark-specific hazard function is

$$\lambda(t, v|Z = z) = f(t, v|Z = z)\, S^{-1}(t|Z = z),$$

where $f(t, v|Z = z)$ is the conditional joint density function of $(T, V^T)^T$ given $Z = z$, and $S(t|Z = z)$ is the conditional marginal survival function of T given $Z = z$. Consequently, the mark-specific vaccine efficacy can be expressed as

$$VE(t, v) = 1 - \frac{f(t, v|Z = 1)}{f(t, v|Z = 0)} \frac{S(t|Z = 0)}{S(t|Z = 1)}. \tag{7.1}$$

For the (T, V)-density ratio, we consider the semiparametric model

$$\frac{f(t, v|Z = 1)}{f(t, v|Z = 0)} = g(t, v, \phi) \tag{7.2}$$

with a pre-specified weight function $g(\cdot, \cdot, \phi)$ indexed by an unknown vector parameter ϕ. The assumption $C \perp\!\!\!\perp (T, V)|Z$ implies that $f(t, v|Z = z) = f(t, v|Z = z, \delta = 1)$ (note that the $T \perp\!\!\!\perp V|Z$ assumption is not required for the equality to hold). Therefore, we can estimate ϕ using the Qin (1998) method of maximum profile likelihood estimation applied to the cohort of infected subjects only. Moreover, unlike the implication of Proposition 2.1, weight functions of the form

$$g(t, v, \phi) = \exp\{\alpha + \widetilde{g}(t, v, \beta)\}, \tag{7.3}$$

with $\phi = (\alpha, \beta^T)^T$ and $\widetilde{g}(t, v, \beta)$ a polynomial function in t and v, do not imply that $T \perp\!\!\!\perp V|Z$. Thus, if $g(t, v, \phi)$ satisfies (7.3), the parameter ϕ is identifiable irrespective of correlation between T and V within treatment groups.

Analogously to Section 2.4.1, the exponential form (7.3) is appealing because it yields model (7.2) that is equivalent to a retrospective logistic regression model specified as $\text{logit}\{P(Z = 1|T = t, V = v, \delta = 1)\} = \alpha^* + \widetilde{g}(t, v, \beta)$ where $\alpha = \alpha^* + \log\{(1 - p_z)/p_z\}$ and $p_z = P(Z = 1|\delta = 1)$ the probability of assignment to treatment among individuals with observed failure.

To estimate $VE(t, v)$ in model (7.1), we propose to use the Kaplan-Meier estimator for the marginal survival function in each treatment group. Model (7.1) relaxes the proportional marginal hazards assumption in exchange for flexible parametric modeling of the failure time in the density ratio model.

It is of note that model (7.1) for $VE(t, v)$ is not equivalent to model (2.5) considered in previous chapters. To illustrate, we consider a weight function satisfying (7.3) with $\widetilde{g}(t, v, \beta) = \beta_1^T v + \beta_2 t$ that yields

$$VE(t, v) = 1 - \mathrm{e}^{\alpha + \beta_1^T v + \beta_2 t} \frac{S(t|Z = 0)}{S(t|Z = 1)}. \tag{7.4}$$

For (7.4) to be equivalent to the model

$$VE(t, v) = 1 - \mathrm{e}^{\alpha + \beta^T v + \gamma}, \tag{7.5}$$

it is needed that $\beta_2 = 0$ and $S(t|Z=0)/S(t|Z=1) = \mathrm{e}^{\gamma}$. The latter condition, however, is incompatible for $\gamma \neq 0$ with the condition $S(t|Z=1) = S(t|Z=0)^{\mathrm{e}^{\gamma}}$ required by model (7.5) as a consequence of the proportional marginal hazards assumption. Further research is necessary to compare finite-sample and asymptotic properties of the two approaches to modeling the $VE(t,v)$ parameter.

7.2.2 Estimation of HIV-1 acquisition time

In RV144, time of HIV-1 acquisition was estimated as the midpoint between the last negative and the first positive result of testing. If the involved sampling times are far apart, the midpoint may be a poor estimate of the acquisition time. It is of interest to develop a statistical procedure able to more accurately estimate the acquisition time, allowing (i) to improve the identification of marks observed prior to vaccine-induced antibody generation, and (ii) to produce auxiliary data with a higher level of correlation with the marks identified in (i).

In recent years high-throughput DNA sequencing technologies (for example, 454 sequencing, Metzker (2010)) have emerged that allow to identify very rare sequence variants. Such technologies allow an improved method of estimating the time of HIV-1 acquisition by studying intra-subject sequence diversity based on a random sample of thousands of HIV-1 sequences identified using deep sequencing at the earliest post-infection time point. To use subject-specific sequence diversity to predict the acquisition time, it is necessary to characterize a region of HIV-1 that provides a predictable "clock" of HIV evolution. For instance, the 500 base long envelope segment HXB2 7125–7624 was shown to be noticeably more diverse within incident cases than the envelope segment HXB2 7625–8124 (Park et al., 2011). It is also advised to exclude T-cell epitopes from the sequenced region as HIV may escape CD8+ T-cell responses within a few weeks post transmission (Goonetilleke et al., 2009). A major limitation of the existing deep sequencing methods is that they produce relatively short DNA segments (400-600 bases long) as compared to Sanger sequencing.

Whether and how this limitation impacts the accuracy of predicting acquisition times needs further exploration.

Intra-subject sequence diversity can be quantified by specifying an aspect of the distribution of all pairwise distances of subject-specific sequences. A common measure of pairwise distance between two aligned sequences of equal length is the number of base (amino acid) mismatches, i.e., their Hamming distance (HD). Park et al. (2011) showed that the mean and variance of the HD distribution failed to distinguish early infections with multiple founder viruses from chronic infections because of a rapid increase in HD average and variance caused by multi-variant transmissions. Nevertheless, a number of very similar sequences in the lineage of each founder virus was observed in early infections with multiple variants which led to the finding that the 10% quantile of the HD distribution, Q_{10}, remains low in early multi-variant infections as opposed to higher values observed in chronic infections. Thus, we hypothesize that the Q_{10} measure may be a potentially strong predictor of the HIV acquisition time. It is also necessary to compare the prediction accuracy using Q_{10} based on the Hamming distance to that based on other pairwise distance measures.

To study the ability of intra-subject sequence diversity to reliably predict the acquisition time, we may use data from the ongoing clinical observational study (RV217) conducted to explore HIV-1 prevalence, incidence, host genetics and acute-infection viral diversity in high-risk cohorts in East Africa and Thailand. To date, over 50 subjects were identified in the acute phase of infection, and thus their acquisition times are well approximated. We may employ deep sequencing methods to assess the acute-phase sequence diversity in these individuals and, subsequently, construct and evaluate a prediction model for the acquisition time.

In this research, the key questions that need to be addressed are:

(i) What region of the HIV-1 genome appropriately captures the intra-subject viral evolution?

(ii) How can the knowledge of longitudinal changes in intra-subject sequence diver-

sity improve the viral evolution model?

(iii) What measure of the pairwise distance and what aspects of the distribution of the pairwise distances most reliably predict the acquisition time?

(iv) How does the number of sequences impact the performance of the prediction model?

(v) How can we best integrate the improved prediction of acquisition time into mark definition for sieve analysis?

7.2.3 Continuous versus discretized genetic distance

A number of trials designed to evaluate vaccine efficacy to prevent infection with various pathogens (e.g., streptococcus pneumoniae or human papilloma virus) only consider discrete mark variables representing several, or sometimes over a hundred, distinct genotypes. In such settings, future research could help clarify how inferences based on continuous-mark approaches (with various degrees of parametrization) compare to those based on the standard Prentice et al. (1978) discrete-mark approach. A parametric structure for the treatment-versus-placebo mark-specific hazard ratio leads to a reduction in the number of model parameters and may result in efficiency gain. The validity of the parametric model can be verified using goodness-of-fit tests (e.g., Qin and Zhang (1997)). Lastly, if the mark of interest is a continuous quantity, treating it as such avoids the problem of data coarsening and biological ambiguity in specifying cut points defining discrete marks.

7.3 Publication plans

We plan to publish three articles involving methodology and applications presented in this dissertation.

Article 1: Mark-specific hazard-ratio model with multivariate continuous marks: an application to vaccine efficacy

Article 1 presents estimation and hypothesis testing methods in the mark-specific hazard-ratio model (2.5) in the presence of complete mark data as described in Chapter 2 and evaluated in Chapter 3, with application to vaccine efficacy. The alternative model (7.1) relaxing the $T \perp\!\!\!\perp V|Z$ assumption is also described. The RV144 V1/V2-focused complete-mark sieve analysis of Chapter 6 is presented here. (submitted to *Biometrics*)

Article 2: Density ratio models with missing data

Article 2 presents a general framework for parameter estimation in density ratio/biased sampling models in the presence of missing data as proposed in Chapter 4 and assessed in Chapter 5. Asymptotic properties of the parameter estimators in Chapter 4 are stated in this article. (manuscript in preparation)

Article 3: Mark-specific vaccine efficacy models with multivariate missing marks

Article 3 presents estimation and hypothesis testing methods about the $VE(t, v)$ parameter using model (2.5) in the presence of missing mark data. Finite-sample performance of the missing-mark vaccine efficacy estimators assessed in Chapter 5 is presented here. In addition, the article contrasts properties of models (2.5) and (7.1) and, for the univariate missing mark setting, includes a comparison with the mark-specific proportional hazards model proposed by Sun and Gilbert (SG) (2012). The Fortran code implementing SG's estimation method is available. Finally, the RV144 V1/V2-focused sieve analysis of Chapter 6 with incomplete marks measured at the HIV diagnosis time is presented here.

BIBLIOGRAPHY

Anderson, J. (1972), "Separate sample logistic discrimination," *Biometrika*, 59, 19–35.

Breslow, N. (1974), "Covariance analysis of censored survival data," *Biometrics*, 30, 89–99.

Buchbinder, S., Mehrotra, D., Duerr, A., Fitzgerald, D., Mogg, R., Li, D., 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," *Lancet*, 372, 1881–1893.

Deeks, S., Schweighardt, B., Wrin, T., Galovich, J., Hoh, R., Sinclair, E., et al. (2006), "Neutralizing Antibody Responses against Autologous and Heterologous Viruses in Acute versus Chronic Human Immunodeficiency Virus (HIV) Infection: Evidence for a Constraint on the Ability of HIV To Completely Evade Neutralizing Antibody Responses." *Journal of Virology*, 80, 6155–6164.

Douek, D., Kwong, P., and Nabell, G. (2006), "The rational design of an AIDS vaccine," *Cell*, 124, 677–681.

Edlefsen, P., Hertz, T., Magaret, C., deCamp, A., Rolland, M., Gottardo, R., et al. (2011), "Sieve analysis of V1V2 sequences in the RV144 Thai trial." *AIDS Vaccine Conference 2011*.

Efron, B. (1977), "Efficiency of Cox's likelihood function for censored data," *Journal of the American Statistical Association*, 72, 557–565.

Flynn, M., Forthal, D., Harro, C., Judson, F., Mayer, K., Para, and the rgp120 HIV Vaccine Study Group (2005), "Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection," *Journal of Infectious Diseases*, 191, 654–665.

Gilbert, P. (2000), "Large sample theory of maximum likelihood estimates in semiparametric biased sampling models," *Annals of Statistics*, 28, 151–194.

— (2004), "Goodness-of-fit tests for semiparametric biased sampling models," *Journal of Statistical Planning and Inference*, 118, 51–81.

Gilbert, P., Grove, D., Gabriel, E., Huang, Y., Gray, G., Hammer, S., et al. (2011), "A sequential Phase 2b trial design for evaluating vaccine efficacy and immune correlates for multiple HIV vaccine regimens," *Statistical Communications in Infectious Diseases*, 3.

Gilbert, P., Lele, S., and Vardi, Y. (1999), "Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials," *Biometrika*, 86, 27–43.

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

Gilbert, P. B. (2010), "Some design issues in phase 2B vs phase 3 prevention trials for testing efficacy of products or concepts," *Statistics in Medicine*, 29, 1061–1071.

Gilbert, P. B., McKeague, I. W., and Sun, Y. (2008), "The 2-sample problem for failure rates depending on a continuous mark: an application to vaccine efficacy," *Biostatistics*, 9, pp. 263–276.

Goonetilleke, N., Liu, M., Salazar-Gonzalez, J., Ferrari, G., Giorgi, E., Ganusov, V., et al. (2009), "The first T cell response to transmitted/founder virus contributes to

the control of acute viremia in HIV-1 infection," *Journal of Experimental Medicine*, 206, 1253–1272.

Grambsch, P. and Therneau, T. (1994), "Proportional hazards tests and diagnostics based on weighted residuals," *Biometrika*, 81, 515–526.

Gray, G., Buchbinder, S., and Duerr, A. (2010), "Overview of STEP and Phambili trial results: Two phase-2b test-of-concept studies investigating the efficacy of MRK adenovirus type 5 gag/pol/nef subtype B HIV vaccine." *Current Opinion in HIV & AIDS*, 5, 357–361.

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

Haynes, B., Gilbert, P., McElrath, M., Zolla-Pazner, S., Tomaras, G., Alam, S., et al. (2012), "Immune-correlates analysis of an HIV-1 vaccine efficacy trial," *New England Journal of Medicine*, 366, 1275–1286.

Hemelaar, J., Gouws, E., Ghys, P. D., and Osmanov, S. (2006), "Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004," *AIDS*, 20, W13–W23.

Horvitz, D. and Thompson, D. (1952), "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, 47, 663–685.

Huang, Y. and Louis, T. A. (1998), "Nonparametric Estimation of the Joint Distribution of Survival Time and Mark Variables," *Biometrika*, 85, pp. 785–798.

Keele, B., Giorgi, E., Salazar-Gonzalez, J., Decker, J., Pham, K., Salazar, M., Sun, C., and Grayson, T. (2008), "Identification and characterization of transmitted

and early founder virus envelopes in primary HIV-1 infection." *Proceedings of the National Academy of Sciences*, 105, 7552–7557.

Korber, B., Gaschen, B., Yusim, K., Thakallapally, R., Kesmir, C., and Detours, V. (2001), "Evolutionary and immunological implications of contemporary HIV-1 variation," *British Medical Bulletin*, 58, 19–42.

Lu, X. and Tsiatis, A. A. (2008), "Improving the efficiency of the log-rank test using auxiliary covariates." *Biometrika*, 95, 679–694.

Lunn, M. and McNeil, D. (1995), "Applying Cox Regression to Competing Risks," *Biometrics*, 51, pp. 524–532.

Mascola, J., Mathieson, B., Zack, P., Walker, M., Halstead, S., and Burke, D. (1993), "Summary report - Workshop on the potential risks of antibody-dependent enhancement in human HIV vaccine trials," *AIDS Research and Human Retroviruses*, 9, 1175–1184.

McKinnon, L., Ball, T., Kimani, J., Wachihi, C., Matu, L., Luo, M., Embree, J., Fowke, K., and Plummer, F. (2005), "Cross-clade CD8(+) T-cell responses with a preference for the predominant circulating clade," *JAIDS-Journal of Acquired Immune Deficiency Syndromes*, 40, 245–249.

Metzker, M. L. (2010), "Sequencing technologies - the next generation," *Nature Reviews Genetics*, 11, 31–46.

Moore, P. L., Ranchobe, N., Lambson, B. E., Gray, E. S., Cave, E., Abrahams, M.-R., et al. (2009), "Limited Neutralizing Antibody Specificities Drive Neutralization Escape in Early HIV-1 Subtype C Infection," *PLoS Pathogens*, 5.

Moss, P., Rowlandjones, S., Frodsham, P., Mcadam, S., Giangrande, P., McMichael, A., et al. (1995), "Persistent High-Frequency of Human Immunodeficiency Virus-

Specific Cytotoxic T-Cells In Peripheral-Blood of Infected Donors," *Proceedings of the National Academy of Sciences of the United States of America*, 92, 5773–5777.

Newey, W. and McFadden, D. (1993), "Estimation in Large Samples," *Handbook of Econometrics*, Vol. 4.

Park, S. Y., Love, T. M. T., Nelson, J., Thurston, S. W., Perelson, A. S., and Lee, H. Y. (2011), "Designing a genome-based HIV incidence assay with high sensitivity and specificity," *AIDS*, 25, F13–F19.

Pierce, D. (1982), "The asymptotic effect of substituting estimators for parameters in certain type of statistics." *Annals of Statistics*, 10, 475–478.

Pitisuttithum, P., Gilbert, P., Gurwith, M., Heyward, W., Martin, M., van Griensven, F., et al. (2006), "Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand," *Journal of Infectious Diseases*, 194, 1661–1671.

Prentice, R. and Pyke, R. (1979), "Logistic disease incidence models and case-control studies," *Biometrika*, 66, 403–411.

Prentice, R. L., Kalbfleisch, J. D., Peterson, A. V., J., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978), "The Analysis of Failure Times in the Presence of Competing Risks," *Biometrics*, 34, pp. 541–554.

Qin, J. (1998), "Inferences for Case-Control and Semiparametric Two-Sample Density Ratio Models," *Biometrika*, 85, pp. 619–630.

Qin, J. and Zhang, B. (1997), "A Goodness-of-Fit Test for Logistic Regression Models Based on Case-Control Data," *Biometrika*, 84, pp. 609–618.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris,

R., et al. (2009), "Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand," *New England Journal of Medicine*, 361, 2209–2220.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994), "Estimation of Regression Coefficients When Some Regressors Are Not Always Observed," *Journal of the American Statistical Association*, 89, pp. 846–866.

Rolland, M., Tovanabutra, S., deCamp, A. C., Frahm, N., Gilbert, P. B., Sanders-Buell, E., et al. (2011), "Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial," *Nature Medicine*, 17, 366–U168.

Rotnitzky, A. and Robins, J. (1995), "Semiparametric regression estimation in the presence of dependent censoring," *Biometrika*, 82, 805–820.

Rubin, D. (1976), "Inference and missing data," *Biometrika*, 63, 581–590.

Simes, R. J. (1986), "An Improved Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 73, pp. 751–754.

Sun, Y. and Gilbert, P. B. (2012), "Estimation of stratified mark-specific proportional hazards models with missing marks." *Scandinavian Journal of Statistics*, 39, 34–52.

Sun, Y., Gilbert, P. B., and McKeague, I. W. (2009), "Proportional hazards models with continuous marks." *Ann. Stat.*, 37, 394–426.

Tchetgen, E. J. T. (2009), "A Simple Implementation of Doubly Robust Estimation in Logistic Regression With Covariates Missing at Random," *Epidemiology*, 20, 391–394.

Thakar, M., Bhonge, L., Lakhashe, S., Shankarkumar, U., Sane, S., Kulkarni, S., Mahajan, B., and Paranjape, R. (2005), "Cytolytic T lymphocytes (CTLs) from HIV-1 subtype C-infected Indian patients recognize CTL epitopes from a conserved

immunodominant region of HIV-1 Gag and Nef," *Journal of Infectious Diseases*, 192, 749–759.

Tomaras, G., Binley, J., Gray, E., Crooks, E., Osawa, K., Moore, P., et al. (2011), "Polyclonal B Cell Responses to Conserved Neutralization Epitopes in a Subset of HIV-1-Infected Individuals," *Journal of Virology*, 85, 11502–11519.

van der Vaart, A. (1998), *Asymptotic statistics.*, Cambridge Series in Statistical and Probabilistic Mathematics, 3. Cambridge: Cambridge Univ. Press.

van der Vaart, A. and Wellner, J. A. (1996), *Weak convergence and empirical processes. With applications to statistics.*, Springer Series in Statistics. New York, NY: Springer. xvi.

van der Vaart, A. W. and Wellner, J. A. (2007), "Empirical processes indexed by estimated functions." *Asymptotics: particles, processes and inverse problems.*, 55, 234–252.

Wang, N., Zhu, T., and Ho, D. (1995), "Sequence diversity of V1 and V2 domains of GP120 from human-immunodeficiency-virus type-1 - lack of correlation with viral phenotype," *Journal of Virology*, 69, 2708–2715.

Wei, X., Decker, J., Wang, S., Hui, H., Kappes, J., Wu, X., et al. (2003), "Antibody neutralization and escape by HIV-1," *Nature*, 422, 307–312.

Zhao, L., Lipsitz, S., and Lew, D. (1996), "Regression analysis with missing covariate data using estimating equations," *Biometrics*, 52, 1165–1182.

Appendix A

PROOF OF THEOREM 2.1

Initially, we establish uniform consistency (Lemma A.1) and weak convergence (Lemma A.2) for the process $\eta_{n,\theta}$ in (2.18) as $n \to \infty$. To this end, consider the classes of functions

$$\mathcal{F}_r = \{f_{\gamma,t,r}(x,z) = f_{1,r}(x,z)f_{2,t}(x,z)f_{3,\gamma}(x,z); |\gamma - \gamma_0| \leq \delta,\ t \in [0,\tau]\}$$

for $r = 0, 1$ and some $\delta > 0$ where

$$\begin{aligned} f_{1,r}(x,z) &= z^r \\ f_{2,t}(x,z) &= I(x \geq t) \\ f_{3,\gamma}(x,z) &= \mathrm{e}^{\gamma z}. \end{aligned}$$

Then $\xi_{n,\gamma}(t) = \mathbb{P}_n f_{\gamma,t,1}$, $\xi_{0,\gamma}(t) = P f_{\gamma,t,1} = E[ZI(X \geq t)\mathrm{e}^{\gamma Z}] \leq \mathrm{e}^{\gamma}$, $\zeta_{n,\gamma}(t) = \mathbb{P}_n f_{\gamma,t,0}$, and $\zeta_{0,\gamma}(t) = P f_{\gamma,t,0} = E[I(X \geq t)\mathrm{e}^{\gamma Z}] \leq \mathrm{e}^{\gamma}$ where

$$\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta^{Kron}_{(X_i,\delta_i,\delta_i V_i,Z_i)}$$

is the empirical measure and δ^{Kron} denotes Kronecker's delta.

Lemma A.1. $\sup_{\gamma,t} |\eta_{n,\gamma}(t) - \eta_{0,\gamma}(t)| \xrightarrow[n\to\infty]{P} 0$ *where the supremum is taken over all values of* γ *and* t *such that* $|\gamma - \gamma_0| \leq \delta$ *for some* $\delta > 0$ *and* $t \in [0,\tau]$.

Proof. Consider a fixed $r \in \{0,1\}$ and the normed space of functions $f : \{0,1\} \times [0,\infty) \mapsto \mathbb{R}_0^+$ with the $L_1(P)$-norm. Trivially, the class of functions described by $f_{1,r}$ has a finite bracketing number as it consists of one bounded function only with a finite norm. The class of monotone functions described by $f_{2,t}$ mapping into $[0,1]$

has a finite bracketing number for every $\varepsilon > 0$ by van der Vaart and Wellner (1996) (henceforth vdV&W), Theorem 2.7.5, page 159. The functions $f_{3,\gamma}$ indexed by $\gamma \in T = [\gamma_0 - \delta, \gamma_0 + \delta]$ are differentiable in γ, and thus by the Mean Value Theorem, for any $\gamma_1, \gamma_2 \in T$, $\gamma_1 \leq \gamma_2$, and some $\gamma \in (\gamma_1, \gamma_2)$,

$$\frac{|f_{3,\gamma_1}(z,x) - f_{3,\gamma_2}(z,x)|}{|\gamma_1 - \gamma_2|} = |z\mathrm{e}^{\gamma z}| \leq z\mathrm{e}^{(\gamma_0+\delta)z} =: F(z).$$

It implies that the functions $f_{3,\gamma}$ are Lipschitz in the index parameter γ, and therefore, by vdV&W Theorem 2.7.11, page 164, for any $\varepsilon > 0$, the upper bound of the bracketing number $N_{[]}(2\varepsilon\|F\|_{P,1}, \widetilde{\mathcal{F}}_3, L_1(P))$ for the class $\widetilde{\mathcal{F}}_3$ of functions $f_{3,\gamma}$ is given by the covering number $N(\varepsilon, T, |\cdot|)$ which is finite since the index set T is compact in the metric space $(\mathbb{R}, |\cdot|)$. Now consider brackets of the form $[l_1 l_2 l_3, u_1 u_2 u_3]$ covering $\mathcal{F}_r$ where the functions l_i, u_i, $i = 1, 2, 3$, are elements of classes of functions described by one of $f_{1,r}$, $f_{2,t}$, and $f_{3,\gamma}$ that define finitely many ε-brackets covering the respective classes. Then, setting $l_1(z,x) = u_1(z,x) = I_{[z^r=1]}$ and using the triangle inequality, we obtain

$$\begin{aligned}
\|u_1u_2u_3 - l_1l_2l_3\|_{P,1} &= \|u_1(u_2u_3 - l_2l_3)\|_{P,1} \\
&\leq \|u_1((u_2 - l_2)u_3 + (u_3 - l_3)l_2)\|_{P,1} \\
&\leq \|(u_2 - l_2)u_3\|_{P,1} + \|(u_3 - l_3)l_2\|_{P,1} \\
&\leq \varepsilon(\|u_3\|_{P,1} + \|l_2\|_{P,1}) \\
&\leq \varepsilon(\mathrm{e}^{\gamma_0+\delta} + 1).
\end{aligned}$$

Thus, for every $\varepsilon > 0$, there exist finitely many $\varepsilon(\mathrm{e}^{\gamma_0+\delta} + 1)$-brackets covering $\mathcal{F}_r$, i.e.,

$$N_{[]}(\varepsilon, \mathcal{F}_r, L_1(P)) < \infty.$$

By vdV&W Theorem 2.4.1, page 122, the class $\mathcal{F}_r$ is P–Glivenko-Cantelli for $r = 0, 1$,

and therefore

$$\begin{aligned}
\sup_{\gamma,t}|\eta_{n,\gamma}(t)-\eta_{0,\gamma}(t)| &= \sup_{\gamma,t}\left|\frac{\xi_{n,\gamma}(t)}{\zeta_{n,\gamma}(t)}-\frac{\xi_{0,\gamma}(t)}{\zeta_{0,\gamma}(t)}\right| \\
&= \sup_{\gamma,t}\frac{|\xi_{n,\gamma}(t)\zeta_{0,\gamma}(t)-\xi_{0,\gamma}(t)\zeta_{0,\gamma}(t)+\xi_{0,\gamma}(t)\zeta_{0,\gamma}(t)-\xi_{0,\gamma}(t)\zeta_{n,\gamma}(t)|}{|\zeta_{n,\gamma}(t)|\,|\zeta_{0,\gamma}(t)|} \\
&\leq \sup_{\gamma,t}\frac{|\zeta_{0,\gamma}(t)|\,|\xi_{n,\gamma}(t)-\xi_{0,\gamma}(t)|+|\xi_{0,\gamma}(t)|\,|\zeta_{n,\gamma}(t)-\zeta_{0,\gamma}(t)|}{|\zeta_{n,\gamma}(t)|\,|\zeta_{0,\gamma}(t)|} \\
&\leq \sup_{\gamma,t}\frac{|\xi_{n,\gamma}(t)-\xi_{0,\gamma}(t)|}{|\zeta_{n,\gamma}(t)|}+\sup_{\gamma,t}\frac{|\zeta_{n,\gamma}(t)-\zeta_{0,\gamma}(t)|}{|\zeta_{n,\gamma}(t)|} \\
&\xrightarrow[n\to\infty]{P} 0
\end{aligned}$$

where, in the last inequality, we used the fact that $|\xi_{0,\gamma}(t)/\zeta_{0,\gamma}(t)| \leq 1$. Note that $\zeta_{n,\gamma}(t) > 0$ and $\zeta_{0,\gamma}(t) > 0$ for all $t \in [0,\tau]$, and thus all considered fractions are well defined. □

Let $\mathbb{G}_n$ be the empirical process defined as $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$.

Lemma A.2. *For $\gamma \in [\gamma_0-\delta, \gamma_0+\delta]$ and $t \in [0,\tau]$,*

$$\sqrt{n}(\eta_{n,\gamma}(t)-\eta_{0,\gamma}(t)) = \frac{1}{\zeta_{0,\gamma}(t)}\,\mathbb{G}_n\left(f_{\gamma,t,1}-\eta_{0,\gamma}(t)f_{\gamma,t,0}\right)+o_p(1)$$

and $\mathbb{G}_n \Rightarrow \mathbb{G}_P$ in $l^\infty(\mathcal{F})$ as $n \to \infty$ where $\mathbb{G}_P$ is the P–Brownian bridge process indexed by the class of functions

$$\mathcal{F} = \{f_{\gamma,t,1}-\eta_{0,\gamma}(t)f_{\gamma,t,0}; |\gamma-\gamma_0| \leq \delta,\ t \in [0,\tau]\}.$$

Proof. We have

$$\begin{aligned}
\sqrt{n}\left(\eta_{n,\gamma}(t)-\eta_{0,\gamma}(t)\right) &= \sqrt{n}\left\{\frac{\xi_{n,\gamma}(t)}{\zeta_{n,\gamma}(t)}-\frac{\xi_{0,\gamma}(t)}{\zeta_{0,\gamma(t)}}\right\}\\
&= \frac{1}{\zeta_{n,\gamma}(t)}\sqrt{n}\left(\xi_{n,\gamma}(t)-\xi_{0,\gamma}(t)\right)+\sqrt{n}\left\{\frac{\xi_{0,\gamma}(t)}{\zeta_{n,\gamma}(t)}-\frac{\xi_{0,\gamma}(t)}{\zeta_{0,\gamma}(t)}\right\}\\
&= \frac{1}{\zeta_{n,\gamma}(t)}\sqrt{n}\left(\xi_{n,\gamma}(t)-\xi_{0,\gamma}(t)\right)\\
&\quad - \frac{\xi_{0,\gamma}(t)}{\zeta_{n,\gamma}(t)\zeta_{0,\gamma}(t)}\sqrt{n}\left(\zeta_{n,\gamma}(t)-\zeta_{0,\gamma}(t)\right)\\
&= \frac{1}{\zeta_{0,\gamma}(t)}\sqrt{n}\left(\xi_{n,\gamma}(t)-\xi_{0,\gamma}(t)\right)\\
&\quad - \frac{\xi_{0,\gamma}(t)}{\zeta_{0,\gamma}(t)^2}\sqrt{n}\left(\zeta_{n,\gamma}(t)-\zeta_{0,\gamma}(t)\right)+o_p(1)\\
&= \frac{1}{\zeta_{0,\gamma}(t)}\mathbb{G}_n\left(f_{\gamma,t,1}-\eta_{0,\gamma}(t)f_{\gamma,t,0}\right)+o_p(1).
\end{aligned}$$

In the second part, we show that the uniform entropy condition holds for the class $\mathcal{F}$ and prove the weak convergence result. Consider a fixed $r\in\{0,1\}$ and the normed space of functions $f:\{0,1\}\times[0,\infty)\mapsto\mathbb{R}_0^+$ with the $L_2(P)$-norm. Trivially, the covering number for the class of functions described by $f_{1,r}$ is 1 for every $\varepsilon>0$ because the class is a one-point set whose element is a square integrable function. Now consider the class $\widetilde{\mathcal{F}}_2$ of functions described by $f_{2,t}$. The envelope function for $\widetilde{\mathcal{F}}_2$ is $F_2(x)\equiv 1$ which is square integrable since $QF_2^2=1$. Further, for fixed $t_1,t_2\in\mathbb{R}$, and a probability measure Q on $(\mathbb{R},\mathcal{B})$,

$$\begin{aligned}
\|f_{2,t_1}-f_{2,t_2}\|_{Q,2}^2 &= \int (f_{2,t_1}-f_{2,t_2})^2 \mathrm{d}Q\\
&= Q(W\geq t_1)-2\,Q(W\geq t_1\vee t_2)+Q(W\geq t_2)\\
&= Q(t_1\wedge t_2\leq W<t_1\vee t_2).
\end{aligned}$$

Hence, if $\|f_{2,t_1}-f_{2,t_2}\|_{Q,2}<\varepsilon$, then there exist $\lceil\frac{1}{\varepsilon^2}\rceil$ balls of the radius ε (with respect to the $L_2(Q)$-norm) that cover $\widetilde{\mathcal{F}}_2$. Thus,

$$N(\varepsilon\|F_2\|_{Q,2},\widetilde{\mathcal{F}}_2,L_2(Q))=N(\varepsilon,\widetilde{\mathcal{F}}_2,L_2(Q))\leq 2\left(\frac{1}{\varepsilon}\right)^2.$$

By vdV&W Theorem 2.6.9, page 142, there exists a constant K_2 such that

$$\log N(\varepsilon, \overline{\text{conv}}\, \widetilde{\mathcal{F}}_2, L_2(Q)) \leq \frac{K_2}{\varepsilon}.$$

Lastly, consider the class $\widetilde{\mathcal{F}}_3$ of functions described by $f_{3,\gamma}$. The collection of all subgraphs (in $\{0,1\}\times\mathbb{R}^+$) of the functions $f_{3,\gamma}$ shatters no two-point set $\{(x_1,y_1),(x_2,y_2)\} \subset \{0,1\}\times\mathbb{R}^+$ because it fails to pick out either the point whose second coordinate equals $\max(y_1,y_2)$ if $x_1 = x_2$ or the point (x_1,y_1) if $x_1 < x_2$ & $y_1 \geq 1$ or the point (x_2,y_2) if $x_1 < x_2$ & $y_1 < 1$. Thus, $\widetilde{\mathcal{F}}_3$ is a VC-class and its VC-index equals 2. By vdV&W Corollary 2.6.12, page 145, for a probability measure Q,

$$\log N(\varepsilon \|F_3\|_{Q,2}, \widetilde{\mathcal{F}}_3, L_2(Q)) \leq \frac{K_3}{\varepsilon}$$

for a constant K_3 and the square integrable envelope function $F_3(z) = \mathrm{e}^{(\gamma_0+\delta)z}$ for $\widetilde{\mathcal{F}}_3$. Now consider a set of ε-balls covering $\mathcal{F}_r$ with centers formed as products of centers of finitely many ε-balls covering $\widetilde{\mathcal{F}}_1$, $\widetilde{\mathcal{F}}_2$, and $\widetilde{\mathcal{F}}_3$. Based on the upper bounds for entropy of the component classes, for every $\varepsilon > 0$, there exists a constant K such that

$$\log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}_r, L_2(Q)) \leq \frac{K}{\varepsilon}$$

where $F(z) = zF_3(z)$ is a square integrable envelope function for $\mathcal{F}_r$. It follows that the class $\mathcal{F}_r$ satisfies the uniform entropy condition

$$\int_0^\infty \sup_Q \sqrt{\log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}_r, L_2(Q))}\mathrm{d}\varepsilon < \infty,$$

and thus, by vdV&W Theorem 2.5.2, page 127, the class $\mathcal{F}_r$ is P–Donsker for $r = 0, 1$. For any $\gamma \in [\gamma_0-\delta, \gamma_0+\delta]$ and $t \in [0,\tau]$, $0 \leq \eta_{0,\gamma}(t) \leq 1$, and therefore, for a probability measure Q, the covering number $N(\varepsilon, \mathcal{F}_{\eta,0}, L_2(Q))$ for the class of functions

$$\mathcal{F}_{\eta,0} = \{\eta_{0,\gamma}(t) f_{0,\gamma,t}; |\gamma - \gamma_0| \leq \delta, t \in [0,\tau]\}$$

is bounded above by $N(\varepsilon, \mathcal{F}_0, L_2(Q))$ for every $\varepsilon > 0$. It follows that the class $\mathcal{F}_{\eta,0}$ is P–Donsker. The classes $\mathcal{F}_1$ and $\mathcal{F}_{\eta,0}$ are uniformly bounded which implies that $\|P\|_{\mathcal{F}_1 \cup \mathcal{F}_{\eta,0}} < \infty$, and thus, by vdV&W Example 2.10.7, page 192, the pairwise sums $\mathcal{F}_1 + \mathcal{F}_{\eta,0}$ also form a P–Donsker class which was to be demonstrated. □

Proof of Theorem 2.1

Consider the following linearization of the estimating functions:

$$\begin{aligned}\sqrt{n}(\Psi_n(\theta_0) - \Psi(\theta_0)) &= \sqrt{n}(\mathbb{P}_n\varphi(\theta_0, \eta_{n,\theta_0}) - P\varphi(\theta_0, \eta_{0,\theta_0})) \\ &= \mathbb{G}_n(\varphi(\theta_0, \eta_{n,\theta_0}) - \varphi(\theta_0, \eta_{0,\theta_0})) + \mathbb{G}_n\varphi(\theta_0, \eta_{0,\theta_0}) \qquad \text{(A.1)} \\ &\quad + \sqrt{n}(P\varphi(\theta_0, \eta_{n,\theta_0}) - P\varphi(\theta_0, \eta_{0,\theta_0}))\end{aligned}$$

where $\mathbb{G}_n$ is the empirical process defined by $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$. The decomposition splits the derivation of the convergence in distribution of (A.1) into three parts. Huber's Z-theorem is employed in the end.

Consider the class of weight functions

$$\mathcal{F}_g = \{\mathrm{e}^{\alpha+\beta' v+v' Bv}; |\alpha - \alpha_0| \le \delta_\alpha, |\beta - \beta_0| \le \delta_\beta, |B - B_0| \le \delta_B\}$$

for $v \in [0,1]^s$, $B = (b_{ij})_{i,j=1}^s$ and $B_0 = (b_{0,ij})_{i,j=1}^s$ upper triangular matrices, and some $\delta_\alpha > 0$, $\delta_\beta > 0$, and $\delta_B > 0$. All inequalities above hold componentwise. Let H be the space of left-continuous, piecewise constant, monotone functions $f : \mathbb{R}_0^+ \mapsto [0, \mathrm{e}^{\gamma_0+\delta_\gamma}]$ for some $\delta_\gamma > 0$. Consider the norm $\|f\|_0^\tau = \sup_{x\in[0,\tau]} |f(x)|$ for $f = f_1 - f_2$, $f_1, f_2 \in H$. Initially, we show that the class of functions

$$\begin{aligned}\mathcal{F}_\varphi = \{\varphi(\theta,\eta);\, \theta = (\alpha,\beta,B,\gamma), |\alpha-\alpha_0| \le \delta_\alpha, |\beta-\beta_0| \le \delta_\beta, |B-B_0| \le \delta_B, |\gamma-\gamma_0| \le \delta_\gamma, \\ \eta = \xi/\zeta,\ \xi,\zeta \in H, \|\xi-\xi_0\|_0^\tau \le \delta, \|\zeta-\zeta_0\|_0^\tau \le \delta,\ \zeta_0(\tau) > \varepsilon > \delta > 0\} \quad \text{(A.2)}\end{aligned}$$

is P–Donsker for weight functions $g(v,\phi)$, $\phi = (\alpha,\beta,B)$, belonging to $\mathcal{F}_g$.

The class $\mathcal{F}_\alpha = \{\mathrm{e}^\alpha; |\alpha - \alpha_0| \le \delta_\alpha\}$ is a VC-class of constant functions and its VC-index equals 2. Thus, the P–Donsker property of $\mathcal{F}_\alpha$ follows by vdV&W Corollary 2.6.12, page 145. For each $i = 1, \ldots, s$, the class $\mathcal{F}_{\beta,i} = \{\mathrm{e}^{\beta_i v_i}; |\beta_i - \beta_{0,i}| \le \delta_{\beta,i}\}$ consists of uniformly bounded monotone functions, and therefore, applying vdV&W Theorem 2.7.5, page 159, is P–Donsker. Since the classes $\mathcal{F}_{\beta,i}$ are uniformly bounded, the class of componentwise products $\mathcal{F}_\beta = \mathcal{F}_{\beta,1} \cdot \ldots \cdot \mathcal{F}_{\beta,s}$ is P–Donsker. The class of quadratic

forms $\widetilde{\mathcal{F}}_B = \{v'Bv; |B - B_0| \leq \delta_B\}$ is a finite-dimensional vector space of measurable functions $f : [0,1]^s \mapsto \mathbb{R}$ with $\dim(\widetilde{\mathcal{F}}_B) = (1+s)s/2$. By vdV&W Lemma 2.6.15, page 146, $\widetilde{\mathcal{F}}_B$ is a VC-class with the VC-index $V(\widetilde{\mathcal{F}}_B) \leq \dim(\widetilde{\mathcal{F}}_B) + 2$. Consider the class $\mathcal{F}_B = \vartheta \circ \widetilde{\mathcal{F}}_B$ where $\vartheta(x) = \mathrm{e}^x$. The map ϑ is monotone, and therefore, applying vdV&W Lemma 2.6.18 (viii), $\mathcal{F}_B$ is P–Donsker. Uniform boundedness of $\mathcal{F}_\alpha$, $\mathcal{F}_\beta$, and $\mathcal{F}_B$ implies that the class of weight functions $\mathcal{F}_g$, formed as componentwise products $\mathcal{F}_\alpha \cdot \mathcal{F}_\beta \cdot \mathcal{F}_B$, is P–Donsker. Further, the classes $\mathcal{F}_\xi = \{\xi \in H; \|\xi - \xi_0\|_0^\tau \leq \delta\}$ and $\mathcal{F}_\zeta = \{\zeta \in H; \|\zeta - \zeta_0\|_0^\tau \leq \delta, \zeta_0(\tau) > \varepsilon > \delta > 0\}$ consist of uniformly bounded monotone functions, and therefore, applying vdV&W Theorem 2.7.5, page 159, are P–Donsker. Consider the map $q(x,y) = x/y$ for $(x,y) \in D = [0, \mathrm{e}^{\gamma_0+\delta_\gamma}] \times (\varepsilon - \delta, \mathrm{e}^{\gamma_0+\delta_\gamma}]$. Since q has a bounded continuous gradient on D, it is Lipschitz. Subsequently, the class $q \circ (\mathcal{F}_\xi, \mathcal{F}_\zeta)$ is P–Donsker by vdV&W Theorem 2.10.6, page 192. The results above imply that the class $\mathcal{F}_\varphi$ is P–Donsker, that is, $\mathbb{G}_n \Rightarrow \mathbb{G}_P$ in $l^\infty(\mathcal{F}_\varphi)$ as $n \to \infty$ where $\mathbb{G}_P$ is the P–Brownian bridge process on $\mathcal{F}_\varphi$.

Further, we have the "consistency" condition

$$\begin{aligned}
\sup_\theta P(\varphi_3(\theta, \eta_{n,\theta}) - \varphi_3(\theta, \eta_{0,\theta}))^2 &= \sup_\theta E\left[\delta(Z - \eta_{n,\theta}(X)) - \delta(Z - \eta_{0,\theta}(X))\right]^2 \\
&= \sup_\theta E\left[\delta\left(\eta_{n,\theta}(X) - \eta_{0,\theta}(X)\right)^2\right] \\
&\leq \sup_{\theta,t} (\eta_{n,\theta}(t) - \eta_{0,\theta}(t))^2 \xrightarrow[n\to\infty]{P} 0
\end{aligned}$$

where the convergence in probability is implied by Lemma A.1. The same convergence holds trivially for the components φ_1 and φ_2 as they do not involve the estimator $\eta_{n,\theta}$. Consequently, by Theorem 2.1 of van der Vaart and Wellner (2007),

$$\sup_\theta |\, \mathbb{G}_n(\varphi(\theta, \eta_{n,\theta}) - \varphi(\theta, \eta_{0,\theta}))| \xrightarrow[n\to\infty]{P} 0$$

where the supremum is taken over all values of θ specified in (A.2). Thus, the probability limit of the first summand in (A.1) is 0.

As a consequence of the P–Donsker property of the class $\mathcal{F}_\varphi$, for the second

summand in (A.1) we obtain

$$\mathbb{G}_n\varphi(\theta_0,\eta_{0,\theta_0}) \xrightarrow[n\to\infty]{\mathcal{D}} \mathbb{G}_P\varphi(\theta_0,\eta_{0,\theta_0}) \sim \mathsf{N}_{d+2}(0,\Sigma)$$

where $d = 1+s+\frac{(s+1)s}{2}$ and $\Sigma = P\varphi(\theta_0,\eta_{0,\theta_0})\varphi^T(\theta_0,\eta_{0,\theta_0}) - P\varphi(\theta_0,\eta_{0,\theta_0})P\varphi^T(\theta_0,\eta_{0,\theta_0})$.

Regarding the last summand in (A.1), consider the transformation $\nu : \eta \mapsto P\varphi(\theta_0,\eta)$ for $\eta \in H$. Because the functions φ_1 and φ_2 do not involve the process η, we focus our attention on merely the third component of the transformation $\nu_3 : \eta \mapsto P\varphi_3(\theta_0,\eta)$. For the difference quotient, we obtain

$$\begin{aligned}\frac{\nu_3(\eta+th_t)-\nu_3(\eta)}{t} &= \frac{E\left\{\delta(Z-\eta(X)-th_t(X))\right\} - E\left\{\delta(Z-\eta(X))\right\}}{t}\\ &= -E\delta h_t(X) \longrightarrow -E\delta h(X) \quad \text{as } t \searrow 0\end{aligned}$$

and for every sequence of functions $h_t \in H$ such that $h_t \to h$ which implies that the map ν_3 is Hadamard-differentiable with the derivative $\nu'_{3,\eta}(h) = -E\delta h(X)$ ($\nu'_{3,\eta}$ is a continuous and linear map between H and $\mathbb{R}$). Using Lemma A.2 and the functional delta method (see, for example, van der Vaart (1998)), we obtain

$$\begin{aligned}&\sqrt{n}\left(\nu_3(\eta_{n,\theta_0}) - \nu_3(\eta_{0,\theta_0})\right)\\ &\quad= \nu'_{3,\eta}\left(\sqrt{n}(\eta_{n,\theta_0}-\eta_{0,\theta_0})\right) + o_p(1)\\ &\quad= -\sqrt{n}\int (\eta_{n,\theta_0}(t)-\eta_{0,\theta_0}(t))\,\mathrm{d}F(t|\delta=1)\times P(\delta=1) + o_p(1)\\ &\quad\xrightarrow[n\to\infty]{\mathcal{D}} \nu'_{3,\eta}\left(\zeta_{0,\theta_0}^{-1}(\cdot)\,\mathbb{G}_P\left(f_{\theta_0,\cdot,1}(x,z) - \eta_{0,\theta_0}(\cdot)f_{\theta_0,\cdot,0}(x,z)\right)\right)\\ &\quad= -\iint \zeta_{0,\theta_0}^{-1}(t)\left(f_{\theta_0,t,1}(x,z) - \eta_{0,\theta_0}(t)f_{\theta_0,t,0}(x,z)\right)\mathrm{d}\mathbb{G}_P\,\mathrm{d}F(t|\delta=1)\times P(\delta=1)\\ &\quad= \mathbb{G}_P\Bigg(\underbrace{\int \zeta_{0,\theta_0}^{-1}(t)\left(f_{\theta_0,t,1}(x,z) - \eta_{0,\theta_0}(t)f_{\theta_0,t,0}(x,z)\right)\mathrm{d}F(t|\delta=1)}_{=:l_{\theta_0}(x,z)}\Bigg)\times P(\delta=1)\end{aligned}$$

where the last two equalities follow by the linearity of the map $f \mapsto \mathbb{G}_P(f)$ and Fubini's Theorem, respectively. Overall we attain

$$\sqrt{n}(\Psi_n(\theta_0)-\Psi(\theta_0)) \xrightarrow[n\to\infty]{\mathcal{D}} Z = \mathbb{G}_P\begin{pmatrix}\varphi_1(\theta_0,\eta_{0,\theta_0})\\ \varphi_2(\theta_0,\eta_{0,\theta_0})\\ \varphi_3(\theta_0,\eta_{0,\theta_0}) + p_\delta l_{\theta_0}\end{pmatrix} \tag{A.3}$$

where $p_\delta = P(\delta = 1)$.

To apply Huber's Z-theorem, it remains to show the condition of asymptotic equicontinuity:

$$\sup_{\theta:\|\theta-\theta_0\|\le\delta_n} \frac{\|\sqrt{n}(\Psi_n - \Psi)(\theta) - \sqrt{n}(\Psi_n - \Psi)(\theta_0)\|}{1+\sqrt{n}\|\theta-\theta_0\|} \xrightarrow{P} 0 \tag{A.4}$$

for every sequence $\delta_n \searrow 0$. The class $\mathcal{F}_\varphi - \varphi(\theta_0, \eta_{0,\theta_0})$ is P–Donsker, and therefore, by vdV&W Lemma 3.3.5, page 311,

$$\sup_{\theta:\|\theta-\theta_0\|\le\delta_n} \frac{\|\mathbb{G}_n\left(\varphi(\theta,\eta_{0,\theta}) - \varphi(\theta_0,\eta_{0,\theta_0})\right)\|}{1+\sqrt{n}\|\theta-\theta_0\|} \xrightarrow{P} 0$$

as $\delta_n \searrow 0$. The weak convergence result for the sequence $\sqrt{n}(\eta_{n,\theta}(t) - \eta_{0,\theta}(t))$ in Lemma A.2, uniform in $t \in [0,\tau]$, and the smoothness of the transformation ν_3 of η (which is based on averaging of the values of η with weights determined by the distribution of X) ensure that

$$\sup_{\theta:\|\theta-\theta_0\|\le\delta_n} \frac{\|\sqrt{n}(P\varphi(\theta,\eta_{n,\theta}) - P\varphi(\theta,\eta_{0,\theta})) - \sqrt{n}(P\varphi(\theta_0,\eta_{n,\theta_0}) - P\varphi(\theta_0,\eta_{0,\theta_0}))\|}{1+\sqrt{n}\|\theta-\theta_0\|} \xrightarrow{P} 0$$

as $\delta_n \searrow 0$. The triangle inequality and the linearity of the map $f \mapsto \mathbb{G}_n f$ imply that, for a sufficiently large n and θ such that $\|\theta - \theta_0\| \le \delta_n$,

$$\begin{aligned}
&\|\sqrt{n}(\Psi_n - \Psi)(\theta) - \sqrt{n}(\Psi_n - \Psi)(\theta_0)\| \\
&\qquad \le \|\mathbb{G}_n\left(\varphi(\theta,\eta_{0,\theta}) - \varphi(\theta_0,\eta_{0,\theta_0})\right)\| \\
&\qquad\quad + \|\sqrt{n}(P\varphi(\theta,\eta_{n,\theta}) - P\varphi(\theta,\eta_{0,\theta})) - \sqrt{n}(P\varphi(\theta_0,\eta_{n,\theta_0}) - P\varphi(\theta_0,\eta_{0,\theta_0}))\|,
\end{aligned}$$

and therefore, the condition (A.4) holds. Finally, using (A.3) and vdV&W Theorem 3.3.1, page 310, we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} -\dot{\Psi}_{\theta_0}^{-1} Z$$

where $\dot{\Psi}_{\theta_0}$ is the continuously invertible derivative of the map $\theta \mapsto \Psi(\theta)$ at θ_0 and its explicit form is specified in the statement of the theorem. □

Appendix B

RV144 SIEVE ANALYSIS USING OTHER V1/V2 AND GP120 DISTANCES

B.1 Distance distribution

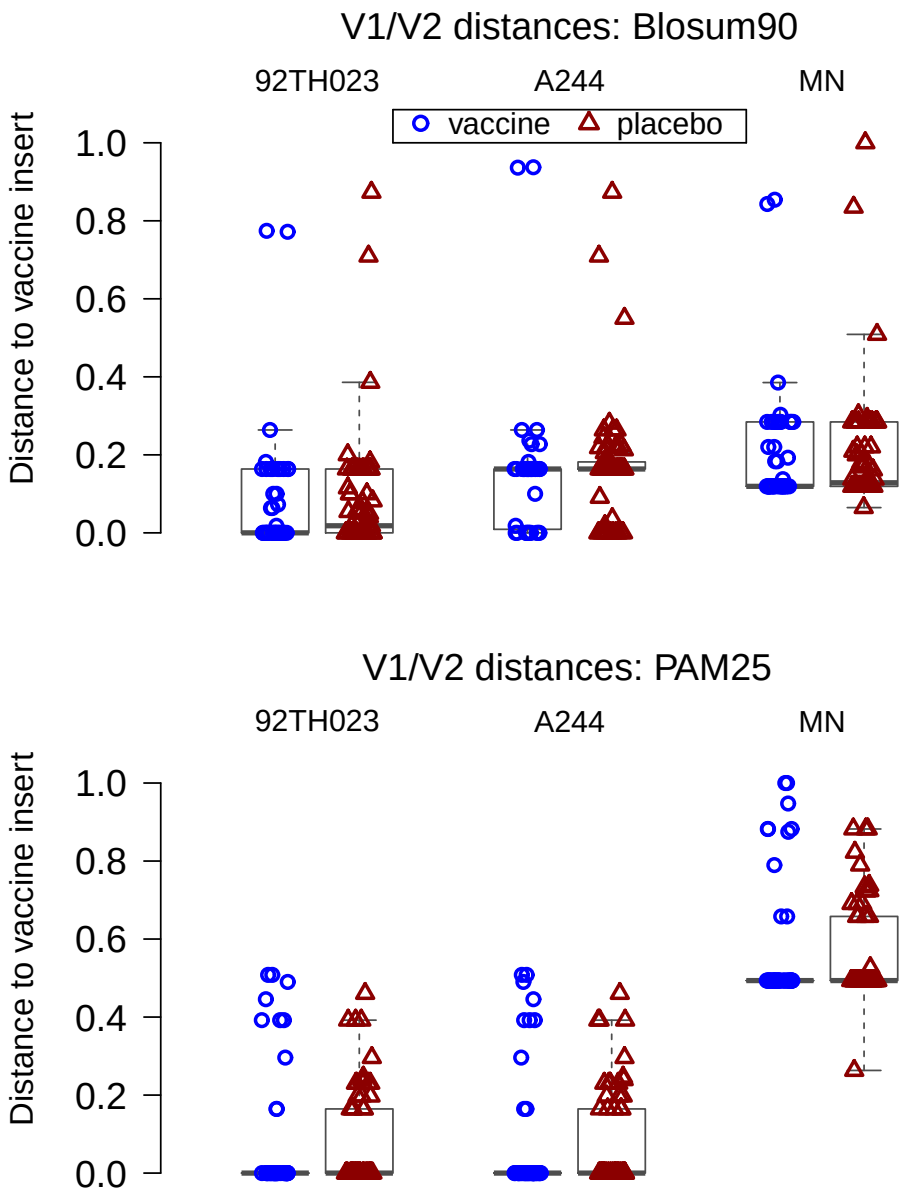


Figure B.1: Distribution of RV144 V1/V2-focused distances using the published set of monoclonal antibody contact sites in the V1/V2 domain.

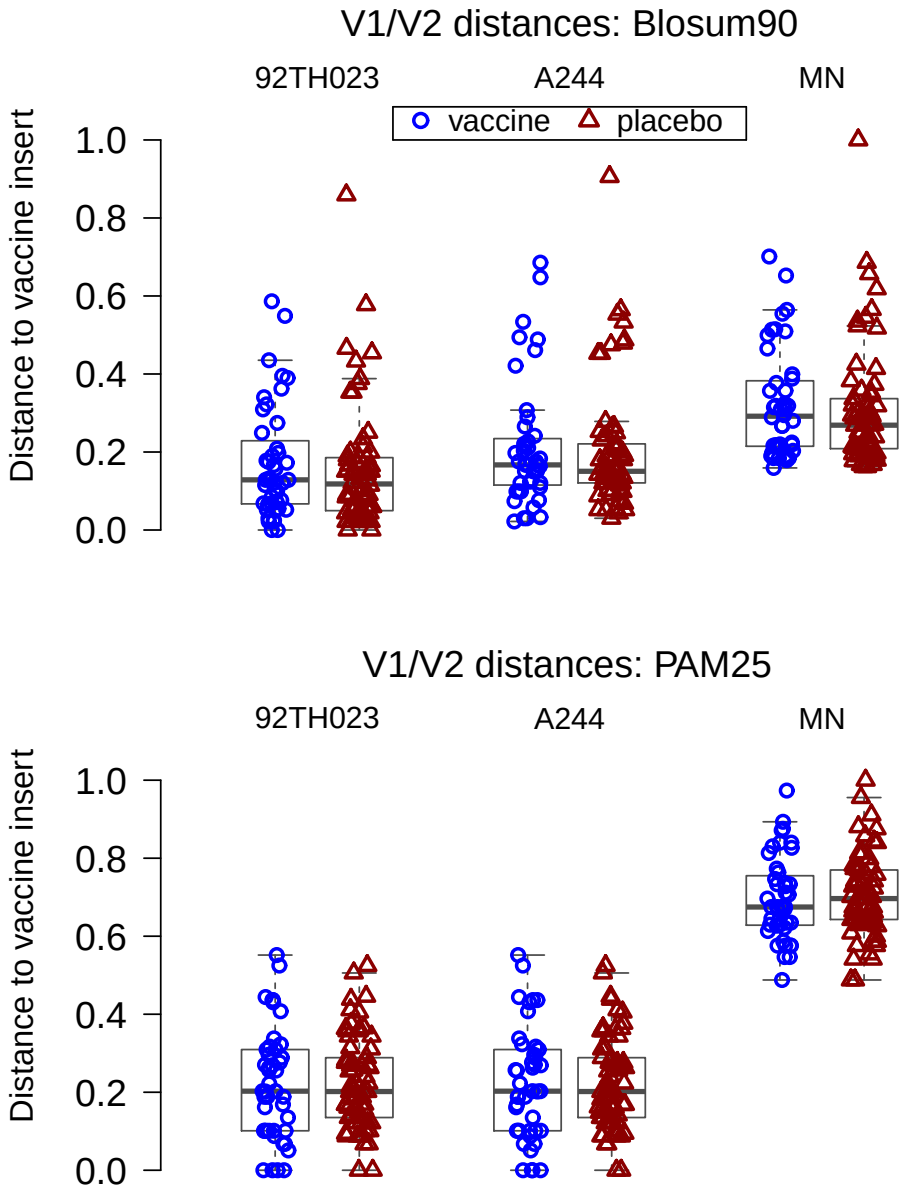


Figure B.2: Distribution of RV144 V1/V2-focused distances using the published set of monoclonal antibody contact sites and other published neutralization relevant contact sites in the V1/V2 domain.

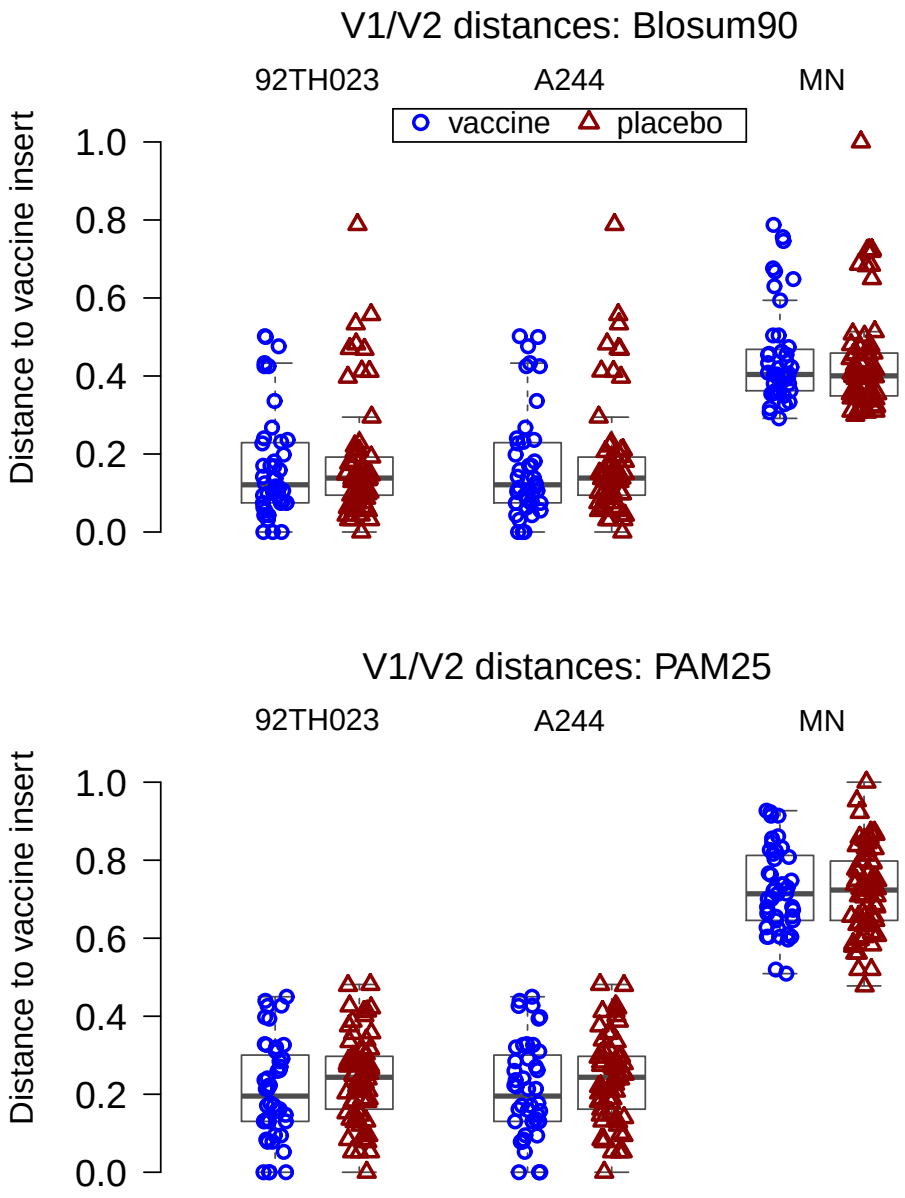


Figure B.3: Distribution of RV144 V1/V2-focused distances using 22 V1/V2 sites with the highest frequency of occurrence in predicted antibody epitopes based on structural biology.

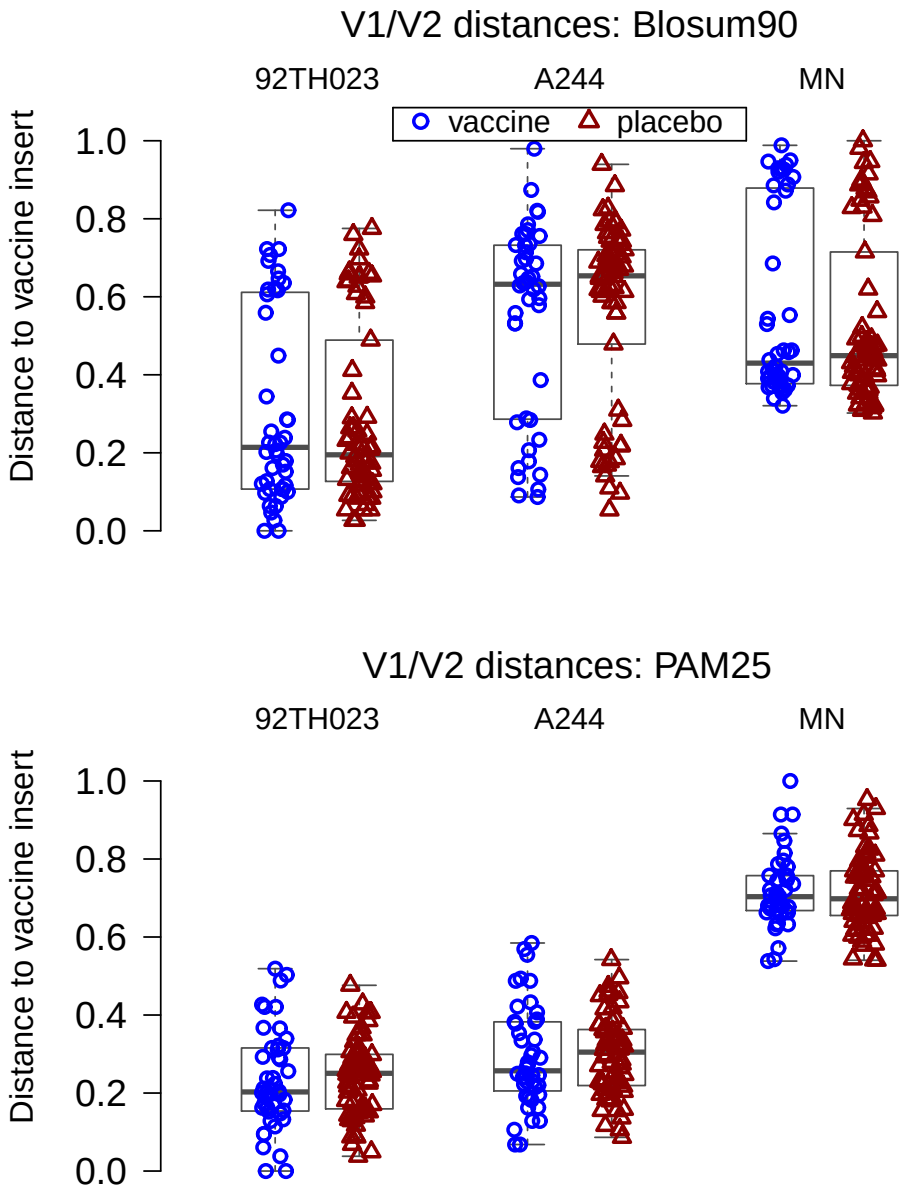


Figure B.4: Distribution of RV144 V1/V2-focused distances using hotspots (Haynes et al., 2012) in the linear peptide binding microarray analysis.

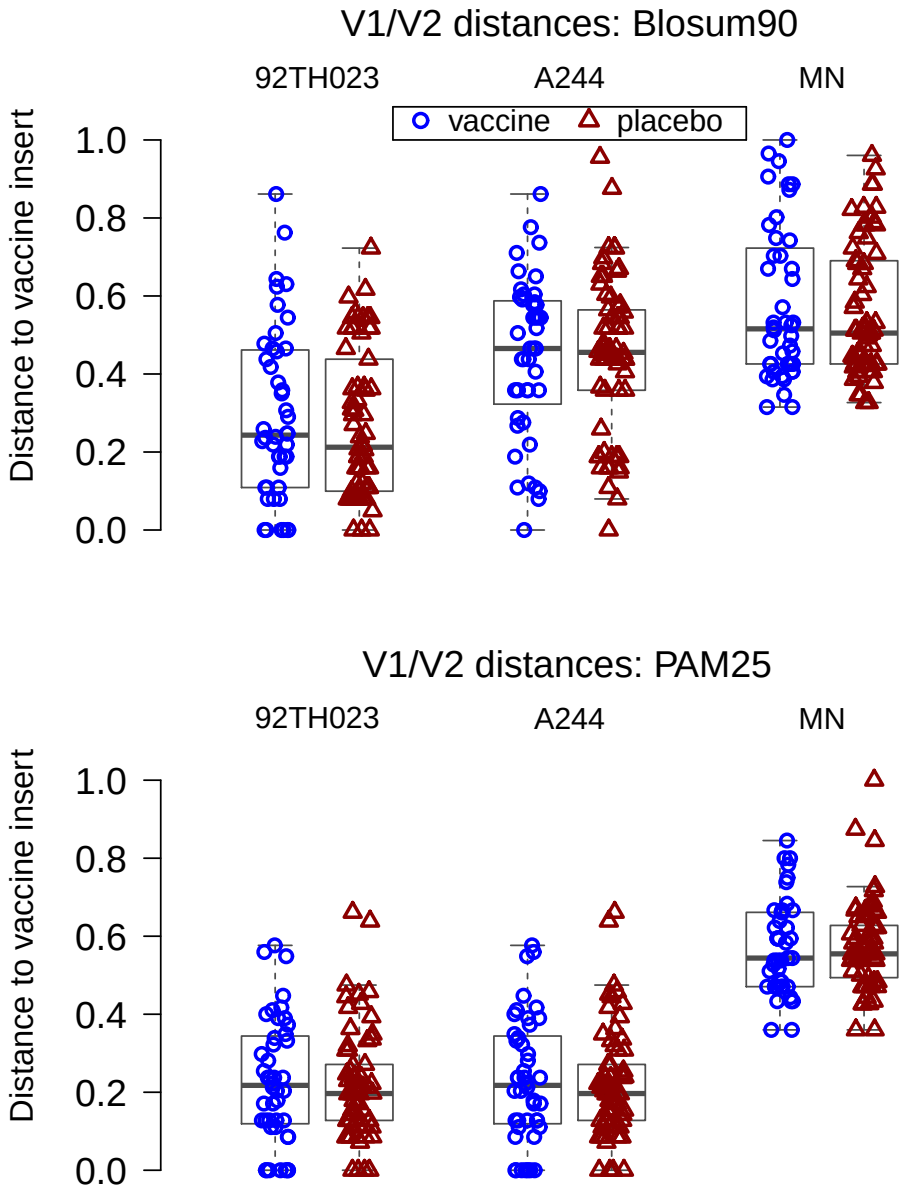


Figure B.5: Distribution of RV144 V1/V2-focused distances using the intersection of the published set of monoclonal antibody and other neutralization relevant V1/V2 contact sites with linear peptide microarray hotspots.

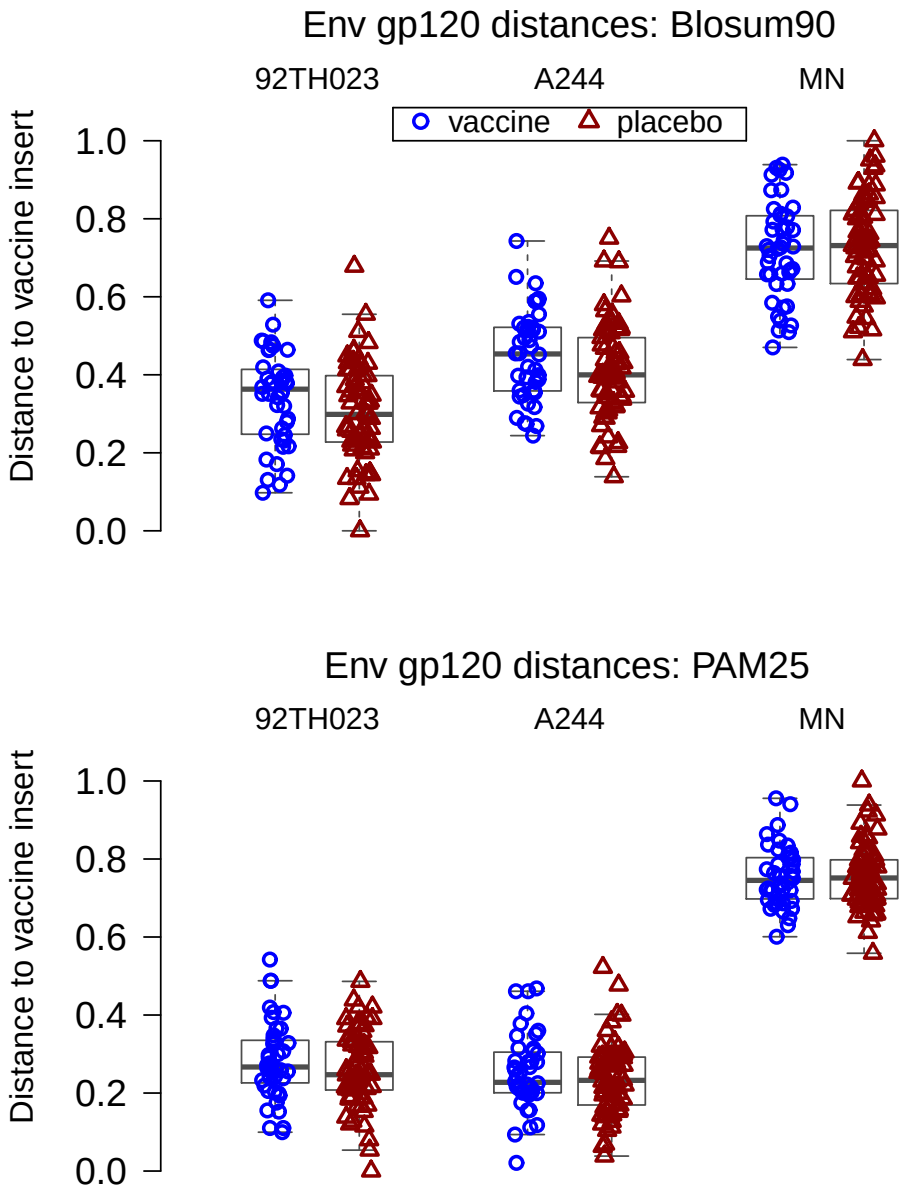


Figure B.6: Distribution of RV144 gp120 distances using the published set of gp120 monoclonal antibody contact sites.

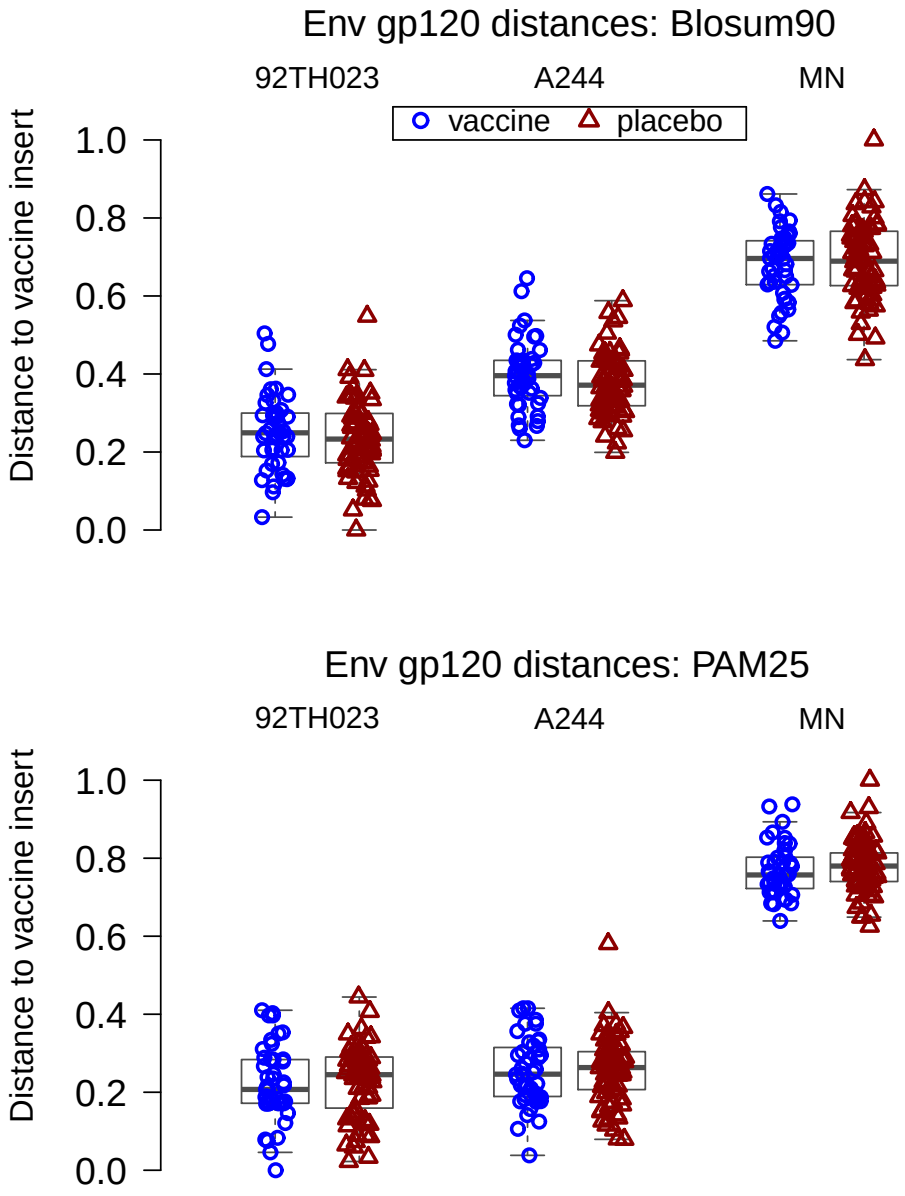


Figure B.7: Distribution of RV144 gp120 distances using the published set of gp120 monoclonal antibody contact sites and other published neutralization relevant contact sites.

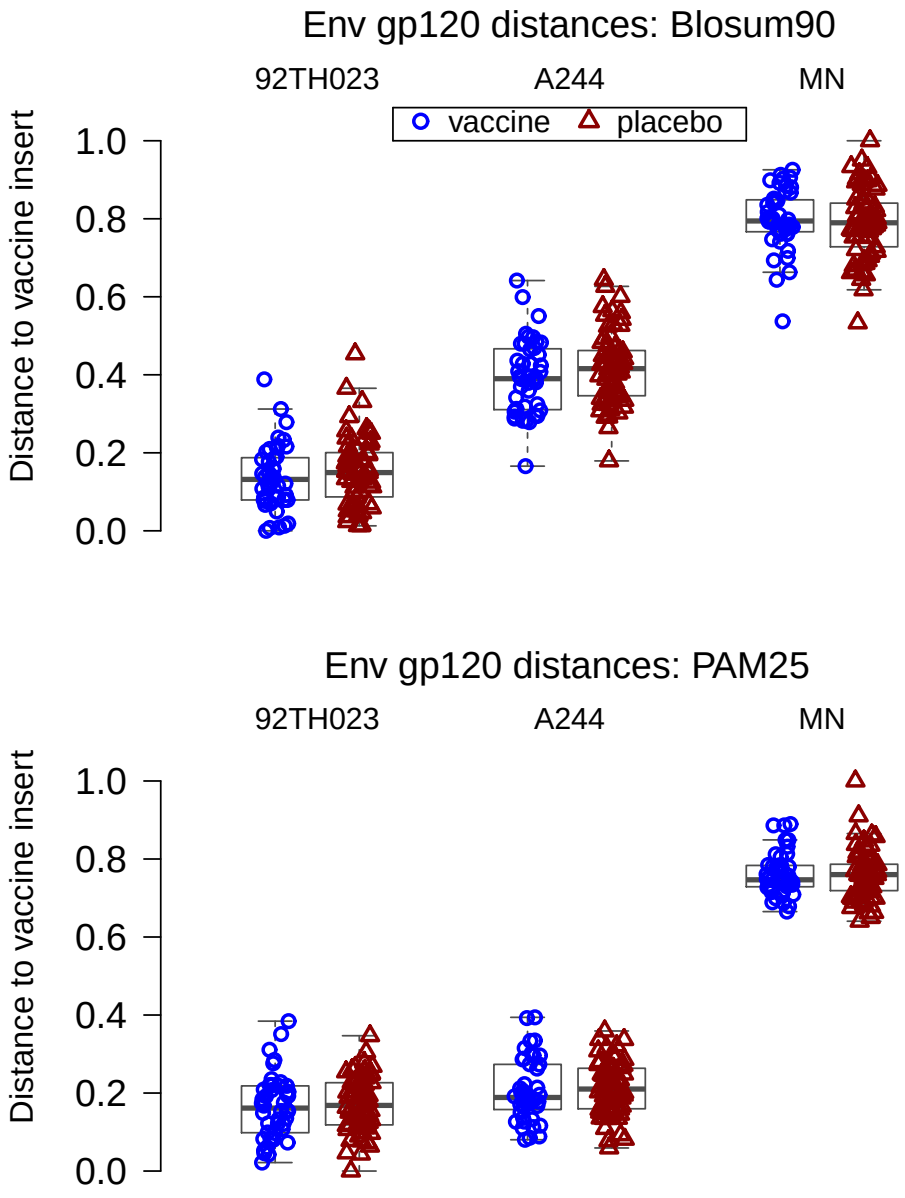


Figure B.8: Distribution of RV144 gp120 distances using hotspots (Haynes et al., 2012) in the linear peptide binding microarray analysis.

B.2 Inference about mark-specific vaccine efficacy

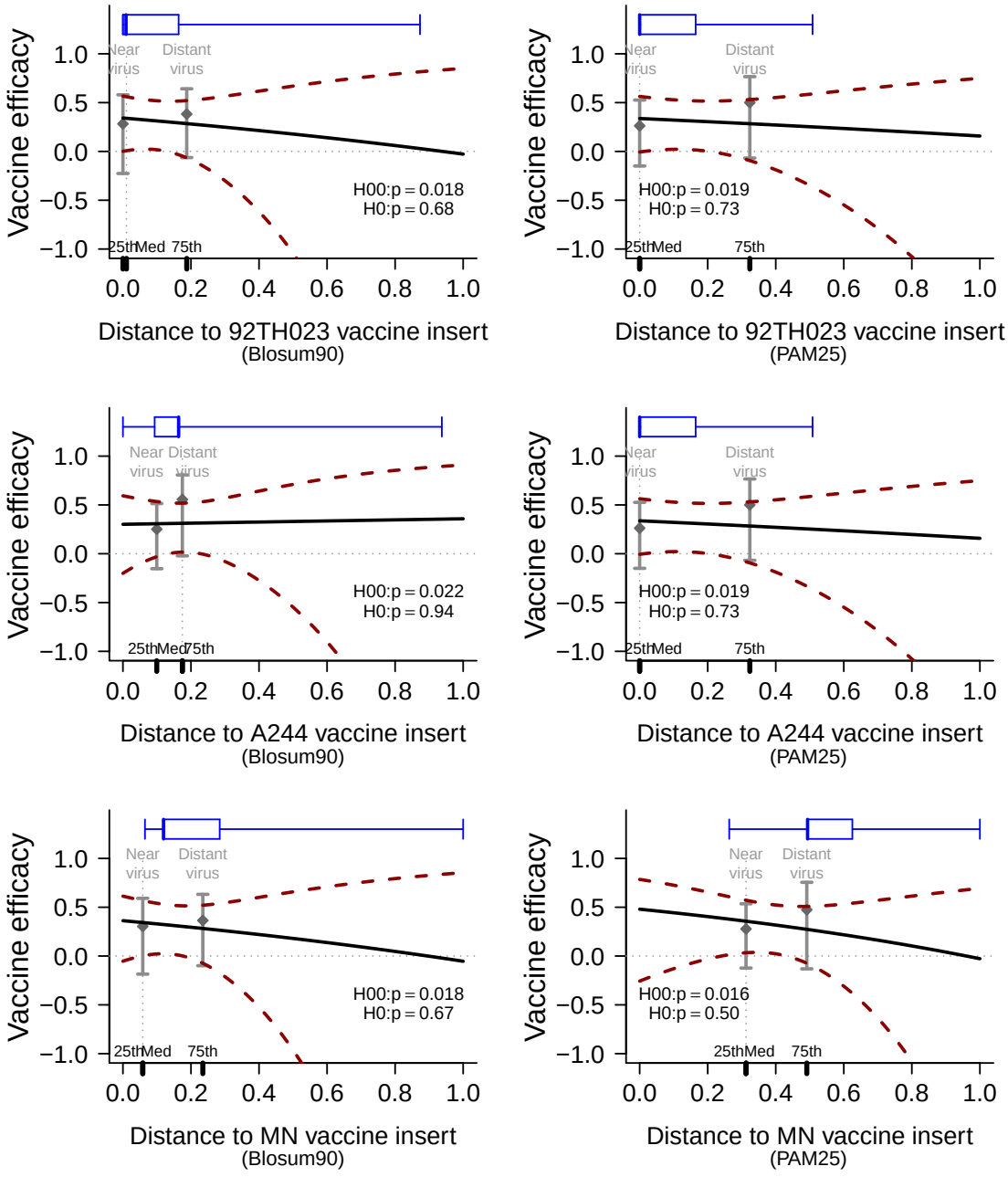


Figure B.9: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for V1/V2-focused distances (rescaled to the 0–1 interval [original distribution in boxplot]) using the published set of monoclonal antibody contact sites. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

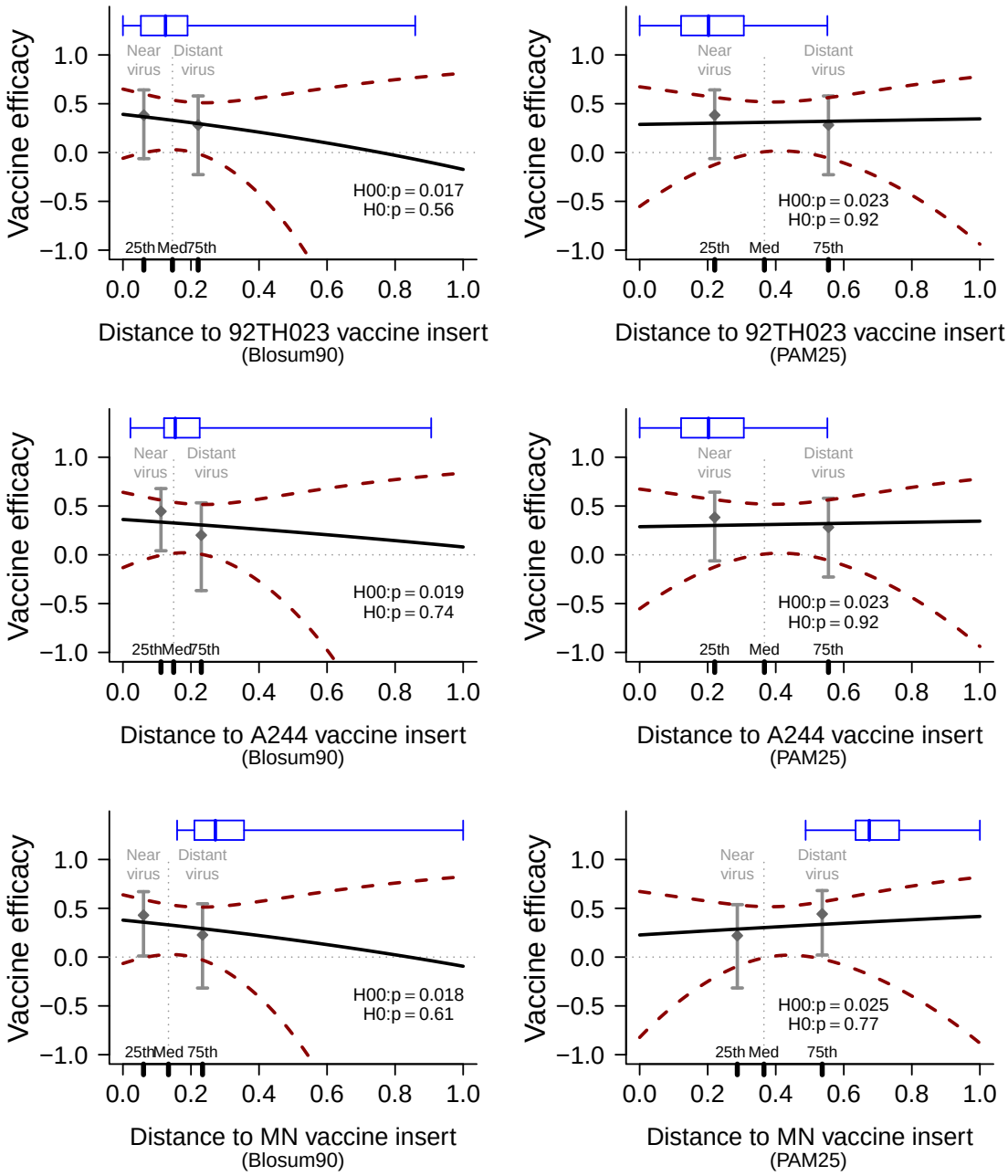


Figure B.10: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for V1/V2-focused distances (rescaled to the 0–1 interval [original distribution in boxplot]) using the published set of monoclonal antibody and other neutralization relevant contact sites. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

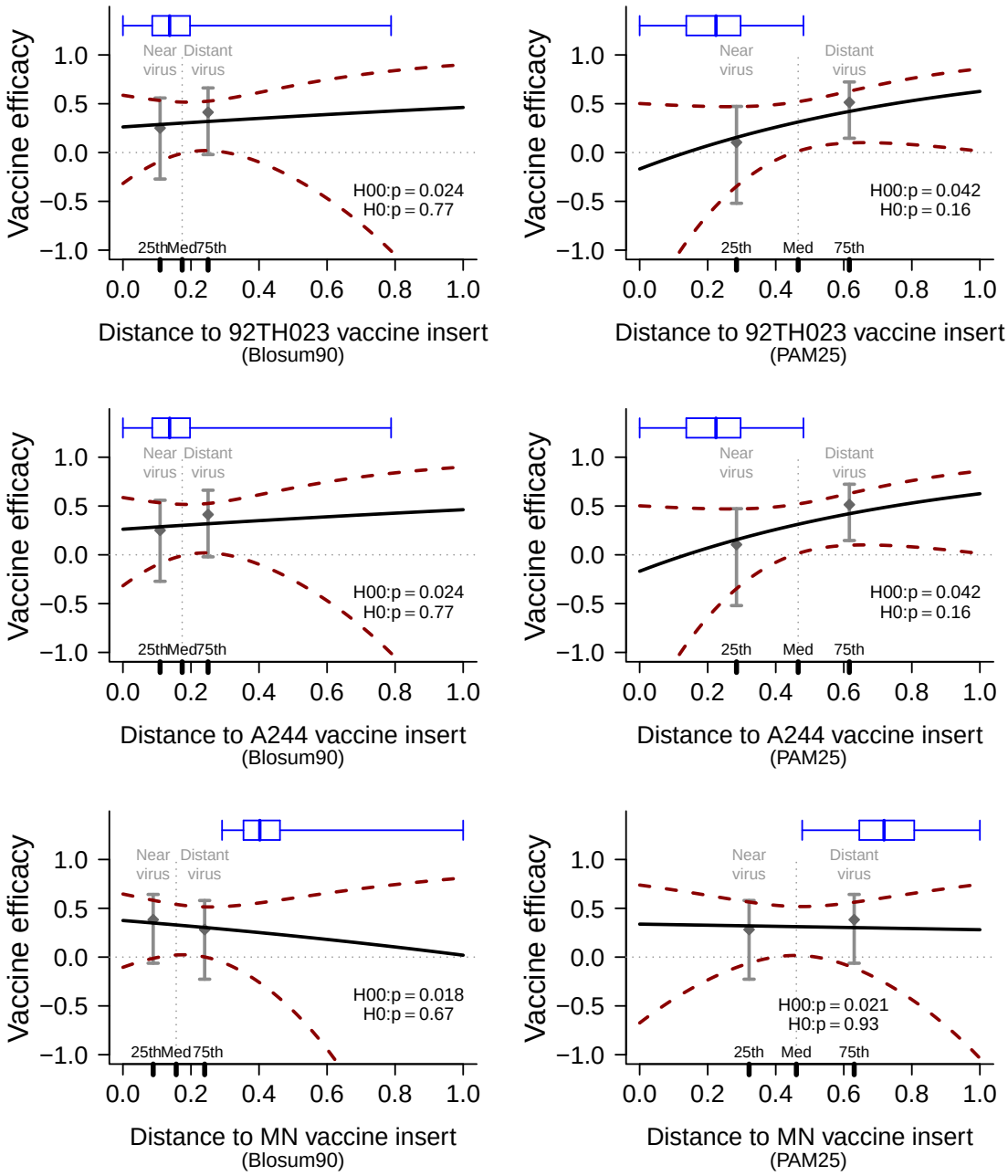


Figure B.11: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for V1/V2-focused distances (rescaled to the 0–1 interval [original distribution in boxplot]) using 22 V1/V2 sites with the highest frequency of occurrence in predicted antibody epitopes based on structural biology. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

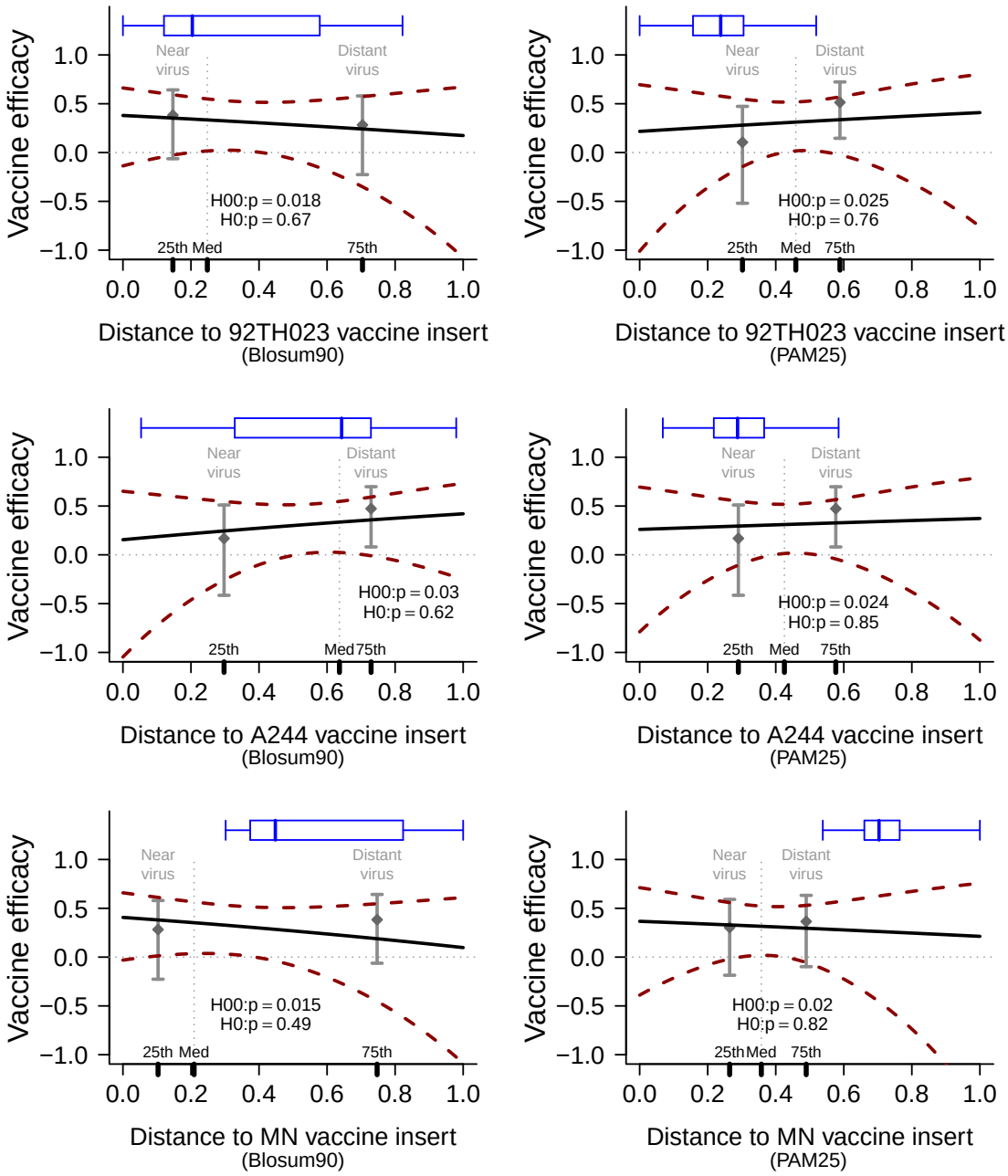


Figure B.12: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for V1/V2-focused distances (rescaled to the 0–1 interval [original distribution in boxplot]) using linear peptide microarray hotspots. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

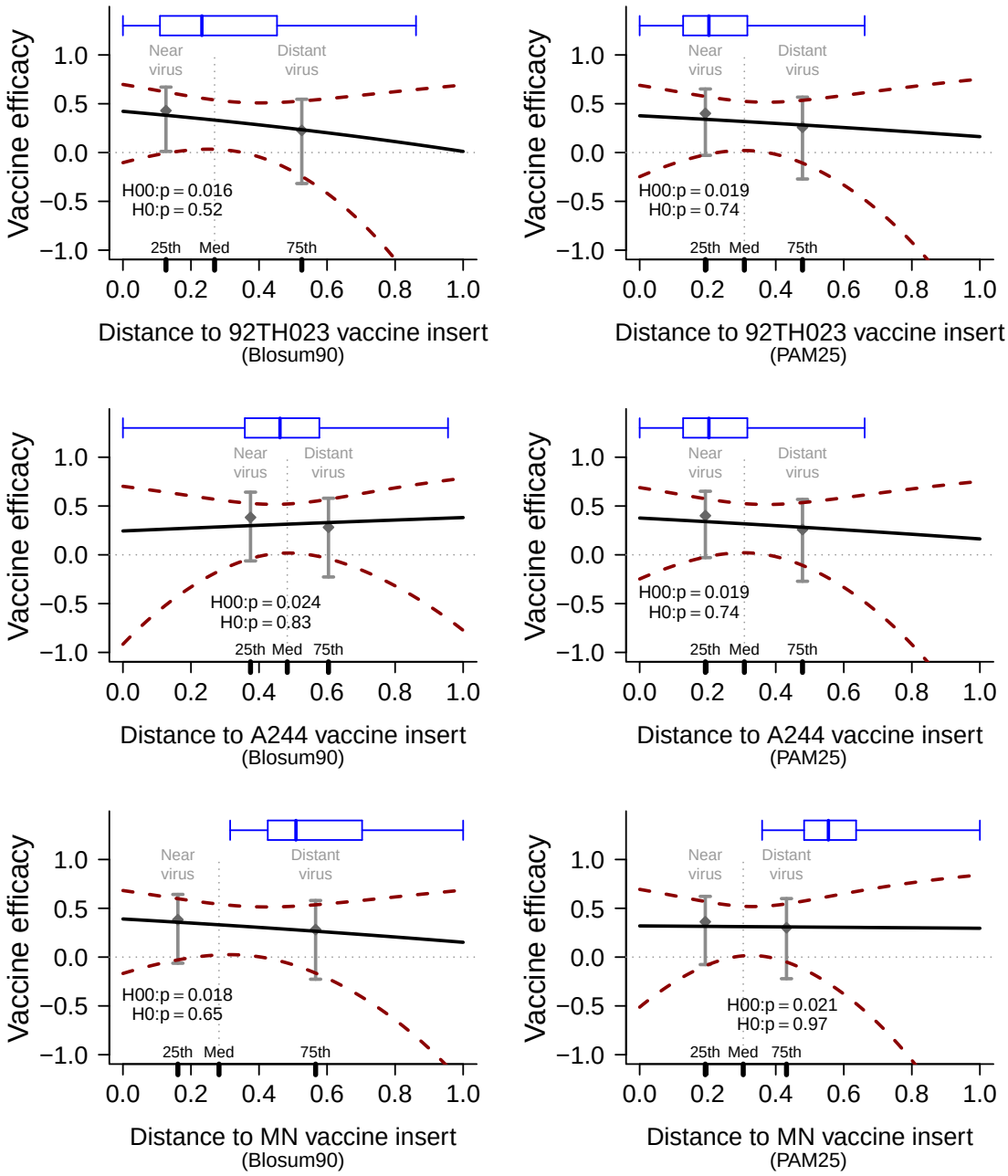


Figure B.13: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for V1/V2-focused distances (rescaled to the 0–1 interval [original distribution in boxplot]) using the intersection of the published set of monoclonal antibody and other neutralization relevant V1/V2 contact sites with linear peptide microarray hotspots. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

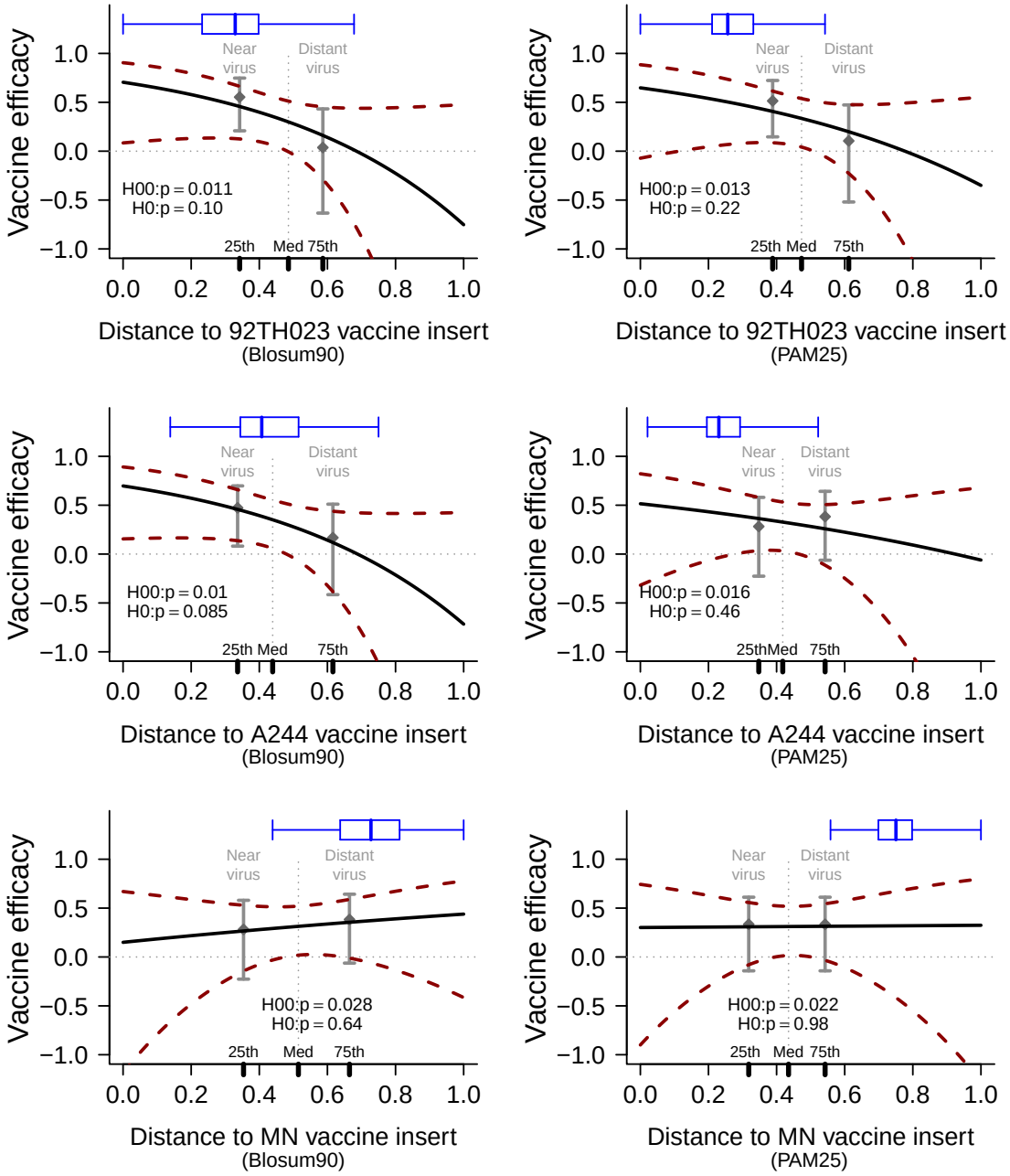


Figure B.14: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for gp120 distances (rescaled to the 0–1 interval [original distribution in box-plot]) using the published set of gp120 monoclonal antibody contact sites. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

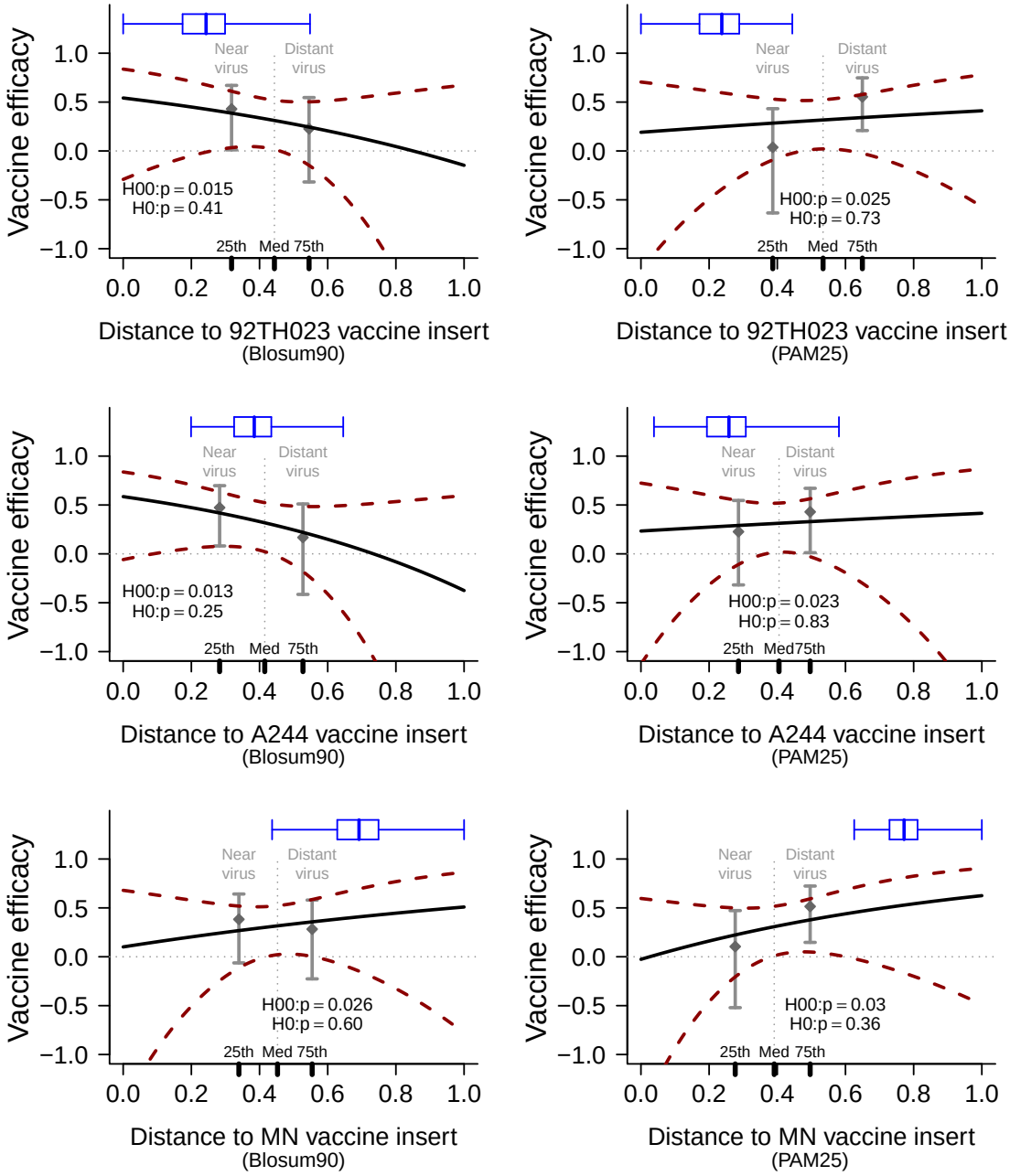


Figure B.15: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for gp120 distances (rescaled to the 0–1 interval [original distribution in boxplot]) using the published set of gp120 monoclonal antibody and other neutralization relevant contact sites. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

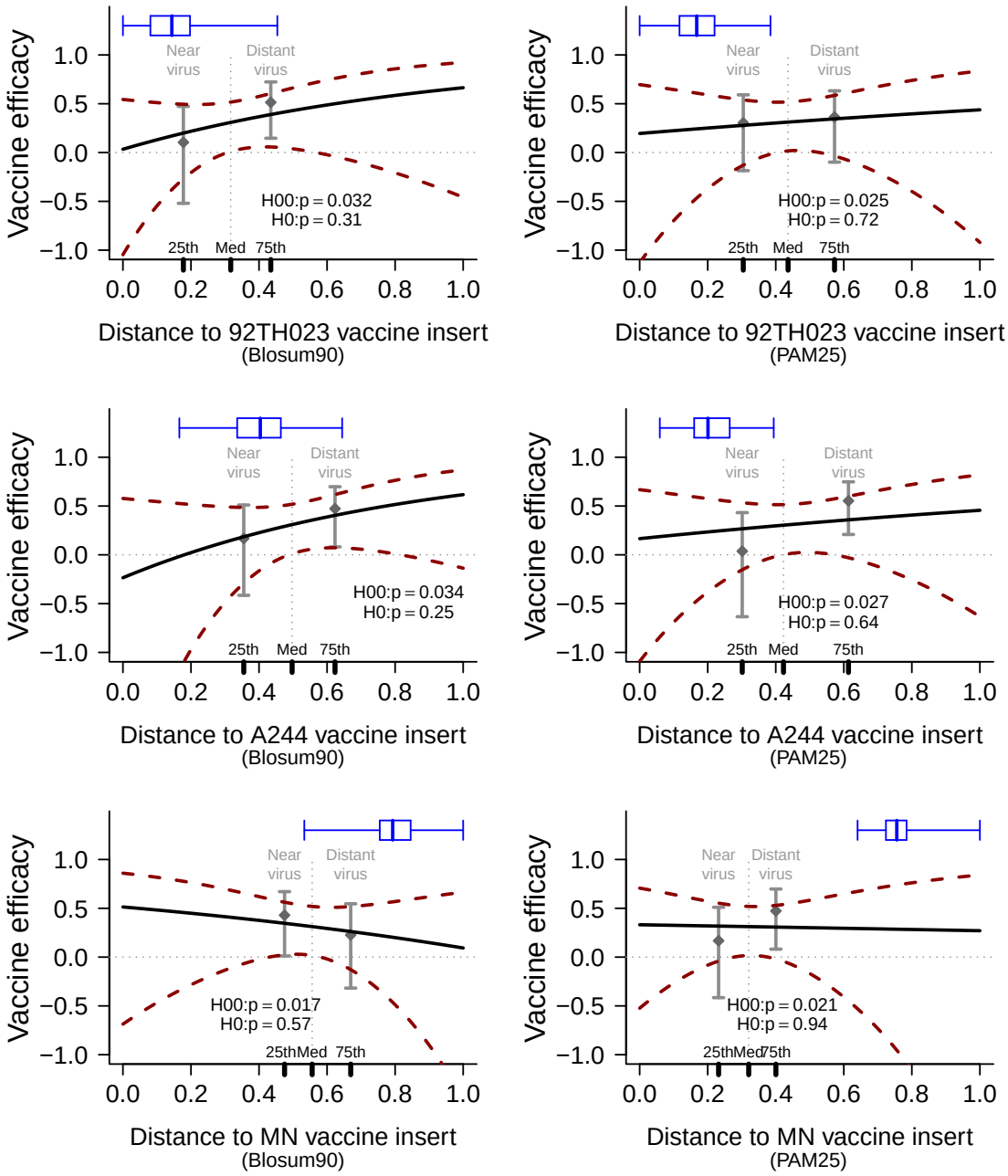


Figure B.16: Estimated mark-specific vaccine efficacy with 95% pointwise confidence bands for gp120 distances (rescaled to the 0–1 interval [original distribution in boxplot]) using linear peptide microarray hotspots. Included are p-values from the weighted Wald test of $\{H_0^0 : VE(v) = 0\}$ and the likelihood ratio test of $\{H_0 : VE(v) = VE\}$, and vaccine efficacy estimates from the competing risks Cox model for the dichotomized mark.

VITA

Michal Juraska was born in Poprad, Slovakia on September 15th, 1982. In 2007, he received a Master of Science in Mathematical Statistics from Charles University in the Czech Republic. In 2012, he earned a Doctor of Philosophy in Biostatistics at the University of Washington. Starting in May 2012, Michal will be a senior biostatistician at Novartis Oncology in Florham Park, New Jersey.