HIV VACCINE TRIALS NETWORK

Methods for Comparing Mark-specific Hazards and Cumulative Incidence Functions Between Two Groups, with Application to HIV Vaccine Trials

Peter Gilbert
Fred Hutchinson Cancer Research Center

Ian McKeague
Florida State University

Yanqing Sun
University of North Carolina at Charlotte

HIV VACCINE TRIALS NETWORK

Outline

- Motivating Example: HIV vaccine efficacy trial
- Statistical Methods
  - Hypothesis testing
  - Estimation
- Illustration

HIV VACCINE TRIALS NETWORK

Example: HIV Vaccine Efficacy Trial

- **Primary objective:** Assess vaccine efficacy (VE) to prevent HIV infection
- **Secondary objective:** Assess if and how VE varies with genotypic/phenotypic characteristics of HIV
- For each infected subject, measure the **distance** V between the infecting virus and the virus(es) represented in the vaccine
- **Available data:**
  - Vaccine group: $(T_{1i}, \delta_{1i}, \delta_{1i}V_{1i}), \quad i = 1, \cdots, n_1$
  - Placebo group: $(T_{2i}, \delta_{2i}, \delta_{2i}V_{2i}), \quad i = 1, \cdots, n_2$

Example: HIV Vaccine Efficacy Trial

- **Case 1:** V a small number of ordered categories
  - E.g.: $V \in \{0, 1, 2, 3+\}$ substitutions/deletions in the HIV V3 loop tip sequence GPGRAF
  - For each strain category j, can study $VE(t, j)$ using cause-specific hazard functions or cumulative incidence functions:

$$VE(t,j) = 1 - \frac{\lambda_{1j}(t)}{\lambda_{2j}(t)} \quad \text{or} \quad VE(t,j) = 1 - \frac{F_{1j}(t)}{F_{2j}(t)}$$

Prentice et al. (1978, Biometrics);
Gilbert (2000, Statistics in Medicine)

Example: HIV Vaccine Efficacy Trial

- **Case 2:** V a large number of ordered categories
  - E.g.: percent amino acid mismatch computed over hundreds or thousands of positions

    $\Rightarrow$ Treat V as continuous, $V \in [0, 1]$
  - Gilbert et al. (1999, Biometrika; 2000, Annals of Statistics) developed semiparametric methods for studying $OR(v)$
    - $OR(v) =$ odds that the infecting strain has distance v for vaccine versus placebo recipients

HIV VACCINE TRIALS NETWORK

Example: HIV Vaccine Efficacy Trial

- **Semiparametric biased sampling model:**

$$F_1(v,\theta) = \frac{\int_0^v w(z,\theta)dF_2(z)}{\int_0^\infty w(z,\theta)dF_2(z)}$$

- **Limitations:**
  - Interpretation conditional on infection
  - $OR(v)$ assumed to satisfy a parametric form
  - Does not account for time to HIV infection

Statistical Methods

- **Objective:** Develop methods for testing and estimation of $VE(t,v)$ defined based on continuous mark-specific hazard and cumulative incidence functions
  - Mark-specific hazard functions:

$$\lambda_k(t,v) = \lim_{h_1,h_2 \to 0} P\{T_k \in [t, t+h_1), V_k \in [v, v+h_2)\}/h_1 h_2$$

  - Mark-specific cumulative incidence functions:

$$F_k(t,v) = \lim_{h_2 \to 0} P\{T_k \leq t, V_k \in [v, v+h_2)\}/h_2$$

- The functions have a *crude* (not *net*) interpretation

HIV VACCINE TRIALS NETWORK

Hypothesis Testing

- Define $VE(t,v) = 1 - \frac{\lambda_1(t,v)}{\lambda_2(t,v)}; \quad VE(t) = 1 - \frac{\lambda_1(t)}{\lambda_2(t)}$
- Test

$$
\begin{aligned}
H_0 : VE(t,v) &= VE(t) \text{ for all } t \in [0,\tau] \\
&\text{versus} \\
H_1 : VE(t,v_1) &\leq VE(t,v_2) \text{ for all } v_1 \leq v_2, t \in [0,\tau] \\
H_2 : VE(t,v_1) &\neq VE(t,v_2) \text{ for some } v_1 \leq v_2, t \in [0,\tau]
\end{aligned}
$$

- $H_0 \Leftrightarrow \lambda_1(t,v)/\lambda_2(t,v)$ does not depend on v

Hypothesis Testing

- Define doubly cumulative mark-specific hazard functions

$$\Lambda_k(t,v) = \int_0^v \int_0^t \lambda_k(s,u)\, ds\, du, \qquad k = 1,2$$

- **Idea of testing procedures:** Compare a nonparametric estimate of $\Lambda_1(t,v) - \Lambda_2(t,v)$ with an estimate under H_0

Hypothesis Testing

- Nelson–Aalen-type estimator (Huang and Louis, 1998, Biometrika):

$$\widehat{\Lambda}_k(t,v) = \int_0^t \frac{N_k(ds,v)}{Y_k(s)},\ t \geq 0,\ v \in [0,1]$$

$$Y_k(t) = \sum_{i=1}^{n_k} I(X_{ki} \geq t)$$

$$N_k(t,v) = \sum_{i=1}^{n_k} I(X_{ki} \leq t, \delta_{ki} = 1, V_{ki} \leq v)$$

Hypothesis Testing

- H_0 holds $\Leftrightarrow$

$$\Lambda_1(t,v) = \int_0^t \frac{\lambda_1(s)}{\lambda_2(s)} \Lambda_2(ds,v)$$

- Under H_0, estimate $\Lambda_1(t,v) - \Lambda_2(t,v)$ by

$$\int_0^t \left[\frac{\widehat{\lambda}_1(s)}{\widehat{\lambda}_2(s)} - 1 \right] \widehat{\Lambda}_2(ds,v)$$

with

$$\widehat{\lambda}_k(t) = \frac{1}{b_k} \int_{u_1}^{u_2} K\left(\frac{t-s}{b_k} \right) d\widehat{\Lambda}_k(s)$$

HIV VACCINE TRIALS NETWORK

Test Process and Test Statistics

- **Test process:**

$$L_n(t,v) = \sqrt{\frac{n_1 n_2}{n}} \int_0^t H_n(s) \left[\widehat{\Lambda}_1(ds,v) - \frac{\widehat{\lambda}_1(s)}{\widehat{\lambda}_2(s)} \widehat{\Lambda}_2(ds,v) \right]$$

- **Test statistics:**

$$\widehat{U}_1 = \sup_{v_1<v_2} \sup_{0<t_1<t_2<\tau} \{L_n(t_2,v_2) - L_n(t_2,v_1) - L_n(t_1,v_2) + L_n(t_1,v_1)\}$$

$$\widehat{U}_2 = \sup_{0\le v\le 1} \sup_{0<t_1<t_2<\tau} |L_n(t_2,v) - L_n(t_1,v)|$$

Theoretical Properties

- **Theorem 1:** Under regularity conditions

$$L_n(t,v) \rightarrow^d L(t,v)$$

  in $D([0,\tau]\times[0,1])$ as $n \rightarrow \infty$

- $\Rightarrow$ Under $H_0, \hat{U}_1 \rightarrow^d U_1$ and $\hat{U}_2 \rightarrow^d U_2$
- Let $c_{1\alpha}$ and $c_{2\alpha}$ be the $(1-\alpha)$ quantile of U_1 and U_2
  - $P(\hat{U}_j > c_{j\alpha}) \rightarrow \alpha$ under H_0

Theoretical Properties

- **Theorem 2:** Under regularity conditions

$$P(\hat{U}_1 > c_{1\alpha}) \rightarrow 1 \text{ under } H_1$$
$$P(\hat{U}_2 > c_{2\alpha}) \rightarrow 1 \text{ under } H_2$$

- Critical values $c_{j\alpha}$ unknown and difficult to obtain
- $\Rightarrow$ Use a resampling procedure to approximate $c_{j\alpha}$

HIV VACCINE TRIALS NETWORK

Monte-Carlo Procedure

- Let $W_{1i} \sim N(0,1), i = 1, \cdots, n_1; W_{2i} \sim N(0,1), i = 1, \cdots, n_2$
- Define a simulated test process $\tilde{L}_n(t,v)$, a function of:
  - W_{1i}, W_{2i}
  - $\widehat{\lambda}_1(t), \widehat{\lambda}_2(t)$
  - A smooth estimate of $\Lambda_2'(t,v) = \frac{d}{ds}\Lambda_2(s,v)|_{s=t}$
- **Theorem 3:** Under regularity conditions, conditional on the observed data sequence

$$\tilde{L}_n(t,v) \rightarrow^d L(t,v)$$

in $D([0,\tau] \times [0,1])$ under H_0 as $n \rightarrow \infty$

HIV VACCINE TRIALS NETWORK

Monte-Carlo Procedure

- Acceptance/Rejection procedure:
  - Compute $\widehat{\tilde{U}}_1$ and $\widehat{\tilde{U}}_2$ based on $\tilde{L}_n(t,v)$
  - Based on B replicates $\widehat{\tilde{U}}_j$, compute

    $\widehat{c}_{j\alpha} = (1-\alpha)^{th}$ percentile of $\widehat{\tilde{U}}_{j1}, \cdots, \widehat{\tilde{U}}_{jB}$
  - Reject H_0 if $\widehat{U}_j > \widehat{c}_{j\alpha}$

HIV VACCINE TRIALS NETWORK

Estimation Procedure

- Sample size too small to reliably estimate

  $VE(t,v) = 1 - \frac{\lambda_1(t,v)}{\lambda_2(t,v)}$

- Alternatively, consider

$$\begin{aligned} VE(t,v) &= 1 - \frac{F_1(t,v)}{F_2(t,v)} \\ &= 1 - \lim_{h \to 0} \frac{P(T \leq t, V \in [v, v+h) | 1)}{P(T \leq t, V \in [v, v+h) | 2)} \end{aligned}$$

HIV VACCINE TRIALS NETWORK

Estimation Procedure

- Estimate $VE(t,v)$ by $1-\frac{\widehat{F}_1(t,v)}{\widehat{F}_2(t,v)}$, where

$$\widehat{F}_k(t,v)=\frac{1}{2b}\int_0^t \frac{\widehat{S}_k(s-)}{Y_k(s)}N_{vk}(ds)$$

$N_{vk}(t)=\sum_{i=1}^{n_k} I(X_i\le t,\delta_i=1,v-b<V_i\le v+b)$

$\widehat{S}_k(t)=$ Kaplan-Meier estimate of $S_k(t)$

- $\widehat{F}_k(t,v)=$ continuous analog of $\widehat{F}_{kj}(t)$ for a discrete mark j

HIV VACCINE TRIALS NETWORK

Estimation Procedure

- $Var\{\widehat{F}_k(t,v)\}$ can be estimated by

$$\frac{1}{(2b)^2}\int_0^t \left[\frac{\widehat{S}_k(s-)}{Y_k(s)}\right]^2 N_{vk}(ds)$$

- 95% pointwise confidence intervals:

$$\widehat{VE}(t,v) \pm 1.96 \times \widehat{Var}\{\widehat{VE}(t,v)\}^{1/2}$$

HIV VACCINE TRIALS NETWORK

Illustration

- First preventive HIV vaccine efficacy trial completed in February 2003
- AIDSVAX, a bivalent recombinant gp120 vaccine, developed and tested by VaxGen, Inc.
- Trial conducted in the U.S./Netherlands/Canada/Carribean, $n = 5403$, 2:1 randomization to vaccine:placebo
- Volunteers tested for HIV infection every 6 months for 3 years
- For HIV infected subjects, the gp120 region of HIV was sequenced

Results of Trial

- **Primary analysis:**

	Number Randomized	Number Infected	Percent Infected
Vaccine	3598	241	6.7%
Placebo	1805	127	7.0%

$$\widehat{VE} = 5.9\%, \quad 95\% \ CI \ \ (-16.7\%, 24.2\%), p = 0.58$$

HIV VACCINE TRIALS NETWORK

Sieve Analysis Illustration

- Define V as the percent mismatch in V1-V2-V3 of the infecting strain relative to the MN vaccine strain
- Two pseudo examples:

  1. (**null case**) Use the real failure times, indicators, and marks, and randomly permute the vaccination statuses to achieve $\approx$ 2:1 vaccine:placebo ratio

Sieve Analysis Illustration

2. (**alternative case**) Use the real failure times, indicators, and marks, and select the vaccination statuses such that

$$P(Z=1|V=v)=\frac{\exp\{\alpha+\beta v\}}{1+\exp\{\alpha+\beta v\}}$$

with α and β chosen such that $P(Z=1|V=\bar{V})=2/3$ and $P(Z=1|V=\max(V))=0.99$

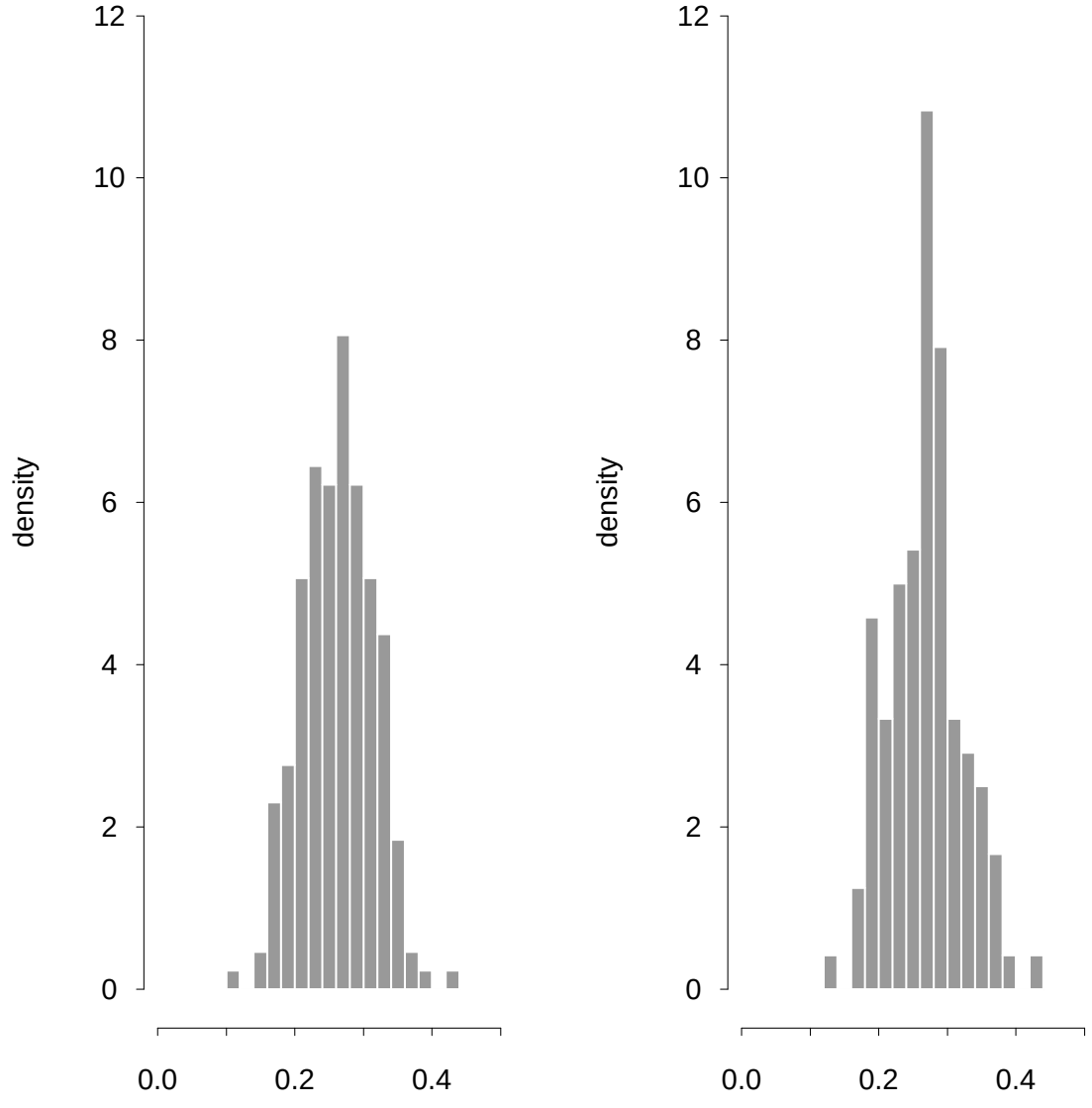
HIV VACCINE TRIALS NETWORK

Example 1 (null case)

	n	mean	range
Vaccine	217	0.348	0.12-0.43
Placebo	120	0.335	0.14-0.44

Example 1 (null case)

Distributions of V1-V2-V3 Strain Distance

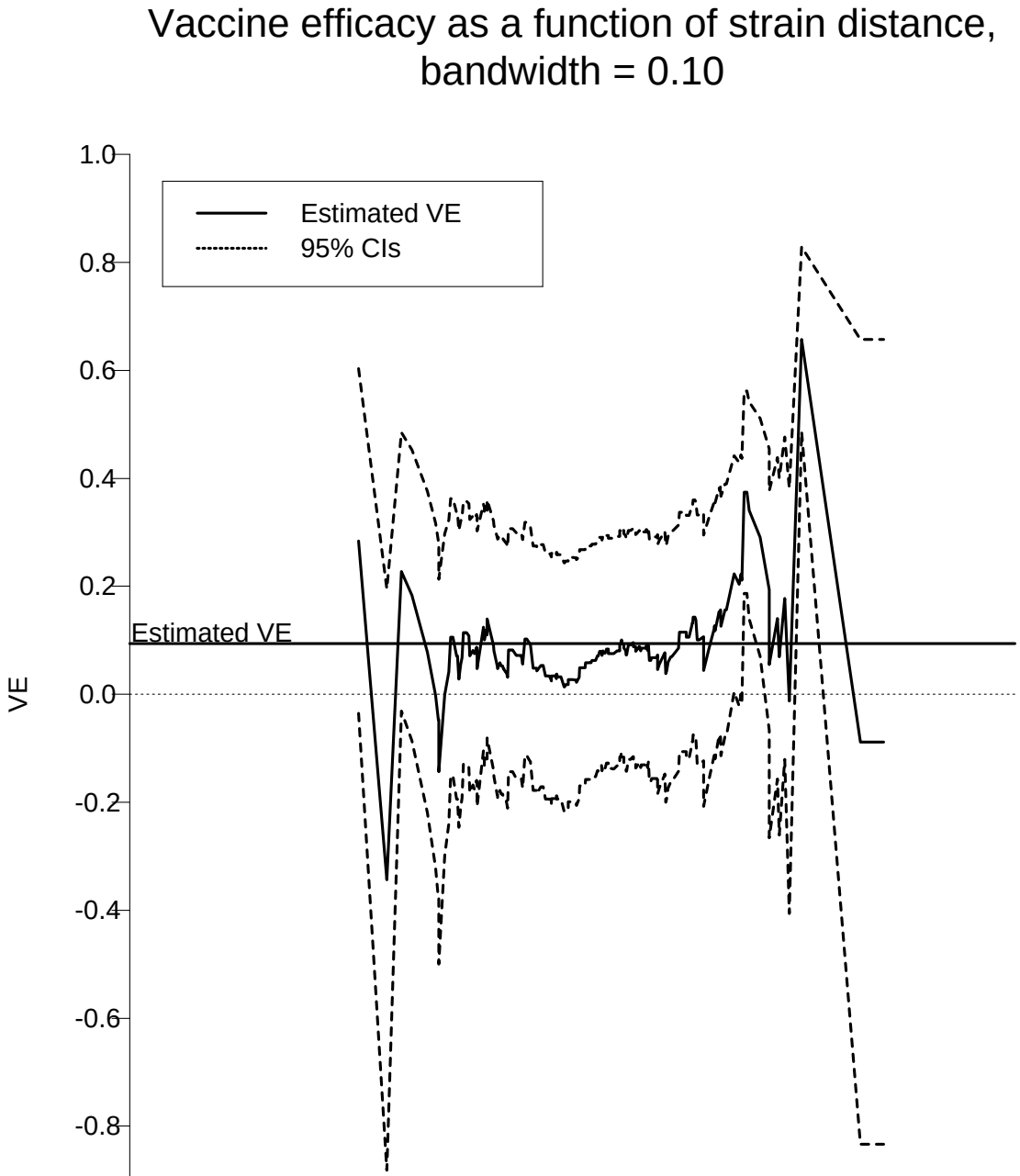


HIV VACCINE TRIALS NETWORK

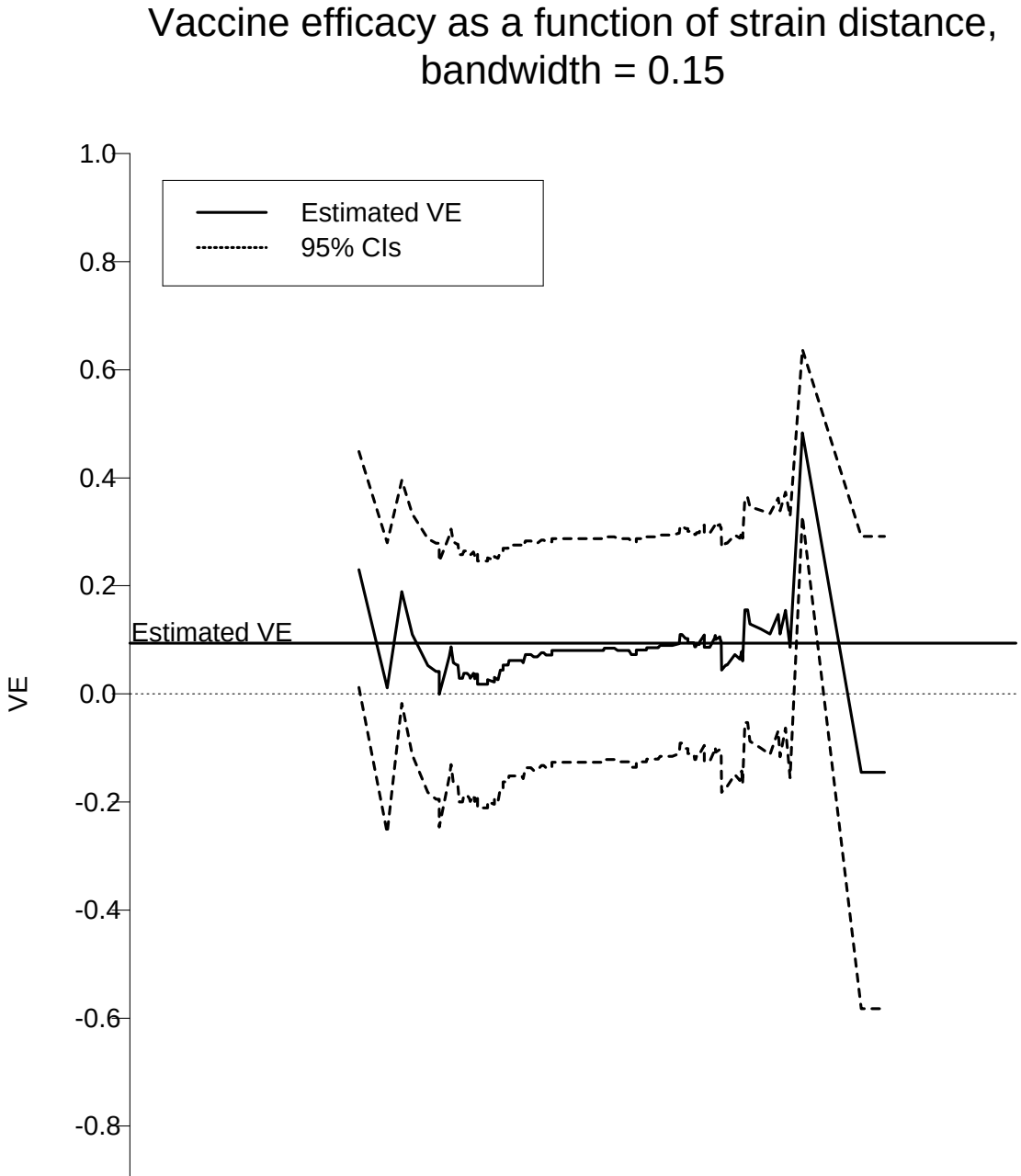
Example 1 (null case)

- Implementation of testing and estimation procedures:
  - Time range 2-36 months
  - bandwidth $= 8.5$ months
  - Distance range 0.12-0.44
  - bandwidth $= 0.10, 0.15, 0.20$
- $\tilde{U}_1 = 0.348, \tilde{U}_2 = 0.335$
  - Based on 1000 simulations, $p_1 = 0.677, p_2 = 0.523$

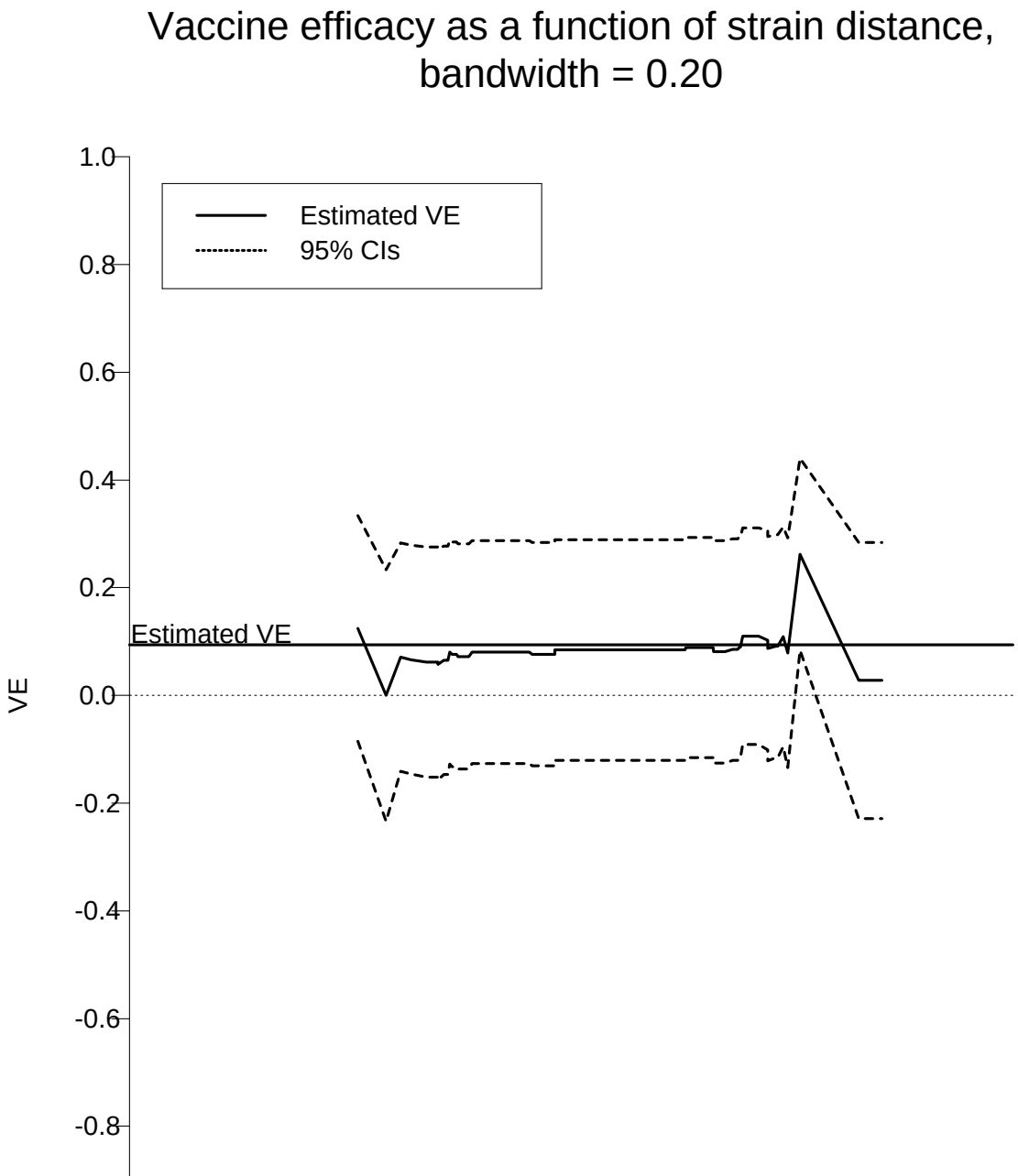
Example 1 (null case)



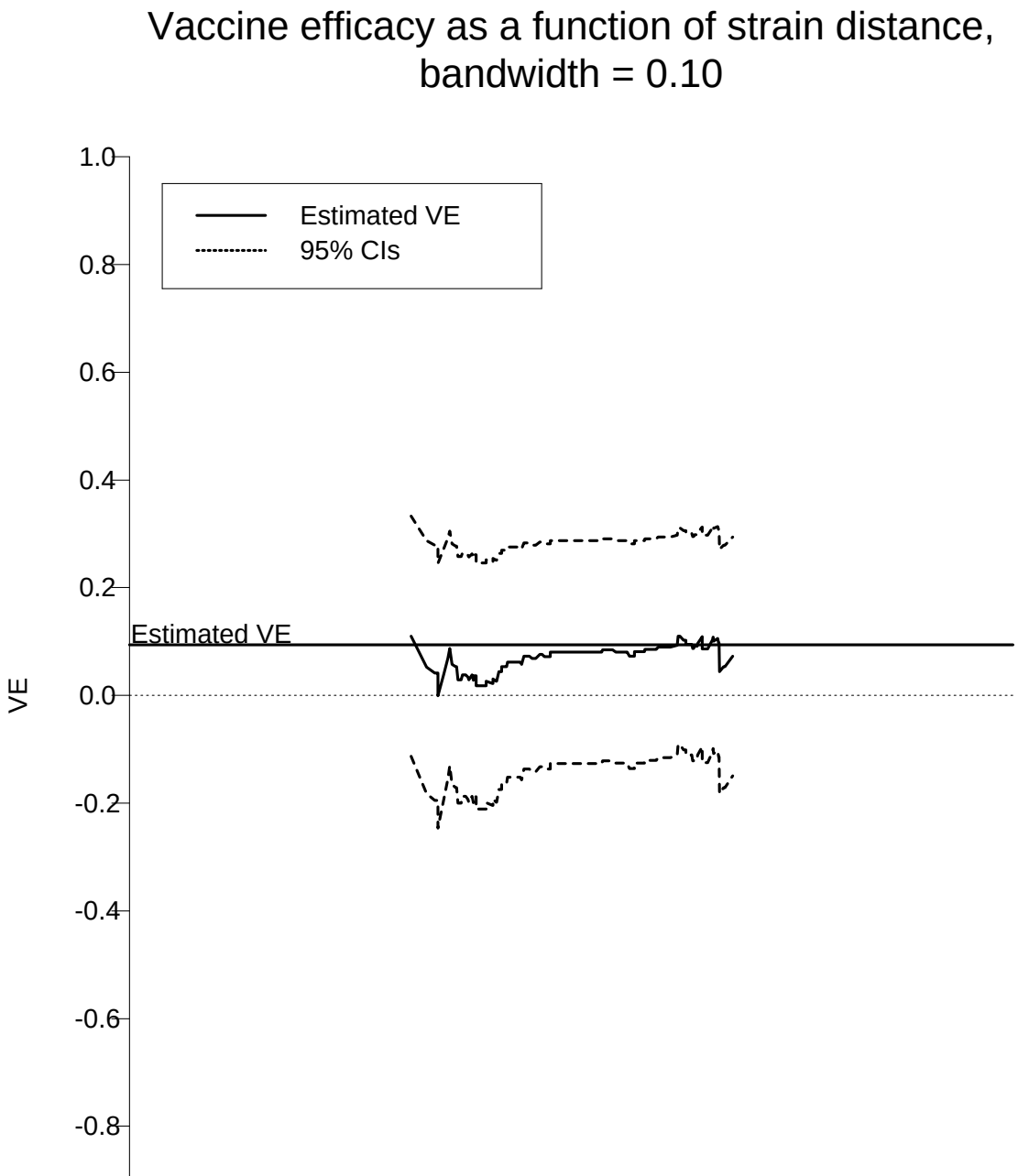
Example 1 (null case)



Example 1 (null case)



Example 1 (null case)



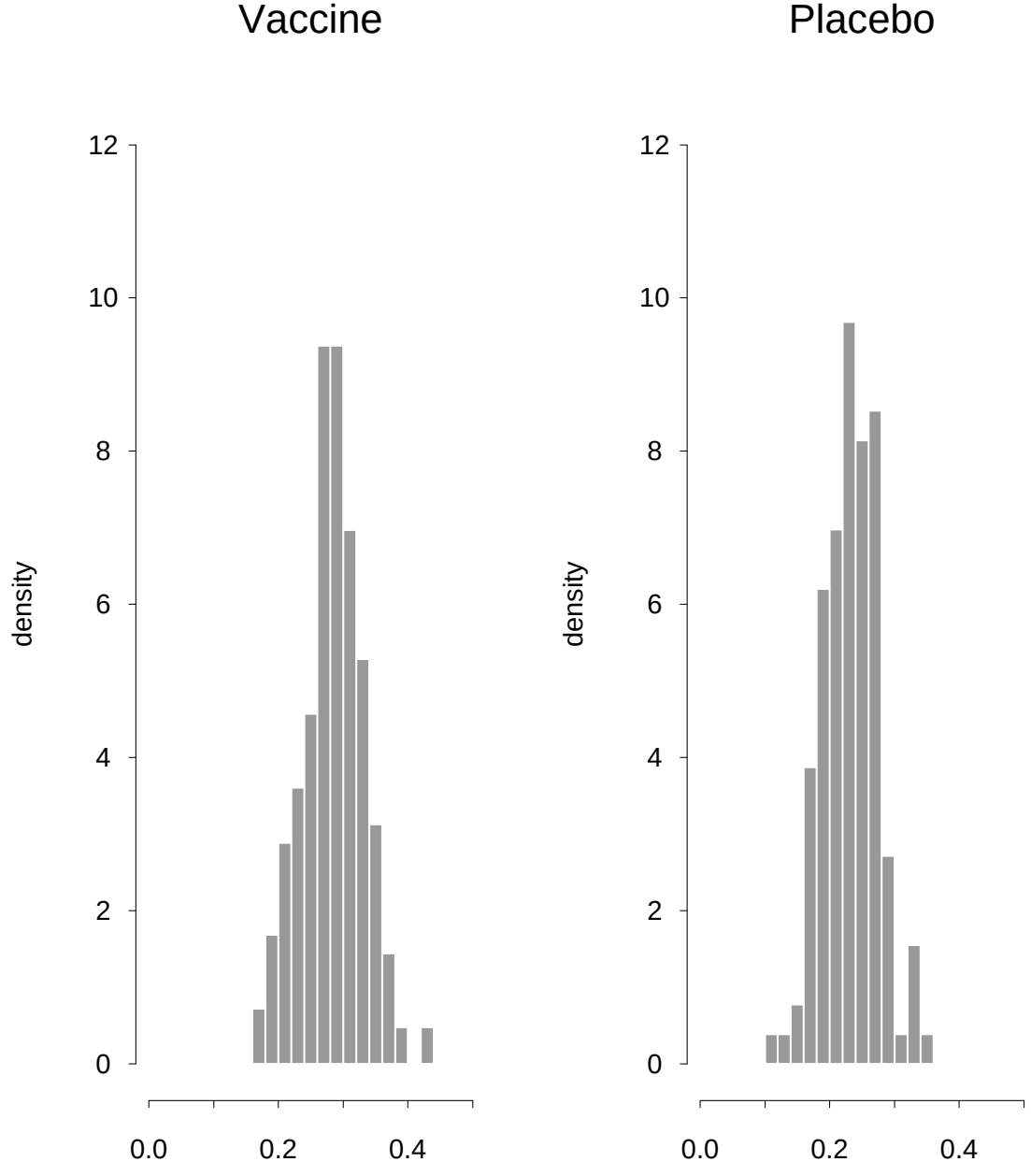
HIV VACCINE TRIALS NETWORK

Example 2 (alternative case)

	n	mean	range
Vaccine	208	0.285	0.17-0.44
Placebo	129	0.232	0.12-0.34

Example 2 (alternative case)

Distributions of V1-V2-V3 Strain Distance



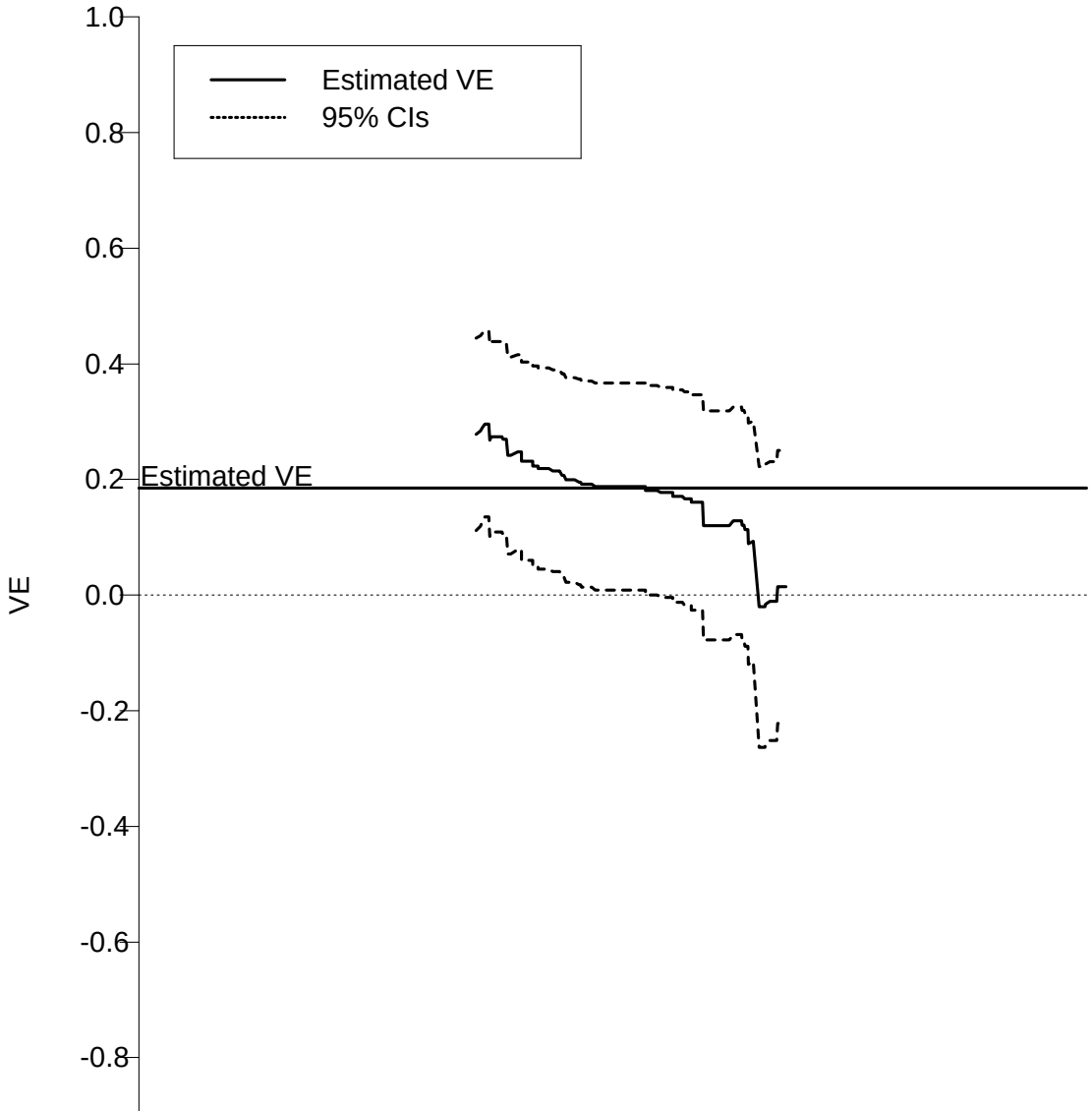
HIV VACCINE TRIALS NETWORK

Example 2 (alternative case)

- Implementation of testing and estimation procedures:
  - Time range 2-36 months
  - bandwidth $= 8.5$ months
  - Distance range 0.15-0.35
  - bandwidth $= 0.15$
- $\tilde{U}_1 = 1.169, \tilde{U}_2 = 1.272$
  - Based on 1000 simulations, $p_1 < 0.001, p_2 < 0.001$

Example 2 (alternative case)

Vaccine efficacy as a function of strain distance,
bandwidth = 0.10



HIV VACCINE TRIALS NETWORK

Future Research

- Simultaneous confidence bands for $VE(t,v)$ in v for t fixed and in t for v fixed
- Study $VE(t,v)$ with covariate adjustment
  - Continuous mark-specific Cox regression model
- Causal inference/Sensitivity analysis to address the fundamental nonidentifiability problem for competing risks data
  - Principal stratification approach (Frangakis and Rubin, 2002, Biometrics)