Module 8:
Evaluating Vaccine Efficacy

Instructors:
Dean Follmann, Peter Gilbert, Betz Halloran, Erin Gabriel, Michael Sachs

Session 6: Effect Modifier Methods for Assessing Immunological Correlates of VE (Part I)

Summer Institute in Statistics and Modeling in Infectious Diseases
University of Washington, Department of Biostatistics

Course materials at:
http://faculty.washington.edu/peterg/SISMID2016.html

July 18–20, 2016

Statistical Center for HIV/AIDS Research & Prevention
SCHARP

FRED HUTCH
CURES START HERE™

Outline of Module 8: Evaluating Vaccine Efficacy

Session 1 (Halloran)	Introduction to Study Designs for Evaluating VE
Session 2 (Follmann)	Introduction to Vaccinology Assays and Immune Response
Session 3 (Gilbert)	Introduction to Frameworks for Assessing Surrogate Endpoints/Immunological Correlates of VE
Session 4 (Follmann)	Additional Study Designs for Evaluating VE
Session 5 (Gilbert)	Methods for Assessing Immunological Correlates of Risk and Surrogate Endpoints
Session 6 (Gilbert)	**Effect Modifier Methods for Assessing Immunological Correlates of VE (Part I)**
Session 7 (Gabriel)	Effect Modifier Methods for Assessing Immunological Correlates of VE (Part II)
Session 8 (Sachs)	Tutorial for the R Package *pseval* for Effect Modifier Methods for Assessing Immunological Correlates of VE
Session 9 (Gilbert)	Introduction to Sieve Analysis of Pathogen Sequences, for Assessing How VE Depends on Pathogen Genomics
Session 10 (Follmann)	Methods for VE and Sieve Analysis Accounting for Multiple Founders

Outline of Session 6

1. Effect Modification/VE Curve Framework
2. Identifiability and Estimation
3. Simulations
4. Discussion

Paper corresponding to this talk: Gilbert and Hudgens (2008, Biometrics)

Notation

- **Throughout consider a 2-arm trial with:**

  Z = treatment assignment (0 or 1)

  S = candidate surrogate endpoint measured at time τ after randomization

  Y = clinical endpoint (0 or 1) measured after time τ [The approach also applies for quantitative **Y**]

  $\mathbf{Y}^{\tau}$ = clinical endpoint (0 or 1) between time 0 and τ

Principal Surrogate Endpoints
(Frangakis and Rubin 2002, *Biometrics*)

- In the VE curve "principal surrogate" framework, the levels of S are not controlled/manipulated/assigned, they are what they happen to be

- **Notation**
  - $S_i(Z)$ = potential immune response endpoint under assignment Z; for $Z = 0, 1$
  - $Y_i(Z)$ = potential clinical endpoint under assignment Z; for $Z = 0, 1$
  - $Y^{\tau}_i(Z)$ = potential clinical endpoint under assignment Z; for $Z = 0, 1$

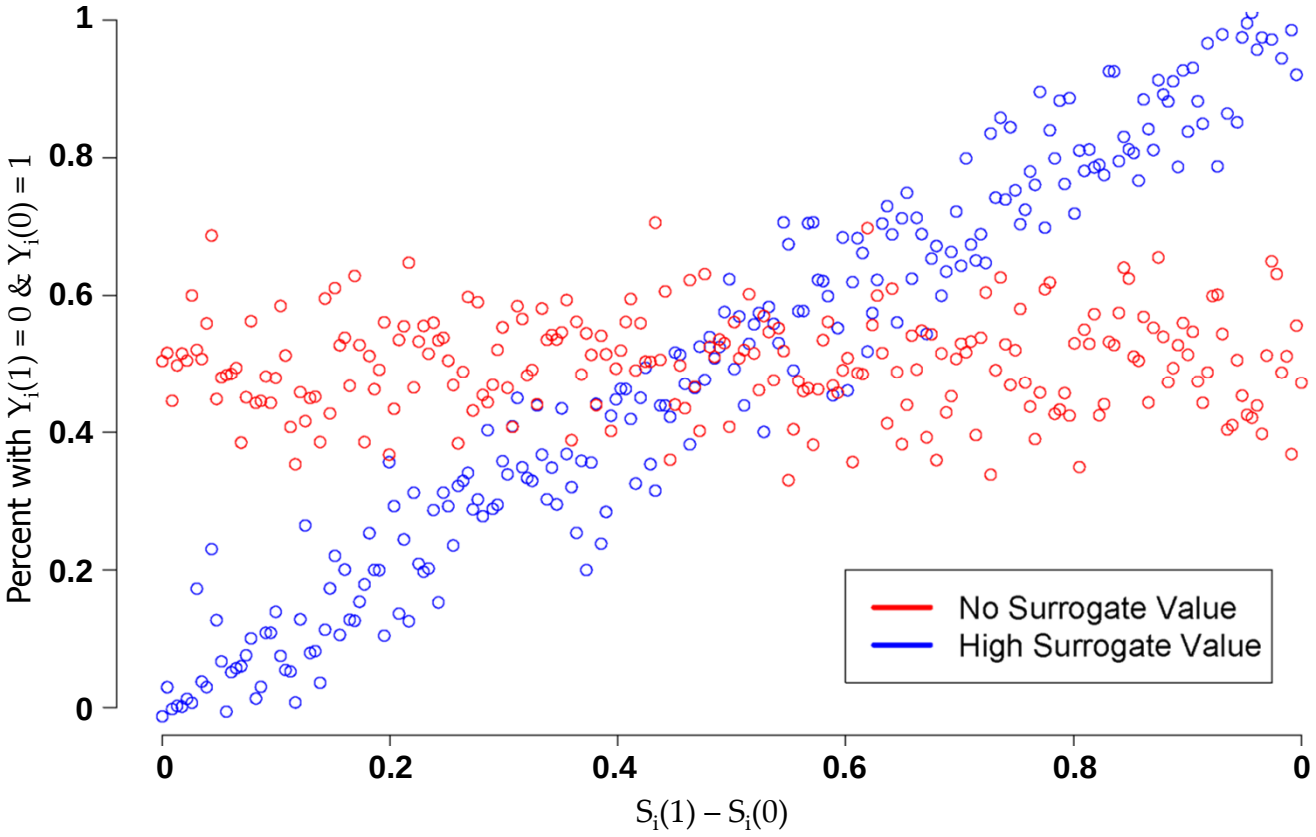
- **Causal Effects**
  - A contrast in $S_i(1)$ and $S_i(0)$ is a causal effect on S for subject i
  - A contrast in $Y_i(1)$ and $Y_i(0)$ is a causal effect on Y for subject I
  - A contrast in $Y^{\tau}_i(1)$ and $Y^{\tau}_i(0)$ is a causal effect on Y^{τ} for subject i

Heuristic of Principal Surrogate Approach

Consider an individual with $Y_i(1) = Y_i(0) = 0$

Define this individual to be protected if $Y_i(1) = 0$ & $Y_i(0) = 1$

Probability of Being Protected as a Function of $S_i(1) - S_i(0)$



Assumptions

A1 Stable Unit Treatment Value Assumption (SUTVA):

$(S_i(1), S_i(0), Y_i(1), Y_i(0), Y^{\tau}_i(1), Y^{\tau}_i(0))$ is independent of the treatment assignments Z_j of other subjects

- **A1** implies "consistency": $\mathbf{(S_i(Z_i), Y_i(Z_i), Y^{\tau}_i(Z_i)) = (S_i, Y_i, Y^{\tau}_i)}$

A2 Ignorable assignments:

Z_i is independent of $(S_i(1), S_i(0), Y_i(1), Y_i(0), Y^{\tau}_i(1), Y^{\tau}_i(0))$

- **A2** holds for randomized blinded trials

A3 No causal effects on Y before the marker is measured (Equal Early Clinical Risk – EECR)

- $P(Y^{\tau}_i(1) = Y^{\tau}_i(0))=1$

Definition of a Principal Surrogate

- Frangakis and Rubin (2002) suggested a surrogate endpoint should satisfy

  Causal Necessity:

  S is necessary for the effect of treatment on the outcome Y in the sense that an effect of treatment on Y can occur only if an effect of treatment on S has occurred

  - $S_i(1) = S_i(0) \Rightarrow Y_i(1) = Y_i(0)$

Updated Definition of a Principal Surrogate
(Gilbert and Hudgens, 2008)

- Restrict to the "always at-risk" cohort with $Y^{\tau}_i(1)=Y^{\tau}_i(0)=0$ throughout
  - Because we assume EECR, simply analyze participants with $Y^{\tau} = 0$
- **Define**

$$\text{risk}_{(1)}(s_1, s_0) = \Pr(Y(1) = 1 \mid S(1) = s_1, S(0) = s_0)$$
$$\text{risk}_{(0)}(s_1, s_0) = \Pr(Y(0) = 1 \mid S(1) = s_1, S(0) = s_0)$$

- **A contrast in** $\text{risk}_{(1)}(s_1, s_0)$ and $\text{risk}_{(0)}(s_1, s_0)$ is a causal effect on Y for the population $\{S(1) = s_1, S(0) = s_0\}$
- **A *principal surrogate* is a biomarker measured at τ satisfying 2 conditions, the first of which is:**

  $\text{risk}_{(1)}(s_1, s_0) = \text{risk}_{(0)}(s_1, s_0)$ for all $s_1 = s_0$
- **This property is** *Average Causal Necessity*:
  - $S(1) = S(0) = s \Rightarrow E[Y(1) \mid S(1) = S(0) = s] = E[Y(0) \mid S(1) = S(0) = s]$
  - i.e., "if there is no vaccine-induced immune response, there is no protection"

Updated Definition of a Principal Surrogate
(Gilbert and Hudgens (2008)

- **The second property** is that the clinical treatment effect [measured by a contrast in $\text{risk}_{(1)}(s_1, s_0)$ and $\text{risk}_{(0)}(s_1, s_0)$] varies widely with the values (s_1, s_0)

  i.e., the variables (S_1, S_0) strongly modify vaccine efficacy

- **Thus,** a **principal surrogate** is defined to be a biomarker satisfying average causal necessity and that is a strong effect modifier

- **Note:** This definition allows for a spectrum of principal surrogates, some more useful than others, depending on the extent to which clinical treatment efficacy varies with (S_1, S_0)
  - Stronger effect modification implies a more useful marker

Causal Effect Predictiveness (CEP) Surface

- Let $h(x, y)$ be a known contrast function with $h(x, x) = 0$
  - e.g., $\mathbf{h(x, y) = x - y}$, $\mathbf{\log(x / y)}$, $\mathbf{1 - x / y}$
- **CEP surface:**

$$\mathbf{CEP^{risk}(s_1, s_0) = h(risk_{(1)}(s_1, s_0), risk_{(0)}(s_1, s_0))}$$

  - E.g., $\mathbf{CEP^{risk}(s_1, s_0) = 1 - risk_{(1)}(s_1, s_0) / risk_{(0)}(s_1, s_0)}$ $\mathbf{[= VE(s_1, s_0)]}$

- Henceforth will call the CEP surface simply the "VE surface"

VE Surface in Terms of Marker Percentiles

- Huang, Pepe, and Feng (2007, *Biometrics*) proposed judging the value of a continuous marker S for predicting disease Y by the *predictiveness curve:*

$$R(v) = Pr(Y = 1 \mid S = F^{-1}(v)) \qquad v \in [0,1], \quad S \sim F$$

- With $S(1) \sim F_{(1)}$, define

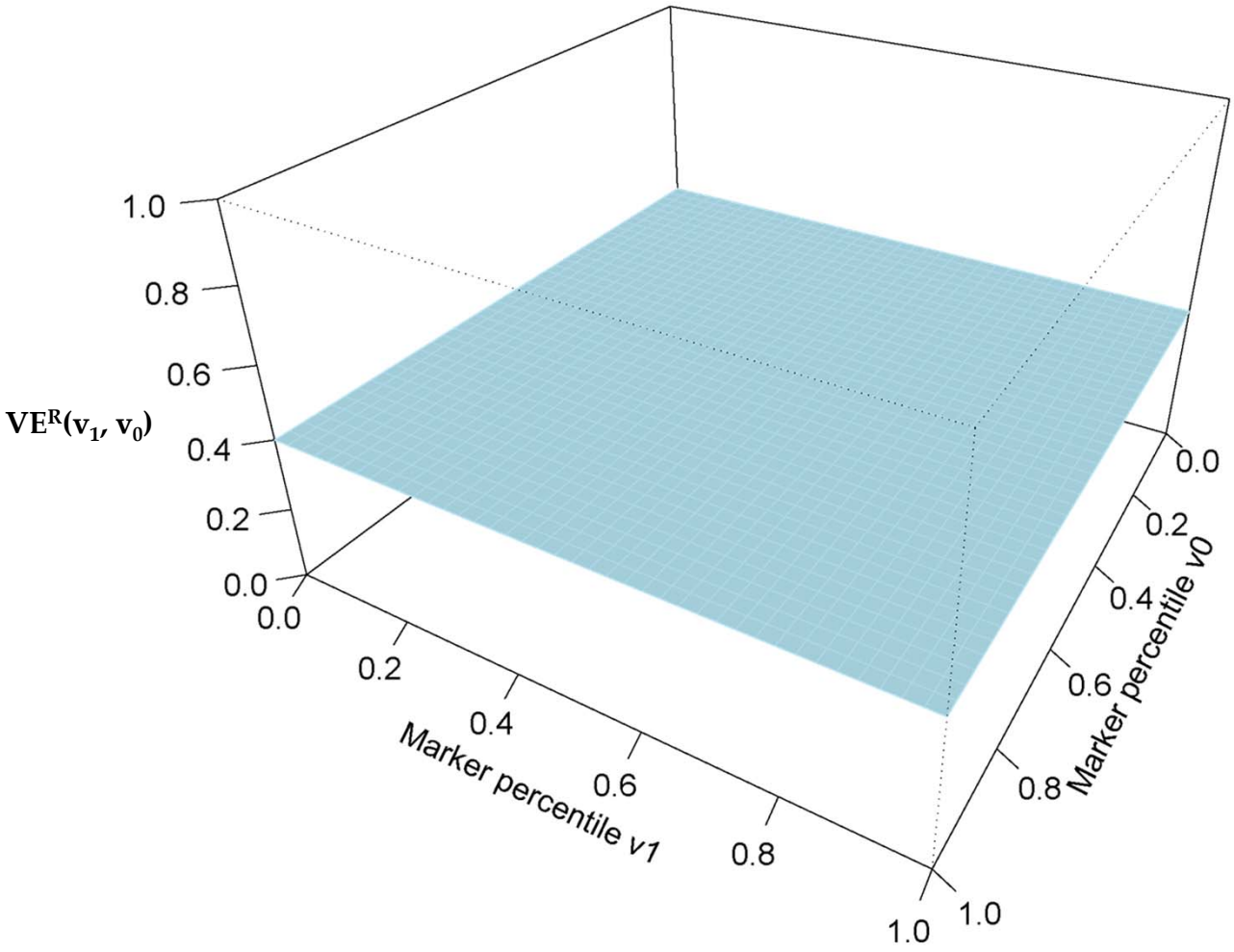
$$R_{(1)}(v_1, v_0) = Pr(Y(1) = 1 \mid S(1) = F^{-1}_{(1)}(v_1), S(0) = F^{-1}_{(1)}(v_0))$$

$$R_{(0)}(v_1, v_0) = Pr(Y(0) = 1 \mid S(1) = F^{-1}_{(1)}(v_1), S(0) = F^{-1}_{(1)}(v_0))$$

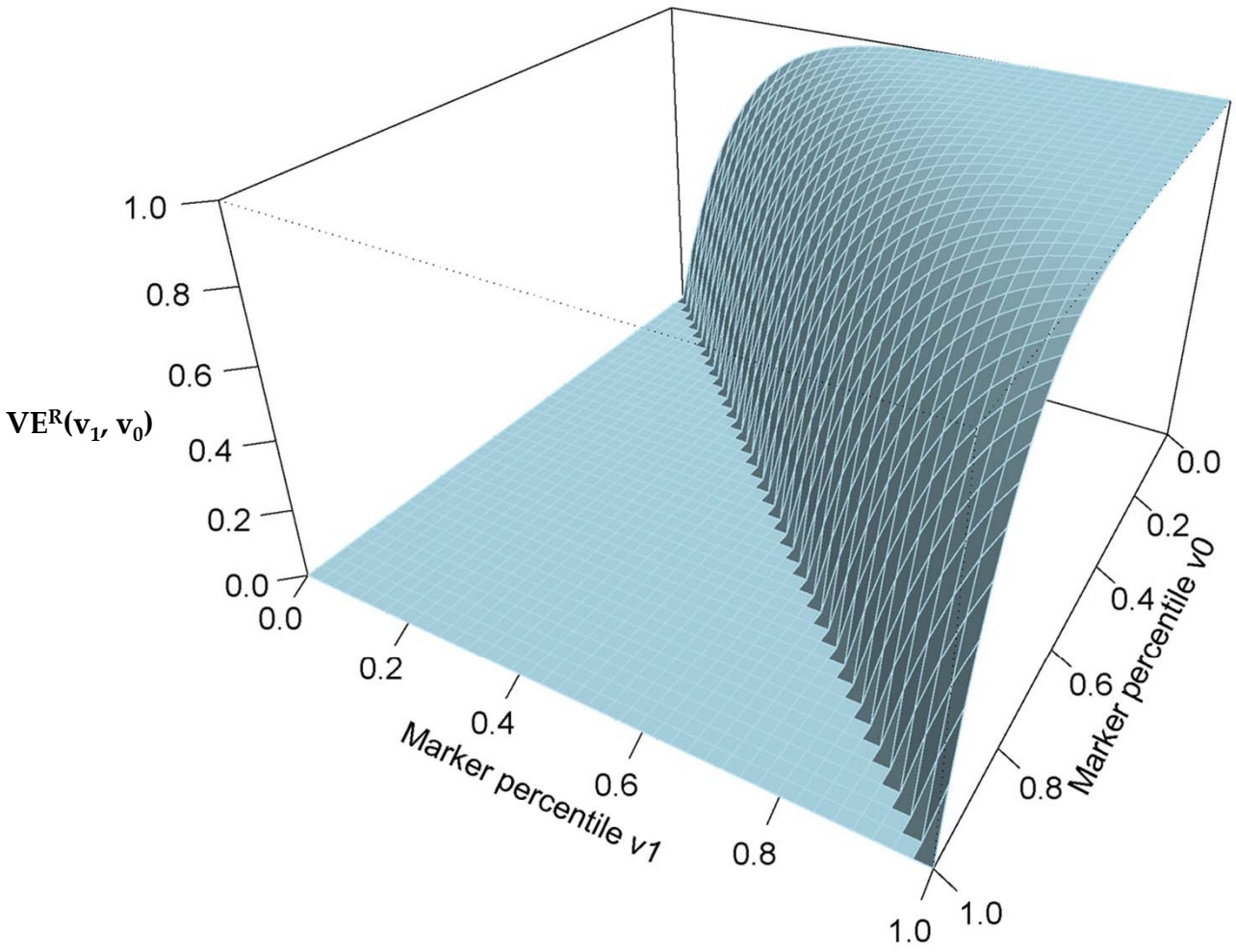
- VE surface:

$$VE^{R}(v_1, v_0) = h(R_{(1)}(v_1, v_0), R_{(0)}(v_1, v_0))$$

$VE^R(v_1, v_0)$ Surface: Biomarker with No Surrogate Value



$VE^R(v_1, v_0)$ Surface: Biomarker with High Surrogate Value



VE Surface in Case Constant Biomarker (CB)

$S_i(0) = c$ for all i

- Case CB typically occurs in vaccine trials where enrolled subjects are naïve to the pathogen under study
- In this case the VE surface is a curve

  $VE^{risk}(s_1, c)$ or $VE^{R}(v_1, F_{(1)}(c))$

- A principal surrogate is a biomarker with

  $VE^{risk}(c, c) = 0$ and

  $VE^{risk}(s_1, c) > 0$ varies markedly with s_1

Marginal VE Curve for the General Case

- Define

$$\text{risk}_{(1)}(s_1) = \Pr(Y(1) = 1 \mid S(1) = s_1)$$

$$\text{risk}_{(0)}(s_1) = \Pr(Y(0) = 1 \mid S(1) = s_1)$$

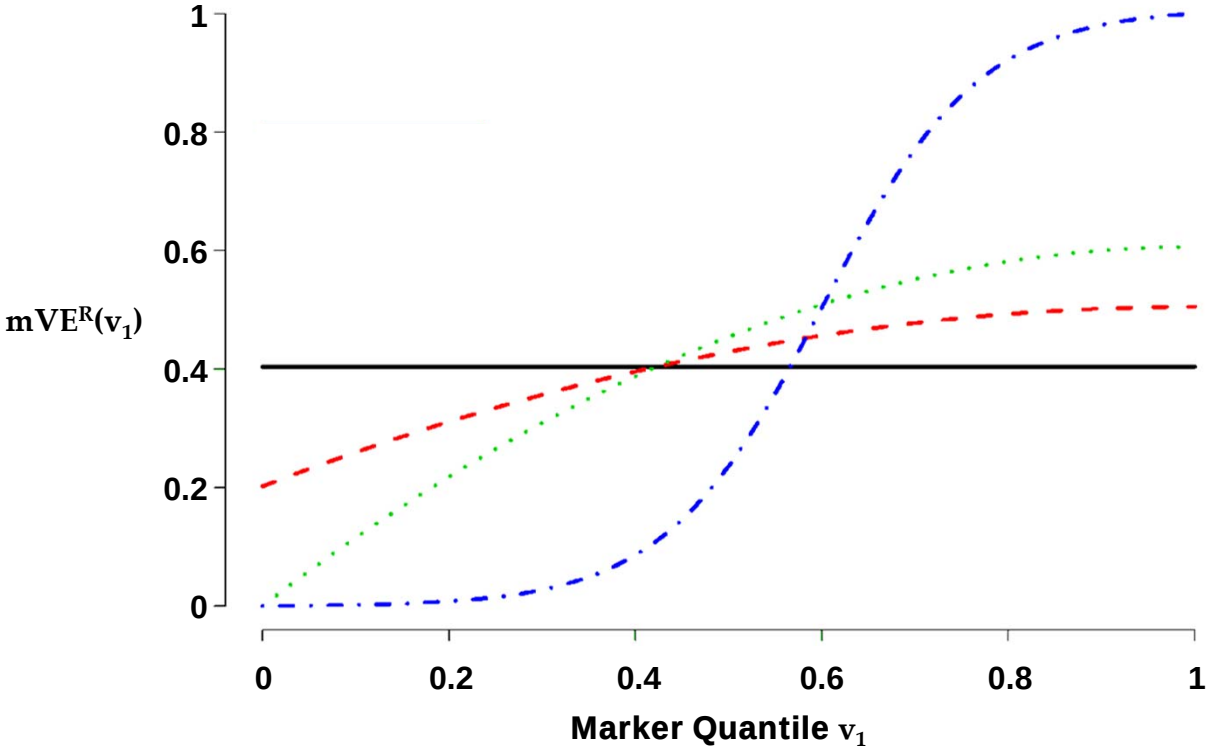
- Marginal VE curve:

$$mVE^R(s_1) = h(\text{risk}_{(1)}(s_1), \text{risk}_{(0)}(s_1))$$

- Marginal VE curve with percentile formulation:

$$mVE^R(v_1) = h(R_{(1)}(v_1), R_{(0)}(v_1))$$

Illustration of Marginal VE Curves



Principal Surrogate Value

- **A biomarker with some surrogate value should have**

$$s_1 \text{ near } s_0 \Rightarrow VE^{risk}(s_1, s_0) \text{ near } 0$$

  - There are some $s_1 \neq s_0$ for which $VE^{risk}(s_1, s_0)$ is far from $risk_{(0)}(s_1, s_0)$

- **Strong Average Causal Sufficiency (Strong ACS):**
  S is sufficient for the effect of treatment on the outcome Y in the sense that an effect of treatment on S implies an effect of treatment on Y
  - $s_1 \neq s_0 \Rightarrow VE^{risk}(s_1, s_0) \neq 0$
  - i.e., "A vaccine effect on the marker implies there is some protection"
  - 1-sided version: $s_1 > s_0 \Rightarrow VE^{risk}(s_1, s_0) > 0$

Connection of ACN and ACS to Prentice (1989) Concept of Specificity & Sensitivity

- Prentice's definition of a valid surrogate re-cast in terms of Specificity and Sensitivity for 1-sided alternatives of interest:
  1. **1-Sided Specificity:** $VE = 0\%$ implies $S(1) =^{d} S(0)$

     - i.e., $S(1) >^{st} S(0)$* implies $VE > 0\%$

  2. **1-Sided Sensitivity:** $VE > 0\%$ implies $S(1) >^{st} S(0)$

     - i.e., $S(1) =^{d} S(0)$ implies $VE = 0\%$
- The Prentice definition equates to 1. and 2. both holding

*$S(1) >^{st} S(0)$ defined as $P(S(1)>s) \geq P(S(0)>s)$ with '>"for some s

Connection of ACN and ACS to Prentice Concept of Specificity & Sensitivity*

- Under Case CB:
  - EECR + ACN ⇒ 1-Sided Sensitivity
  - EECR + ACN + 1-sided Strong ACS ⇒ 1-Sided Specificity

- In General Case:
  - EECR + ACN Does Not ⇒ 1-Sided Sensitivity even under the 2 extra conditions below
  - EECR + ACN + 1-sided Strong ACS ⇒ 1-Sided Sensitivity under either of 2 extra conditions

Cond 1: $P(S(1) \geq S(0))=1$; Cond 2: No harm for any subgroup, $CEP(\mathbf{s_1}, \mathbf{s_0}) \geq \mathbf{0}$

- Special case of binary S, EECR, Case CB: ACN & 1-sided Strong ACS if and only if 1-Sided Specificity & 1-Sided Sensitivity
  - 1:1 correspondence of 2 principal surrogate criteria and the Prentice definition

*Results from Gilbert, Gabriel, Huang, Chan (2015, *J Causal Inference*)

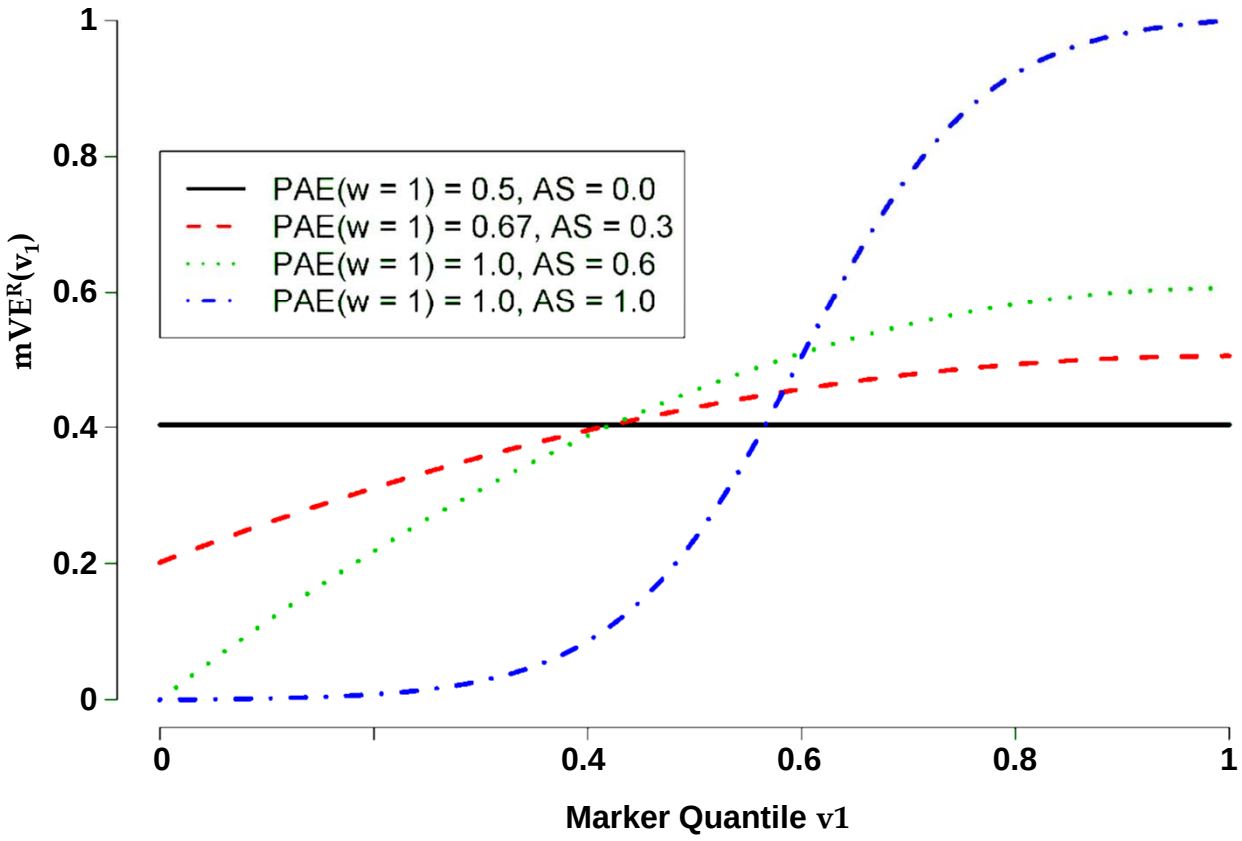
Summary Measures of Surrogate Value

- Focus on the 1-sided setting where interest is in assessing if greater vaccine-induced immune responses predict beneficial VE > 0
- Following Frangakis and Rubin (2002), consider 'dissociative' and 'associative' effects
  - Dissociative effect = no treatment effect on marker but a treatment effect on the clinical endpoint
  - Associative effect = treatment effect on the marker and on the clinical endpoint
  - If a marker is valuable as a surrogate, then few subjects will have dissociative effects and many will have associative effects

Summary Measures of Surrogate Value

- Define the expected dissociative effect (EDE) and the expected associative effect (EAE)
  - $EDE = E[CEP^{risk}(S(1),S(0)) | S(1) = S(0)]$
  - $EAE(w) = E[w(S(1),S(0))CEP^{risk}(S(1),S(0)) | S(1) > S(0)]$
- Based on these, define summary measures of surrogate value (proportion associative effect and associative span)
  - $PAE(w) = |EAE(w)| / [|EDE| + |EAE(w)|]$
  - $AS = |EAE(w)| - |EDE|$
    - $PAE(w) > 0.5$; $AS > 0$ suggests some surrogate value

Summary Measures of Surrogate Value



Black: no surrogate value

Blue: Perfect surrogate value

Green and Red: Partial Surrogate value

Green and Blue: Satisfy Average Causal Necessity

Challenge to Evaluating a Principal Surrogate: Missing Data

- The VE surface is not identified from data collected in a randomized trial with standard design
  - Only one of $(S_i(1), Y_i(1), Y^{\tau}_i(1))$ or $(S_i(0), Y_i(0), Y^{\tau}_i(0))$ is observed from each subject
- Accurate prediction/modeling of the missing potential outcomes is required to estimate the VE surface (and the marginal VE curve)

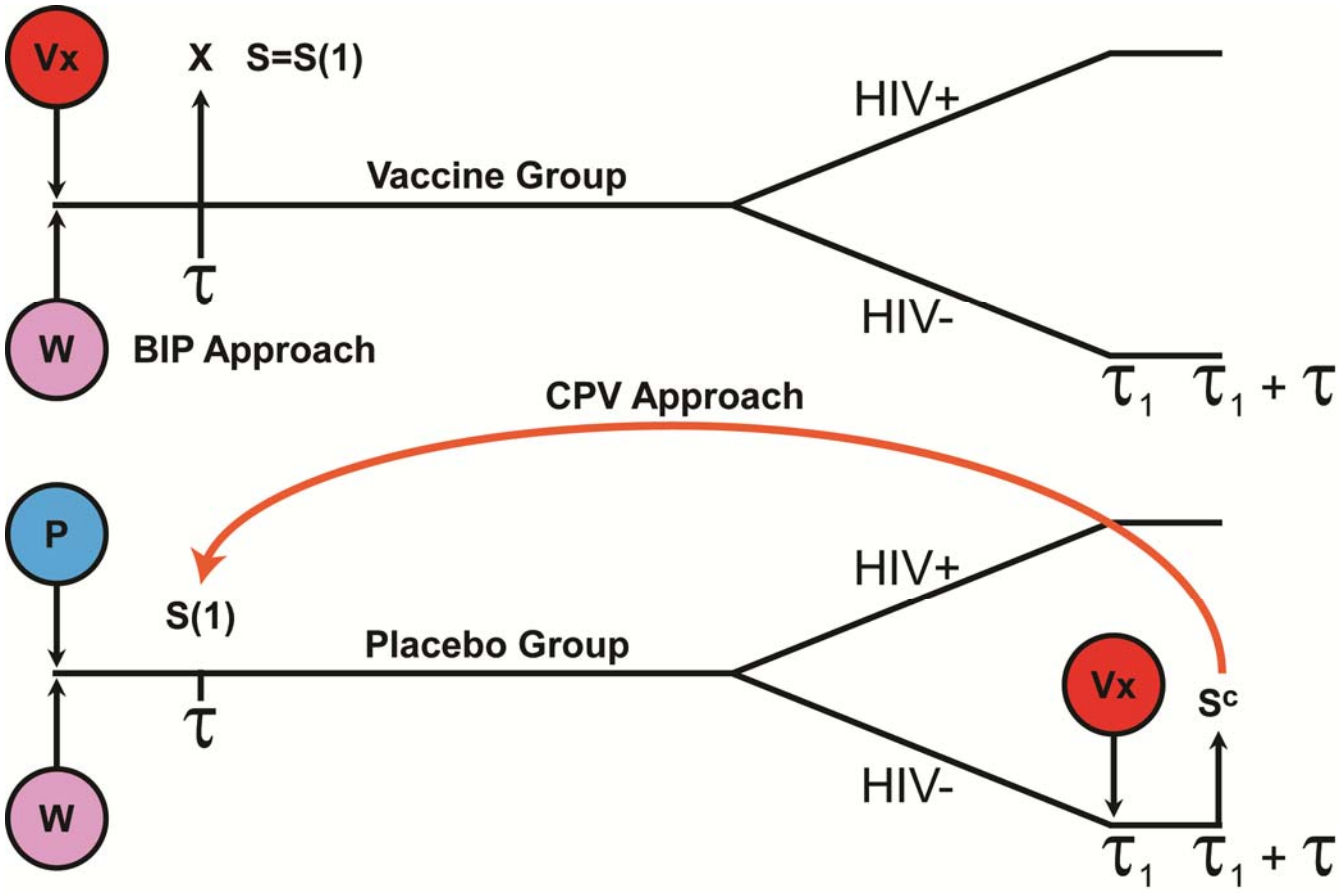
Outline of Session 6

1. Effect Modification/VE Curve Framework
2. **Identifiability and Estimation**
3. Simulations
4. Discussion

Dean Follmann's Augmented Vaccine Efficacy Trial Designs

- Follmann (2006, *Biometrics*) proposed augmented vaccine trial designs for aiding inference on the VE curve
- Two strategies for predicting S(1) for placebo recipients
  - Baseline Immunogenicity Predictor (BIP)
  - Closeout Placebo Vaccination (CPV)
- Follmann developed estimation approaches for augmented designs with BIP, CPV, or both
- Gilbert and Hudgens (2008) considered the BIP approach only

Schematic of Baseline Predictor and Closeout Placebo Vaccination Trial Designs*

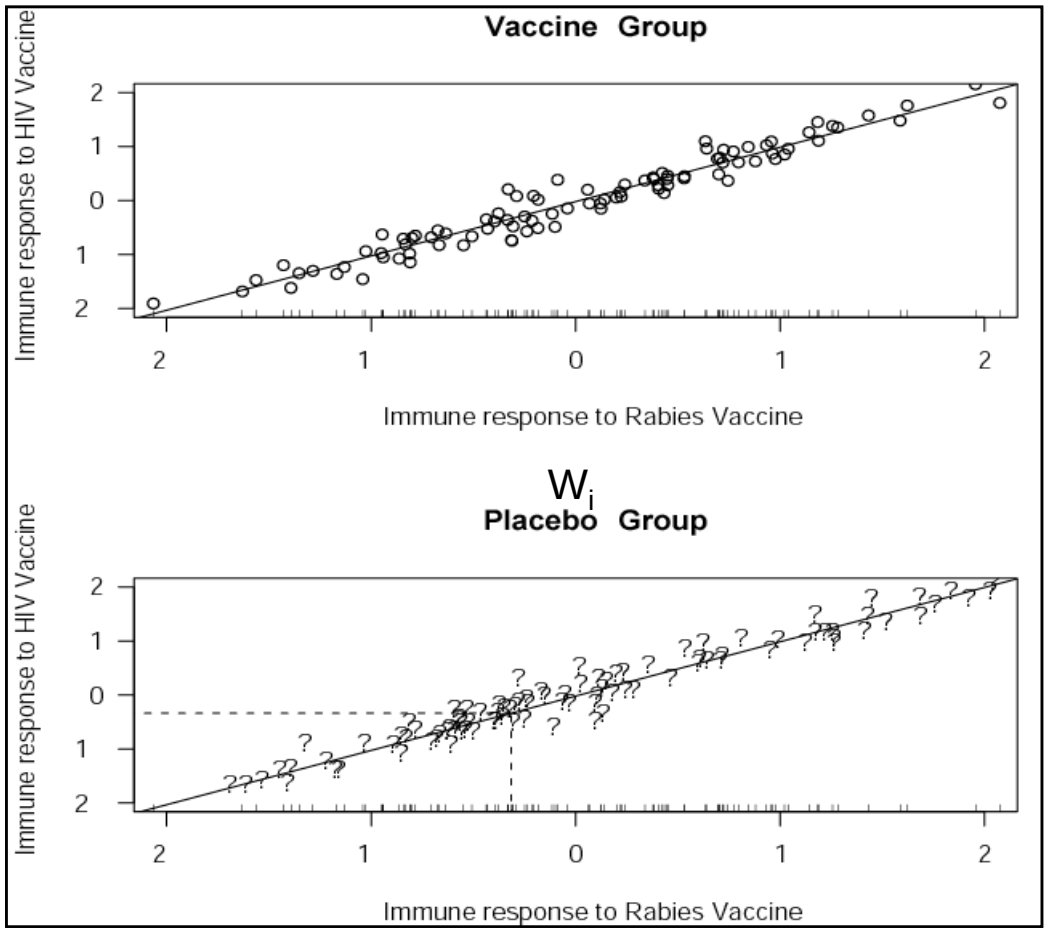


*Proposed by Follmann (2006, Biometrics)

Closeout Placebo Vaccination

- At the end of the trial, inoculate a random sample of uninfected placebo recipients with HIV vaccine
- Measure the immune response on the same schedule as it was measured for vaccine recipients
- Assume the measurement is what we would have seen, had we inoculated during the trial

Baseline Immunogenicity Predictor (BIP) Approach*



$S_i(1)$

W_i

$S_i(1)$

W_i

Evaluate correlation of $\mathbf{W}$ and $\mathbf{S(1)}$ in vaccine group with $\Upsilon^{\tau} = 0$

Predict $\mathbf{S(1)}$ from vaccine group model and $\mathbf{W}$ in placebos with $\Upsilon^{\tau} = 0$

*Figure from Follmann (2006, Biometrics)

Baseline Immunogenicity Predictor (BIP) Approach

- Needed condition for the BIP approach:

$$W, S(1) \mid Z = 1, Y^{\tau} = 0 \;\; =^{d} \;\; W, S(1) \mid Z = 0, Y^{\tau} = 0 \qquad (*)$$

- This holds by A2, A2 (randomization), and A3 (EECR)
  - A3 needed for $Y^{\tau} = 0$ to equate to $Y^{\tau}(1) = Y^{\tau}(0) = 0$
- Without A3, (*) may not hold, in which case it is not valid to use a regression model to fill in the S(1)'s of placebo recipients based on their W's
- (Recall that all conditional distributions implicitly condition on $Y^{\tau} = 0$)

Build on 2-Phase Sampling / Nested Case-Control Methods

- **2-phase sampling / nested case-control**
  - $(W, S(1))$ measured in
    - All infected vaccine recipients
    - Sample of uninfected vaccine recipients
  - W measured in
    - All infected placebo recipients
    - Sample of uninfected placebo recipients
- **2-Phase designs** (E.g., Kulich and and Lin, 2004, *JASA;* Breslow et al., 2009, *AJE, Stat Biosciences*)
  - Phase 1: Measure inexpensive covariates in all subjects
  - Phase 2: Measure expensive covariates X in a sample of subjects
- **Our application**
  - Vaccine Group: Exactly like 2-phase design with $X = (W, S(1))$
  - Placebo Group: Like 2-phase design with $X = (W, S(1))$ and $S(1)$ missing

Inverse Probability Weighted (IPW) 2-Phase Methods Do Not Apply: Hence we use a Full Likelihood-Based Method

- None of the case-cohort/2-phase methods described in Session 5 apply to this problem
  - The reason: they are all IPW-based methods, using score equations that sum over subjects with phase-2 data only, which assume that every subject has a positive probability that S(1) is observed
- However all placebo subjects have zero-probability that S(1) is observed
- To deal with this problem, we use full likelihood methods, for which the score equations sum over all subjects

Maximum Estimated Likelihood with BIP

(Build on Pepe and Fleming, 1991, *JASA*)

- Posit models for $\text{risk}_{(1)}(s_1,0;\beta)$ and $\text{risk}_{(0)}(s_1,0;\beta)$
- Vaccine arm:
  - $(W_i, S_i(1))$ measured: Likld contribn $\mathbf{risk_{(1)}(S_i(1), 0; \beta)}$
  - $(W_i, S_i(1))$ not measured: $\mathbf{\int risk_{(1)}(s_1, 0; \beta)\, dF(s_1)}$
- Placebo arm:
  - W_i measured: Likld contribn $\mathbf{\int risk_{(0)}(s_1, 0; \beta)\, dF^{S|W}(s_1 | W_i)}$
  - W_i not measured: $\mathbf{\int risk_{(0)}(s_1, 0; \beta)\, dF(s_1)}$
- $L(\beta, F^{S|W}, F) = \prod_i \{[\text{risk}_{(1)}(S_i(1),0;\beta)^{Yi}\,(1 - \text{risk}_{(1)}(S_i(1),0;\beta))^{1-Yi}]^{Zi}\}^{\delta i}$ [Vx subcohort]

$\times \{[\int \text{risk}_{(0)}(s_i,0;\beta) dF^{S|W}(s_1|W_i)^{Yi}\,(1 - \int \text{risk}_{(0)}(s_1,0;\beta))\, dF^{S|W}(s_1|W_i)^{1-Yi}]^{1-Zi}\}^{\delta i}$ [Plc subcohort]

$\times \{[\int \text{risk}_{(1)}(s_i,0;\beta) dF(s_1)^{Yi}\,(1 - \int \text{risk}_{(1)}(s_1,0;\beta))\, dF(s_1)^{1-Yi}]^{Zi}\}^{1-\delta i}$ [Vx not subcohort]

$\times \{[\int \text{risk}_{(0)}(s_i,0;\beta) dF(s_1)^{Yi}\,(1 - \int \text{risk}_{(0)}(s_1,0;\beta))\, dF(s_1)^{1-Yi}]^{1-Zi}\}^{1-\delta i}$ [Plc not subcohort]

Maximum Estimated Likelihood Estimation (MELE)

- Likelihood $L(\beta, F^{S|W}, F)$
  - β is parameter of interest [CEP surface and marginal CEP curve depend only on β]
  - $F^{S|W}$ and F are nuisance parameters

Step 1: Choose models for $F^{S|W}$ and F and estimate them based on vaccine arm data

Step 2: Plug the consistent estimates of $F^{S|W}$ and F into the likelihood, and maximize it in β

- e.g., EM algorithm

Step 3: Estimate the variance of the MELE of β, accounting for the uncertainty in the estimates of $F^{S|W}$ and F

- Bootstrap

Modeling Approach 1
(Fully Parametric)

- **Assume:**
  - $F^{S|W}$ has a specified parametric distribution
  - $S(1)$ is continuous subject to "limit of detection" left-censoring:
  - $S(1) = \max(S^*(1), 0)$, where $S^*(1)$ has a continuous cdf
  - **A4-P:** Structural models for $\text{risk}_{(z)}$ (for $z=0, 1$)
  - $\text{risk}_{(z)}(s_1, 0, w; \beta_z) = g(\beta_{z0} + \beta_{z1} s_1 + \beta^T_{z2} w)$, g a known link
- **Example:**

  $F^{W|X}$ normal, $F^{S|W}$ censored normal with left-censoring below 0, A4-P holds with $g = \Phi$, the standard normal cdf
- **No interactions assumption (untestable):** One of the components of β^T_{12} equals the corresponding component of β^T_{02} (untestable)

Modeling Approach 1
(Fully Parametric)

- **Interpretation:**
  - With $h(x, y) = g^{-1}(x) - g^{-1}(y)$

$$VE^{risk}(s_1, 0, w) = (\beta_{10} - \beta_{00}) + (\beta_{11} - \beta_{01})s_1 + (\beta_{12} - \beta_{02})^T w$$

- Under assumption of no interactions between Z and W:

$$VE^{risk}(s_1, 0) = (\beta_{10} - \beta_{00}) + (\beta_{11} - \beta_{01})s_1$$

= W-adjusted VE-curve

Parametric Approach:
Interpretation of Parameters

$VE^{risk}(s_1, 0) = (\beta_{01} - \beta_{00}) + (\beta_{11} - \beta_{10})s_1$

- **S** satisfies average causal necessity $\Leftrightarrow$ $\beta_{01} = \beta_{00}$
- $\beta_{11} = \beta_{10}$ indicates a positive treatment effect on **S** does not predict a beneficial clinical effect
- $\beta_{11} < \beta_{10}$ indicates it does predict a beneficial clinical effect (i.e., some effect modification)

- A 'good' surrogate has $|\beta_{01} - \beta_{00}|$ near 0 and $|\beta_{11} - \beta_{10}|$ large

Modeling Approach 2
(Fully Nonparametric)

- **Assume:**
  - S and W categorical with J and K levels; $S_i(0)=1$ for all i
  - Nonparametric models for $P(S(1)=j, W=k)$
  - **A4-NP:** Structural models for $risk_{(z)}$ (for $z=0, 1$)

$$risk_{(z)}(j, 1, k; \beta) = \beta_{zj} + \beta'_k \text{ for } j=1, \ldots, J; k=1, \ldots, K$$

Constraint: $0 \le \beta_{zj} + \beta'_k \le 1$ and $\Sigma_k \beta'_k = 0$ for identifiability

- **No interactions assumption:** W has the same association with risk for the 2 study groups (untestable)

Modeling Approach 2
(Fully Nonparametric)

- **Interpretation:**
  - With $h(x, y) = \log (x / y)$

    $CEP^{risk}(j, 1) = \log (\text{avg-risk}_{(1)}(j, 1) / \text{avg-risk}_{(0)}(j, 1))$

    where $\text{avg-risk}_{(z)}(j, 1) = (1/K) \Sigma_k \text{risk}_{(z)}(j, 1, k; \beta)$ for $z=0, 1$

    $VE(j, 1) = 1 - \exp\{CEP^{risk}(j, 1)\}$

Interpretation
(Fully Nonparametric)

- With $VE(j, 1) = 1 - \text{avg-risk}_{(1)}(j, 1) / \text{avg-risk}_{(0)}(j, 1)$:
  - **S** satisfies ACN and 1-sided Strong ACS if

    $VE(1, 1) = 0$ and **$VE(j, 1) > 0$** for all $j > 1$

- A biomarker with *some* value as a surrogate will have
  - **$VE(1, 1)$** near **0**
  - **$VE(j, 1)$** > 0 for some $j > 1$
- The most useful marker will also have $VE(j, 1)$ large for some $j > 1$ [strong effect modification]

Modeling Approach 2
(Fully Nonparametric)

- Wald tests for whether a biomarker has any surrogate value
  - Under the null, $\mathrm{PAE}(w) = 0.5$ and $\mathrm{AS} = 0$
  - $Z = (\text{Est. } \mathrm{PAE}(w) - 0.5)/\ \text{s.e.}(\text{Est. } \mathrm{PAE}(w))$
  - $Z = \text{Est. AS}/\ \text{s.e.}(\text{Est. AS})$
    - Estimates obtained by MELE; bootstrap standard errors
- For nonparametric case A4-NP, test $H0: VE^{risk}(j, 1) = 0$ vs $H1: VE^{risk}(j, 1)$ increases in j (like the Breslow-Day trend test)
  - $T = \Sigma_{j>1}(j-1)\ \{\text{Est. } \beta_{0j} - (\text{Est. } \beta_{0j} + \text{Est. } \beta_{1j})(\text{Est. } \mu_{z0}/(\text{Est. } \mu_{z0} + \text{Est. } \mu_{z1}))\}$ divided by bootstrap s.e.

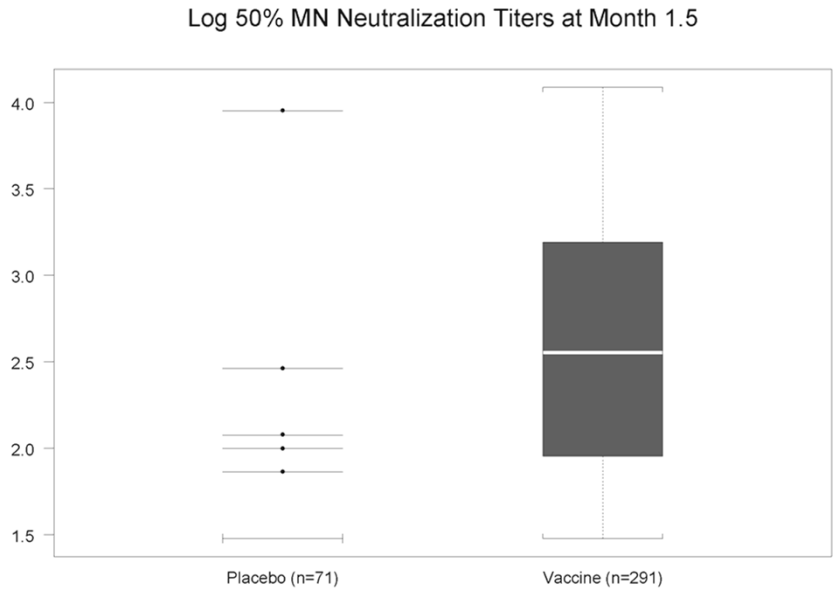
    $\text{Est. } \mu_z = (1/J)\ \Sigma_j\ \beta_{zj}$

Outline of Session 6

1. Effect Modification/VE Curve Framework
2. Identifiability and Estimation
3. **Simulations**
4. Discussion

Simulation Plan
(Based on Vax004 HIV Vaccine Efficacy Trial)

- **Biomarker of interest:** S = 50% neutralization titer against the recombinant gp120 molecule (MN strain) measured at the month 1.5 visit



- 66 of 71 placebo recipients had S left-censored below the LLOQ = 1.65
- Range of S is $[1.65, 4.09]$; rescale to $[0, 1]$ so that $S_i(0) = 0$ [Case CB holds]

Simulation Plan

- **Step 1:** For all $N=5403$ subjects, generate $(W_i, S_i(1))$ from a bivariate normal with means (0.41, 0.41), sds (0.55, 0.55), correlation
  $\rho = 0.5, 0.7, \text{or } 0.9$
  - sd of 0.55 chosen to achieve the observed 23% rate of left-censoring
  - Values of $W_i, S_i(1) < 0$ set to 0; values > 1 set to 1

- **Step 2:** Bin W_i and $S_i(1)$ into quartiles
  - Under model A4-NP generate $Y_i(Z)$ from a Bernoulli$(\beta_{zj} + \beta'_k)$ with the parameters set to achieve:
    - $P(Y(1) = 1) = 0.0$ and $P(Y(0) = 1) = 0.134$ (overall VE = 50%)
    - The biomarker has either (i) no or (ii) high surrogate value

Simulation Plan

- Recall $CEP^{risk}(j, 1) = \log (\text{avg-risk}_{(1)}(j, 1) / \text{avg-risk}_{(0)}(j, 1))$
- Scenario (i) (no surrogate value)
  - $CEP^{risk}(j, 1) = -0.69$ for $j = 1, 2, 3, 4$
  - i.e., $VE(j, 1) = 0.50$ for $j = 1, 2, 3, 4$
- Scenario (ii) (high surrogate value)
  - $CEP^{risk}(j, 1) = -0.22, -0.51, -0.92, -1.61$ for $j = 1, 2, 3, 4$
  - i.e., $VE(j, 1) = 0.2, 0.4, 0.6, 0.8$ for $j = 1, 2, 3, 4$

Simulation Plan

- Step 3: Create nested case-control sampling (3:1 control: case)
  - **Vaccine group:** $(W, S(1))$ measured in all infected ($n=241$) and a random sample of 3 x 241 uninfected
  - **Placebo group:** W measured in all infected ($n=127$) and a random sample of 3 x 127 uninfected
- The data were simulated to match the real Vax004 trial as closely as possible

Questions Evaluated by the Simulations

- Bias of the MELEs of
  - β_{zj}
  - $CEP^{risk}(j, 1)$ [equivalent to $VE(j, 1)$]
  - AS
  - $PAE(w)$ for $w(j) = 1,\ j,\ I(j=4)$
- Coverage probabilities of bootstrap percentile CIs for the above parameters
- Power of Wald tests and of the test-for-trend

Model A4-NP Simulation Results*

Table 1

Model A4-NP simulation results for the nonparametric MELEs $\widehat{CEP}^{risk}(j,1;\beta) = \log(\widehat{\beta}_{1j}/\widehat{\beta}_{0j})$ *for* $j = 1,\cdots,4$[a]

Cor.		No Surrogate Value Scenario						High Surrogate Value Scenario				
ρ	Parameter	Bias	SE	SEE	CP	Power	Parameter	Bias	SE	SEE	CP	Power
0.5	$CEP^{risk}(1,1) = -0.69$	-0.04	0.42	0.41	0.98	0.45	$CEP^{risk}(1,1) = -0.22$	-0.06	0.67	0.65	0.98	0.12
	$CEP^{risk}(2,1) = -0.69$	0.11	0.91	0.90	0.99	0.09	$CEP^{risk}(2,1) = -0.51$	0.09	0.96	0.93	1.00	0.04
	$CEP^{risk}(3,1) = -0.69$	0.13	0.88	0.87	0.99	0.06	$CEP^{risk}(3,1) = -0.92$	0.15	0.94	0.93	1.00	0.09
	$CEP^{risk}(4,1) = -0.69$	0.09	0.80	0.72	0.98	0.18	$CEP^{risk}(4,1) = -1.61$	-0.03	0.65	0.66	0.98	0.66
0.7	$CEP^{risk}(1,1) = -0.69$	-0.03	0.30	0.29	0.96	0.62	$CEP^{risk}(1,1) = -0.22$	-0.03	0.45	0.47	0.97	0.13
	$CEP^{risk}(2,1) = -0.69$	0.09	0.80	0.77	0.99	0.17	$CEP^{risk}(2,1) = -0.51$	0.06	0.87	0.84	0.99	0.08
	$CEP^{risk}(3,1) = -0.69$	-0.02	0.82	0.79	1.00	0.11	$CEP^{risk}(3,1) = -0.92$	-0.02	0.83	0.83	0.99	0.17
	$CEP^{risk}(4,1) = -0.69$	0.06	0.73	0.64	0.97	0.22	$CEP^{risk}(4,1) = -1.61$	0.00	0.47	0.48	0.96	0.82
0.9	$CEP^{risk}(1,1) = -0.69$	0.00	0.19	0.19	0.95	0.90	$CEP^{risk}(1,1) = -0.22$	-0.01	0.28	0.27	0.94	0.18
	$CEP^{risk}(2,1) = -0.69$	0.02	0.48	0.48	0.96	0.37	$CEP^{risk}(2,1) = -0.51$	0.01	0.66	0.59	0.95	0.26
	$CEP^{risk}(3,1) = -0.69$	-0.02	0.68	0.63	0.96	0.27	$CEP^{risk}(3,1) = -0.92$	0.00	0.62	0.58	0.95	0.40
	$CEP^{risk}(4,1) = -0.69$	-0.01	0.53	0.50	0.96	0.32	$CEP^{risk}(4,1) = -1.61$	-0.03	0.39	0.36	0.95	0.99

[a] ρ is the linear correlation of the simulated bivariate normal variables latent to the quartilized variables W and $S(1)$. Bias is the median bias. SE is the empirical standard error of $\widehat{CEP}^{risk}(j,1)$. SEE is the median of the bootstrap standard error estimates based on 500 bootstrap replicates. CP is the empirical coverage of bootstrap percentile 95% confidence intervals for $\widehat{CEP}^{risk}(j,1)$. Power refers to power of the Wald test to reject $H_0 : CEP^{risk}(j,1) = 0$. 1000 simulations were done to compute the table elements for each model.

*In Gilbert and Hudgens (2008, *Biometrics)*

Model A4-NP Simulation Results*

Table 2

Model A4-NP simulation results for the nonparametric MELEs $\widehat{PAE}^{\omega}$ and $\widehat{AS}$, with $h(x,y) = \log(x/y)$[a]

Cor.		No Surrogate Value Scenario						High Surrogate Value Scenario				
ρ	Parameter	Bias	SE	SEE	CP	Power	Parameter	Bias	SE	SEE	CP	Power
0.5	$PAE^{\omega_1} = 0.50$	-0.13	0.22	0.21	0.95	0.03	$PAE^{\omega_1} = 0.82$	-0.21	0.23	0.23	0.98	0.15
	$PAE^{\omega_2} = 0.50$	-0.12	0.21	0.20	0.96	0.02	$PAE^{\omega_2} = 0.84$	-0.18	0.19	0.20	0.97	0.21
	$PAE^{\omega_3} = 0.50$	0.03	0.21	0.20	0.99	0.04	$PAE^{\omega_3} = 0.88$	-0.11	0.17	0.19	0.99	0.51
	$AS = 0.00$	0.07	0.53	0.55	0.99	0.04	$AS = 1.39$	-0.22	0.70	0.71	0.98	0.51
0.7	$PAE^{\omega_1} = 0.50$	-0.09	0.19	0.19	0.94	0.02	$PAE^{\omega_1} = 0.82$	-0.12	0.18	0.20	0.97	0.27
	$PAE^{\omega_2} = 0.50$	-0.08	0.17	0.17	0.94	0.02	$PAE^{\omega_2} = 0.84$	-0.10	0.15	0.17	0.97	0.39
	$PAE^{\omega_3} = 0.50$	0.02	0.20	0.19	0.99	0.04	$PAE^{\omega_3} = 0.88$	-0.06	0.12	0.14	0.98	0.75
	$AS = 0.00$	0.04	0.50	0.49	0.99	0.05	$AS = 1.39$	-0.14	0.51	0.55	0.96	0.70
0.9	$PAE^{\omega_1} = 0.50$	-0.03	0.13	0.14	0.96	0.02	$PAE^{\omega_1} = 0.82$	-0.04	0.14	0.15	0.96	0.56
	$PAE^{\omega_2} = 0.50$	-0.02	0.13	0.14	0.96	0.02	$PAE^{\omega_2} = 0.84$	-0.04	0.11	0.12	0.96	0.75
	$PAE^{\omega_3} = 0.50$	0.01	0.19	0.17	0.98	0.08	$PAE^{\omega_3} = 0.88$	-0.02	0.09	0.10	0.97	0.94
	$AS = 0.00$	0.02	0.50	0.46	0.98	0.08	$AS = 1.39$	-0.03	0.45	0.43	0.96	0.94

[a] ρ is the linear correlation of the simulated bivariate normal variables latent to the quartilized variables W and $S(1)$. Bias is the median bias. SE is the empirical standard error of $\widehat{PAE}^{\omega}$ and $\widehat{AS}$. SEE is the median of the bootstrap standard error estimates based on 500 bootstrap replicates. CP is the empirical coverage of bootstrap percentile 95% confidence intervals for PAE^{ω} and AS. Power is for 1-sided tests of $H_0 : PAE^{\omega} = 0.5$ versus $H_1 : PAE^{\omega} > 0.5$ or $H_0 : AS = 0$ versus $H_1 : AS > 0$ at level $\alpha = 0.05$. For the PAE weights, $\omega_1(j,1) = 1$, $\omega_2(j,1) = j$, and $\omega_3(j,1) = I[j = J = 4]$. 1000 simulations were done to compute the table elements for each model.

Trend tests: Power 0.83, 0.99, > 0.99 for ρ = 0.5, 0.7, 0.9

*In Gilbert and Hudgens (2008, *Biometrics*)

Additional Simulation Study

- Evaluate the performance of the MELE method with binned covariates when the data were generated from the continuous model A4-P:
  - $\text{risk}_{(z)}(s_1, 0, w; \beta_z) = \Phi(\beta_{z0} + \beta_{z1} s_1 + \beta_{z3} w)$
    - **Vaccine group:** Set $(\beta_{10}, \beta_{11}, \beta_{13}) = (-1.21, -0.67, -0.1)$ [based on a probit regression fit to the Vax004 data]
    - **Placebo group:** Set $(\beta_{00}, \beta_{01}, \beta_{03})$ such that VE = 50%, $\beta_{03} = \beta_{13}$ and either
      (i) $\beta_{01} = \beta_{11}$ (**no surrogate value**)
      (ii) $\beta_{01} = 0$ (**high surrogate value**)
    - With $h(x, y) = \Phi^{-1}(x) - \Phi^{-1}(y)$:

      (i): $\text{CEP}^{\text{risk}}(s_1, 0) = \beta_{10} - \beta_{00} = -0.11$ $[\text{AS} = 0; \text{PAE}(w) = 0.5]$

      (ii): $\text{CEP}^{\text{risk}}(s_1, 0) = \beta_{10} - \beta_{00} + (\beta_{11} - \beta_{01})s_1 = -0.11 - 0.67\, s_1$
      $[\text{AS} = 0.67, \text{PAE}(w) = 0.82\text{-}0.88]$

Results: Additional Simulation Study*

Table 3

Model A4-P (probit) model simulation results for the nonparametric MELEs $\widehat{PAE}^{\omega}$ and $\widehat{AS}$, with $h(x,y) = \Phi^{-1}(x) - \Phi^{-1}(y)$[a]

Cor.		No Surrogate Value Scenario						High Surrogate Value Scenario				
ρ	Parameter	Bias	SE	SEE	CP	Power	Parameter	Bias	SE	SEE	CP	Power
0.5	$PAE^{\omega_1} = 0.50$	-0.20	0.25	0.23	0.94	0.03	$PAE^{\omega_1} = 0.82$	-0.25	0.24	0.23	0.96	0.12
	$PAE^{\omega_2} = 0.50$	-0.19	0.23	0.22	0.94	0.03	$PAE^{\omega_2} = 0.85$	-0.24	0.22	0.22	0.94	0.15
	$PAE^{\omega_3} = 0.50$	0.01	0.21	0.21	1.00	0.05	$PAE^{\omega_3} = 0.88$	-0.17	0.20	0.20	0.97	0.31
	$AS = 0.00$	0.01	0.29	0.31	1.00	0.03	$AS = 0.67$	-0.26	0.39	0.36	0.93	0.30
0.7	$PAE^{\omega_1} = 0.50$	-0.14	0.21	0.21	0.92	0.02	$PAE^{\omega_1} = 0.82$	-0.14	0.20	0.21	0.96	0.21
	$PAE^{\omega_2} = 0.50$	-0.14	0.20	0.19	0.92	0.02	$PAE^{\omega_2} = 0.85$	-0.15	0.17	0.19	0.96	0.28
	$PAE^{\omega_3} = 0.50$	-0.02	0.21	0.20	0.99	0.04	$PAE^{\omega_3} = 0.88$	-0.11	0.17	0.17	0.97	0.50
	$AS = 0.00$	-0.03	0.27	0.26	0.99	0.04	$AS = 0.67$	-0.22	0.29	0.29	0.91	0.47
0.9	$PAE^{\omega_1} = 0.50$	-0.06	0.16	0.16	0.92	0.03	$PAE^{\omega_1} = 0.82$	-0.07	0.16	0.17	0.97	0.45
	$PAE^{\omega_2} = 0.50$	-0.07	0.15	0.16	0.91	0.02	$PAE^{\omega_2} = 0.85$	-0.08	0.14	0.15	0.96	0.55
	$PAE^{\omega_3} = 0.50$	-0.05	0.20	0.18	0.98	0.04	$PAE^{\omega_3} = 0.88$	-0.05	0.13	0.12	0.96	0.75
	$AS = 0.00$	-0.08	0.24	0.22	0.98	0.05	$AS = 0.67$	-0.16	0.22	0.21	0.85	0.76

[a] ρ is the linear correlation of the simulated bivariate normal variables W and $S(1)$. Bias is the median bias. SE is the empirical standard error of $\widehat{PAE}^{\omega}$ and $\widehat{AS}$. SEE is the median of the bootstrap standard error estimates based on 500 bootstrap replicates. CP is the empirical coverage of bootstrap percentile 95% confidence intervals for PAE^{ω} and AS. Power is for 1-sided tests of $H_0 : PAE^{\omega} = 0.5$ versus $H_1 : PAE^{\omega} > 0.5$ or $H_0 : AS = 0$ versus $H_1 : AS > 0$ at level $\alpha = 0.05$. For the PAE weights, $\omega_1(j,1) = 1$, $\omega_2(j,1) = j$, and $\omega_3(j,1) = I[j = J = 4]$. 1000 simulations were done to compute the table elements for each model.

*In Gilbert and Hudgens (2008, *Biometrics*)

Conclusions of Simulation Study

- The MELE method of Gilbert and Hudgens performs well for realistically-sized Phase 3 vaccine efficacy trials, if there are baseline covariates that explain at least 50% of the variation in Y
- This underscores the importance of developing baseline predictors of immunogenicity endpoints
- Importantly, the good performance depends on the assumptions A3 and A4, which are not fully testable (more in discussion)
- R code for the nonparametric method of Gilbert and Hudgens (2008) is implemented in the R package *pseval* available at CRAN

Remarks on Power for Evaluating a Principal Surrogate Endpoint

- What about adding CPV to BIP?
- If the BIP is high quality (e.g., $\rho > 0.50$), then the BIP design is quite powerful with modest/moderate gain by adding CPV
- However, crossing over placebo subjects to the vaccine arm has additional value beyond efficiency improvement:
  - Helps in diagnostic tests of structural modeling assumptions (A4)
  - May help accrual and enhance ethics
  - May adaptively initiate crossover, after some overall VE > 0 is established (Gilbert et al., 2011, *Statistical Communications in Infectious Diseases*)

Outline of Session 6

1. Effect Modification/VE Curve Framework
2. Identifiability and Estimation
3. Simulations
4. **Discussion**

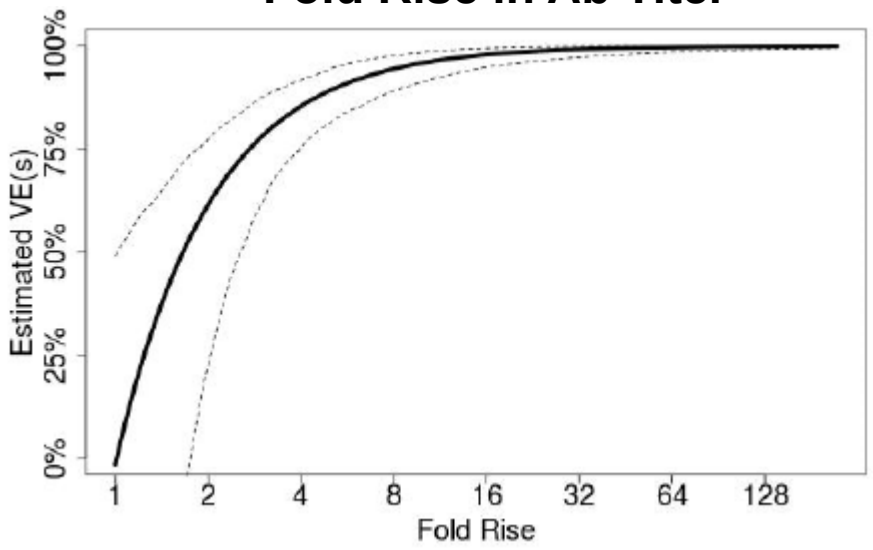
Some Avenues for Identifying Good BIPs

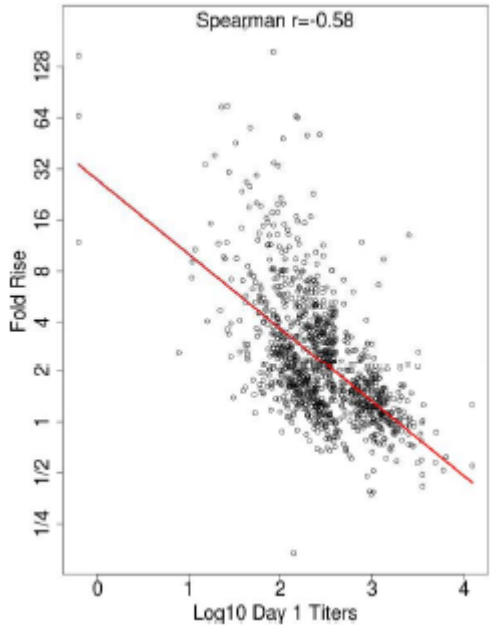
- Demographics
  - Age, gender, geography
- Host immune genetics
  - E.g., HLA type for predicting epitope-specific T cell responses (MHC binding prediction servers)
- Add beneficial licensed vaccines to efficacy trials and use known correlates of protection as BIPs (Follmann's [2006] original proposal)
  - The HVTN is exploring this strategy in a Phase 1 trial (HVTN 097) in preparation for VE trials
- Measure the marker S at baseline
  - Used successfully in varicella zoster (Gilbert, Gabriel et al., 2014, *JID*) and influenza VE trials
- Systems vaccinology analyses
  - E.g., Gene expression, cell populations; used successful for influenza vaccination (Tsang et al., 2014, *Cell*)

Example of a Successful BIP: Varicella Zoster Vaccine

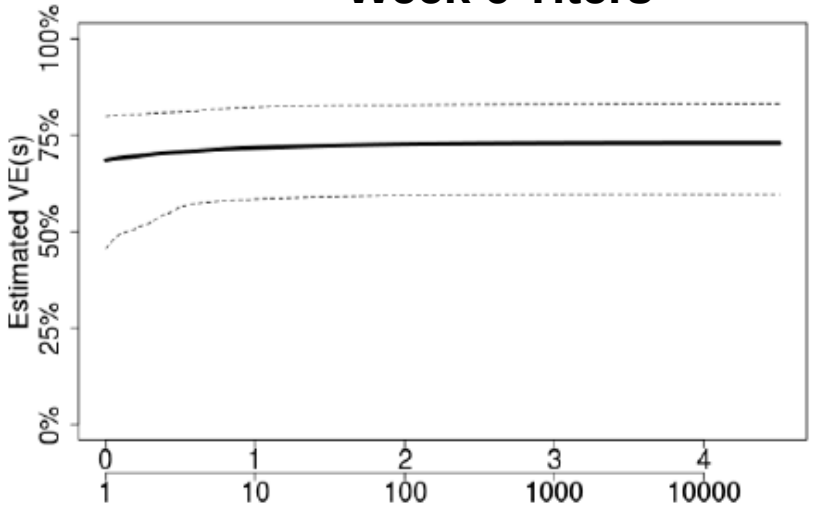
[Gilbert, Gabriel, et al., 2014]

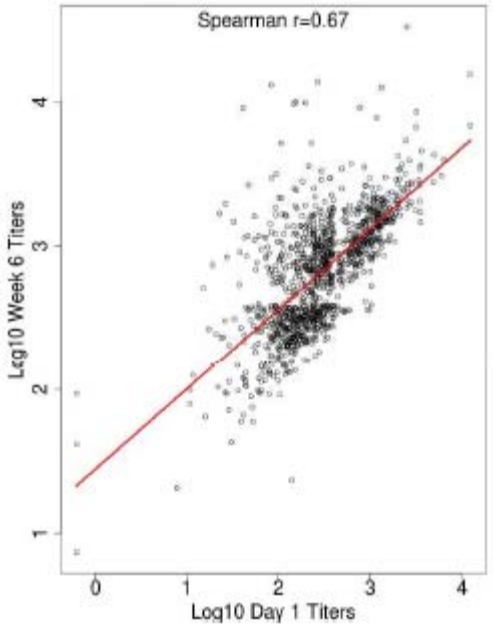
Fold Rise in Ab Titer





Week 6 Titers





Tetanus and Hepatitis B Vaccination in HVTN 097

(Ongoing Phase 1 Trial in South Africa)

Group	N	Month 0 (Day 0)	Month 1 (Day 28)	Month 2 (Day 56)	Month 4 (Day 112)	Month 7 (Day 196)	Month 7.5 (Day 210)	Month 8.5 (Day 238)	Month 13 (Day 394)
1	60	Tetavax®	ALVAC	ALVAC	ALVAC + AIDSVAX® B/E	ALVAC + AIDSVAX® B/E	ENGERIX-B®	ENGERIX-B®	ENGERIX-B®
2	20	Placebo	ALVAC	ALVAC	ALVAC + AIDSVAX® B/E	ALVAC + AIDSVAX® B/E	Placebo	Placebo	Placebo
3	20	Tetavax®	ALVAC	Placebo	Placebo	Placebo	ENGERIX-B®	ENGERIX-B®	ENGERIX-B®

ENGERIX-B® is a licensed hepatitis B vaccine

- Assess known correlates of protection as BIPs for a set of HIV-vaccine induced responses
- Antibodies to tetanus toxoid antigen and to hepatitis B surface antigen

Elaborations of the Original Methods

- Huang and Gilbert (2011, *Biometrics*) used the same VE estimands and assumptions as Gilbert and Hudgens (2008), with 3 extensions:
  - Relaxed the parametric assumptions on the distribution of (W, S)
  - Studied the method for using multiple immune biomarkers [e.g., assess if 2 immune response biomarkers provide superior surrogate value compared to 1]
  - Developed a new summary measure of surrogate value for 1 or more immune response biomarkers (standardized total gain)
- Huang, Gilbert, and Wolfson (2013, *Biometrics*) developed an improved 'pseudo-score' method incorporating BIP and/or CPV that is sometimes the method of choice
- Similar methods with a time-to-event clinical endpoint have been developed
  - Qin et al. (2008, *Annals of Applied Statistics*); Miao et al. book chapter (2013)
    - Cox proportional hazards model with discrete and continuous failure times
  - Gabriel and Gilbert (2014, *Biostatistics*)
    - Weibull model with continuous failure time
    - Allows for time-varying VE and time-varying surrogate value
  - Gabriel, Sachs, and Gilbert (2015, *Stat Med*)
    - Weibull model with continuous failure time
    - Compare and combine multiple biomarkers

Critical Assumptions for the Methods

- **Key assumptions of all published methods** [except one paper relaxed **A3** noted below]
  - **A3:** No causal effects on Y before the marker is measured [$P(Y^{\tau}(1) = Y^{\tau}(0))=1$]
    - Will approximately hold for some trials [e.g., if τ is near baseline]
    - Important to develop sensitivity analysis methods that account for departures from A3 [addressed in Wolfson and Gilbert (2010, *Biometrics*)]
  - **A4:** Structural models for $\text{risk}_{(z)}()$ functions, for z = 0, 1
    - The model for $\text{risk}_{(1)}()$ is fully testable
    - The model for $\text{risk}_{(0)}()$ is not fully testable
      - For each specific surrogate endpoint evaluation problem requires careful thought, accounting for biological knowledge
      - Use of closeout-placebo vaccination helps in testing modeling assumptions for $\text{risk}_{(0)}()$ [discussed in several papers]
  - Consistency of the MELE also depends on consistent estimation of the nuisance parameters– at least these assumptions are fully testable

Appendix:

R Tutorial

R Tutorial:
Application of Gilbert and Hudgens

- R code for implementing the nonparametric method of Gilbert and Hudgens (2008)

  A4-NP: Structural models for $\text{risk}_{(z)}$ (for z=0, 1)

  $$\text{risk}_{(z)}(j, 1, k; \beta) = \beta_{zj} + \beta'_k \text{ for } j=1, \ldots, J; k=1, \ldots, K$$

  Constraint: $0 \le \beta_{zj} + \beta'_k \le 1$ and $\Sigma_k \beta'_k = 0$ for identifiability

- Recall the setting for which this method applies:
  - Constant Biomarkers (no or minimal variation in S in placebo recipients)
  - The BIP design is used with the baseline immunogenicity predictor W a categorical variable
  - The biomarker to evaluate as a specific SoP, S, is categorical
- R code at: http://faculty.washington.edu/peterg/SISMID2016.html

R Tutorial:
Application of Gilbert and Hudgens

- Exercise: Apply the Gilbert and Hudgens method to the same data-set that was assessed earlier for evaluating a CoR
- W = a binned/discretized version of the infectivity assay result
- S = a binned/discretized version of the MN Neuts measurement, and of the CD4 Blocking measurement
- Feel free to try one or another discretizations
  - E.g., cut W and S into quartiles; or cut S into 2 parts in the search of a 'threshold of protection'

R Tutorial:
Application of Gilbert and Hudgens

- Suggest to perform the set of analyses that were done on the simulated data-sets described earlier
  - For each level j estimate β_{1j}, β_{0j} and hence estimate the parameter of interest
  - $VE(j, 1) = 1 - \log(\text{avg-risk}_{(1)}(j, 1) / \text{avg-risk}_{(0)}(j, 1))$
  - $= 1 - \log(\Sigma_k [\beta_{1j} + \beta'_k] / [\Sigma_k \beta_{0j} + \beta'_k])$
  - Estimate AS and PAE(w) for w(j) = 1 and for w(j) = j
  - Compute 95% confidence intervals for each of the above parameters
  - Compute p-values for testing
    - H0j: VE(j, 1) = 0 vs H1j: VE(j, 1) > 0, for j = 1, …, J
    - H0: AS = 0 vs H1: AS > 0
    - H0: PAE(w) = 0.5 vs H1: PAE(w) > 0.5
    - H0: VE(j, 1) = 0 for all j vs H1: VE(j, 1) monotone non-decreasing in j with some < [trend test]

R Tutorial:
Application of Gilbert and Hudgens

- Is there evidence that either MN Neuts or CD4 Blocking levels have some value as a principal surrogate endpoint?
- If so, what quality is the principal surrogate endpoint? How to interpret the results in terms of effect modification?