Module 8:
Evaluating Immune Correlates of Protection

Instructors: **Ivan Chan, Peter Gilbert, Paul T. Edlefsen, Ying Huang**

Session 9:
Evaluating a Specific Surrogate of Protection Part I

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

July 15-17, 2013

Statistical Center for HIV/AIDS Research & Prevention SCHARP

1

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Outline of Module 8

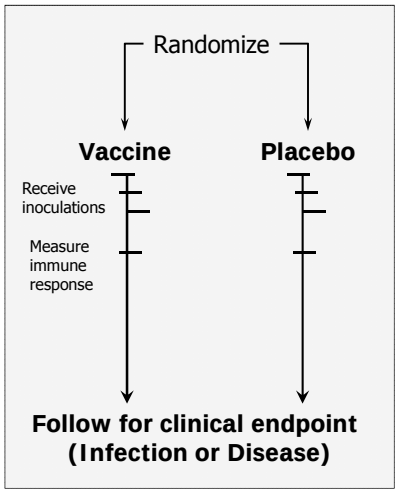
Session 1 (Chan)	Introduction to Vaccines and Basic Concepts
Session 2 (Gilbert)	Introduction to Immune Correlates of Protection
Session 3 (Chan)	Evaluating Correlates of Protection using Individual, Population, and Titer-Specific Approaches
Session 4 (Gilbert)	Continuation of Session 2; plus Evaluating a Correlate of Risk (CoR)
Session 5 (Chan)	Use of Statistical Models in Assessing Correlates of Protection
Session 6 (Edlefsen)	Introduction to Sieve Analysis
Session 7 (Gilbert)	Thai Trial Case Study (Including Sieve Analysis)
Session 8 (Chan)	Validation using Prentice Criteria, Design Considerations
Session 9 (Gilbert)	**Evaluating a Specific Surrogate of Protection Part I (Gilbert and Hudgens, 2008)**
Session 10 (Huang)	Evaluating a Specific Surrogate of Protection Part II (Huang and Gilbert, 2011; Huang, Gilbert and Wolfson, 2013)

Statistical Center for HIV/AIDS Research & Prevention SCHARP

2

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Preventive Vaccine Efficacy Trial



- **Primary Objective**
  - Assess VE: Vaccine Efficacy to prevent infection or disease with a pathogen
- **Secondary Objective**
  - Assess vaccine-induced immune responses as surrogate endpoints for infection or disease

Statistical Center for HIV/AIDS Research & Prevention SCHARP

3

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Three Tiers of Surrogate Endpoint Evaluation*

	Definition	Framework for Empirical Assessment
Correlate of Risk Tier 1	The biomarker correlates with the clinical endpoint measuring vaccine efficacy	Vaccine efficacy trials/ epidemiological studies
Specific Surrogate of Protection Tier 2	Vaccine effects on the biomarker predict vaccine efficacy, for the **same setting** as the efficacy trial	Single large efficacy trial or multiple similar trials
General Surrogate of Protection Tier 3	A specific SoP that reliably predicts vaccine efficacy in **different settings** (e.g., across vaccine lots, vaccine formulations, human populations, viral populations)	Multiple diverse efficacy and/or post-licensure trials

***Proposed in Qin, Gilbert, McElrath, Corey, Self (2007, JID)**

Statistical Center for HIV/AIDS Research & Prevention SCHARP

4

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Outline of Session 9

1. Overview of Four Approaches to Defining and Evaluating a Surrogate Endpoint
2. Principal Stratification Approach
3. Identifiability and Estimation
4. Simulations
5. Discussion
6. R Tutorial

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

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Part 1:

Overview of Four Approaches to Defining and Evaluating a Surrogate Endpoint

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

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) [The approach also applies for quantitative Y]

Four Frameworks for Surrogate Endpoints
(Joffe and Greene, 2008, Biometrics)

- **Causal-effects paradigm**

  "for a good surrogate, the effect of treatment on the surrogate, combined with the effect of the surrogate on the clinical outcome, allow prediction of the effect of treatment on the clinical outcome"

  1. Prentice/statistical surrogate — *Valid replacement endpoint*
  2. Controlled natural direct and indirect effects — *Mediation*

- **Causal-association paradigm**

  "for a good surrogate, the effect of treatment on the surrogate is associated with its effect on the clinical outcome"

  3. Principal stratification — *Association of individual-level treatment effects*
  4. Meta-analysis — *Association of group-level treatment effects*

1. Prentice / Statistical Surrogate Framework

- **In the introductory talk, we noted 3 challenges posed to this framework**
  1. Hard to evaluate operational criteria for vaccine trials for which there is ~no variability of the immunological biomarker in the placebo group
  2. For validity must include in the regression model all common causes (simultaneous predictors) of the biomarker and the clinical outcome
  3. For validity must include in the regression model all common causes (simultaneous predictors) of clinical risk before and after the biomarker is measured
- **Let's elaborate on points 2. and 3.**
  - Grateful to Marshall Joffe and Tom Greene for their 2008 Biometrics paper that helpfully elucidates point 2

Prentice (1989, Stat Med): Criteria for a Surrogate Endpoint

- **Definition of a Surrogate Endpoint**

  "a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint"

- **Main Operative Criteria**
  1. The surrogate and clinical endpoints are correlated, in each treatment arm
  2. All of the treatment effect on the clinical endpoint is mediated through the surrogate: $Y \perp Z \mid S$
     - Necessary condition for full mediation 2.:
       $$\Pr(Y = 1 \mid S = s, Z = 1) = \Pr(Y = 1 \mid S = s, Z = 0) \textbf{ for all } s \quad (*)$$
     - A biomarker satisfying (*) is a *Statistical Surrogate*
       [Frangakis and Rubin, 2002, Biometrics]

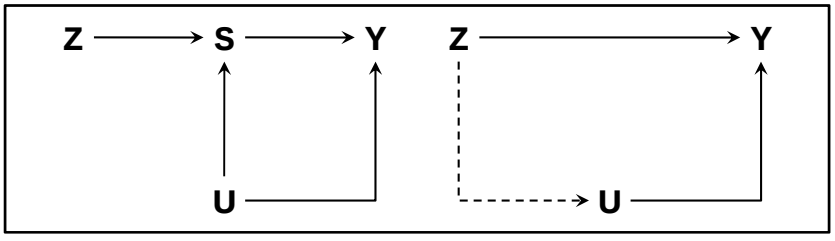
Statistical Surrogate Criteria in Terms of GLMs

- **Effect of Z on Y:**
  - logit(E[Y|Z]) = b0 + Z b1
  - b1 = overall treatment effect on clinical outcome
- **Effect of (Z, S) on Y:**
  - logit(E[Y|Z, S]) = a0 + Z a1 + S a2
  - a1 = treatment effect on clinical outcome controlling for S
  - Full mediation condition holds if a1 = 0
- **More generally, 1 – a1/b1** is the proportion of the treatment effect explained (PTE) by S

  [Freedman, Graubard, Schatzkin, 1992, Stats Med]

Unmeasured Common Causes (Graphical)

- a1=0 means that E[Y|Z=1,S=s] = E[Y|Z=0,S=s] for all s
- The groups {Z=1, S=s} and {Z=0, S=s} are apples and oranges if there are unmeasured common causes of S and Y (Pearl, 2000, *Causality*)



If a common cause U is unaccounted for, then conditioning on S induces an association of Z and U, and thereby of Z and Y

What if There are Unmeasured Common Causes of S and Y?

- If a common cause U is unaccounted for, then conditioning on S induces an association of Z and U, and thereby of Z and Y
  - Similar to the 'no unmeasured confounders' assumption routinely made for causal analyses in epidemiological studies
- Thus, even if S mediates the entire effect of Z on Y and Z has no direct effect on Y controlling for S, it does not follow that

$$Z \perp Y \mid S$$ (or, equivalently, that a1 = 0)

[this phenomenon discussed broadly including in Rosenbaum (1984, JRRS-B), Robins (1986, Math Modeling), Pearl (2000, Causality textbook), Frangakis and Rubin (2002, Biometrics)]

What if There are Unmeasured Common Causes of S and Y?

- **Practical point 1:** When checking the full mediation condition include all baseline covariates X that may predict both S and Y
  - Effect of Z on Y controlling for (X,S)

    $\text{logit}(E[Y \mid Z,S]) = a0 + Z\,a1 + S\,a2 + X\,a3$

  a1 = treatment effect on clinical outcome controlling for X and S

  a1 = 0 indicates full mediation IF X captures all common causes

When designing a trial plan to collect the putative simultaneous predictors!

- **Practical point 2:** Acknowledge there may be residual confounding, which may make important a sensitivity analysis

What if Some Subjects Experience Y=1 Before S is Measured?

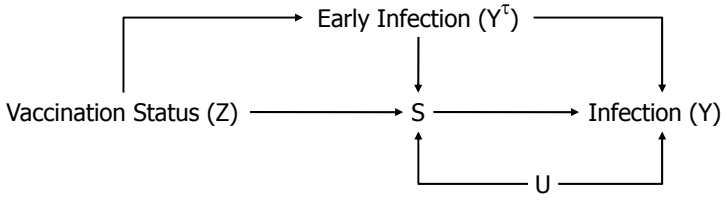
- For simplicity Joffe and Greene (2008) assumed S and Y are both measured once, at fixed times, with S measured before Y, and S and Y are never missing
- In practice, typically some (or many) subjects experience Y=1 before S is measured
  - e.g., VaxGen HIV vaccine efficacy trial (Flynn et al., 2005)
    - S is measured at month 6.5 post-randomization
      - 62 of the 368 total HIV infections (17%) occurred prior to month 6.5
  - e.g., RV144 HIV vaccine efficacy trial (Rerks-Ngarm et al., 2009)
    - S is measured at month 6 post-randomization
      - 15 of the 125 total HIV infections (12%) occurred prior to month 6

Unmeasured Simultaneous Predictors of Early and Later Clinical Risk

- **Y^{τ} = Indicator of infection before the biomarker is measured**
- **For validity the statistical surrogate approach assumes no unmeasured simultaneous predictors V of Y^{τ} and Y**
  - That is, for validity there cannot be any unaccounted for subject characteristics that predict both early and later clinical risk
  - Practically speaking, this means that for validity one must control for all clinical prognostic factors

Summary on No Unmeasured Simultaneous Predictors

- **For validity the statistical surrogate approach assumes both:**
  - No unmeasured simultaneous predictors V of Y^{τ} and Y (i.e., of early and later clinical risk)
  - No unmeasured simultaneous predictors U of S and Y



Statistical Center for HIV/AIDS Research & Prevention SCHARP

17

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

2. Controlled Natural Direct / Indirect Effects Framework (Mediation)

- Literature with key contributors Judea Pearl, James Robins, Sanders Greenland
  - Robins and Greenland (1992, *Epidemiology*)
  - Pearl (2001, *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*)
- Need potential outcomes notation (Neyman, 1923; Rubin, 1974)

Statistical Center for HIV/AIDS Research & Prevention SCHARP

18

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Potential Outcomes/Counterfactuals

- **Potential Outcomes/Counterfactuals Framework**
  - Use of counterfactuals by the Father of Probability: Blaise Pascal, *The Pensees (1660),* No. 413, Lafuma Edition

"Cleopatra's nose: if it had been shorter the whole face of the earth would have been different."

- Most historians acknowledge that Marc Antony's falling in love with Cleopatra played a major role in the fall of the Roman Republic
- For many, counterfactuals are a natural way of thinking

Potential Outcomes/Counterfactuals

- **Notation**
  - $\mathbf{S_i(Z)}$ = potential immune response endpoint under assignment **Z;** for **Z** = 0, 1
  - $\mathbf{Y_i(Z)}$ = potential clinical endpoint under assignment **Z;** for **Z** = 0, 1
- **Individual Causal Effects**
  - A contrast in $\mathbf{S_i(1)}$ and $\mathbf{S_i(0)}$ is a causal effect on **S** for subject **i**
  - A contrast in $\mathbf{Y_i(1)}$ and $\mathbf{Y_i(0)}$ is a causal effect on **Y** for subject **i**
- **Average causal effects:** $E[S_i(0) - S_i(1)]$, $E[Y_i(0) - Y_i(1)]$

2. Controlled Natural Direct and Indirect Causal Effects (Mediation)

- **Notation**
  - $Y_i(Z, S)$ = potential clinical endpoint under assignment to Z and to S
- **Individual Causal Effects**
  - A contrast in $Y_i(z, s)$ and $Y_i(z', s')$ is a causal effect on Y for subject i
  - Direct effect **(at s):** $Y_i(0, s) - Y_i(1, s)$ **[hold S fixed at s]**
  - Indirect effect **(at s):** Overall effect – direct effect
    $[Y_i(0) - Y_i(1)] – [Y_i(0, s) - Y_i(1, s)]$
- **Average Causal Effects**
  - Direct effect (at s): $E[Y_i(0, s) - Y_i(1, s)]$
  - Indirect effect (at s): $E[Y_i(0) - Y_i(1)] – E[Y_i(0, s) - Y_i(1, s)]$

2. Controlled Natural Direct and Indirect Causal Effects (Mediation)

- Average Causal Effects
  - Direct effect (at s): $E[Y_i(0, s) - Y_i(1, s)]$
  - Indirect effect (at s): $E[Y_i(0) - Y_i(1)] – E[Y_i(0, s) - Y_i(1, s)]$
- A valid surrogate in this paradigm has no direct effect for all s
  - i.e., $E[Y_i(0, s) - Y_i(1, s)] = 0$ for all s
  - That is, S **fully mediates** the effect of Z on Y
  - i.e., "the treatment effect on the clinical endpoint is fully through the surrogate/fully mediated by the treatment effect on the surrogate endpoint"
- A useful conceptual framework, decomposing the overall effect into component effects

2. Natural Direct / Indirect Effects Framework (Mediation)

- Of the 4 frameworks, this one may be the best suited for assessing mediation (e.g., as argued by papers of Tyler VanderWeele)
- However, this approach requires **conceivability of manipulating** a placebo recipient's biomarker level to what it would have been had s/he been assigned the vaccine
  - In trials of subjects without prior exposure to the pathogen: Inconceivable
  - In trials of subjects with prior exposure: May be conceivable in rare instances, but more likely inconceivable due to heterogeneity of host genetics and other host factors
  - Where it is conceivable, it is still challenging to assess mediation because unverifiable assumptions are needed (and thus sensitivity analysis is warranted)
- Gilbert, Hudgens, and Wolfson (2011, International Journal of Biostatistics) discuss the conceivability and utility of this approach

Part 2:

Principal Stratification Approach
(Gilbert and Hudgens 2008, Biometrics)

Principal Surrogate Endpoints
(Frangakis and Rubin 2002, Biometrics)

- In the principal stratification 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
- **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

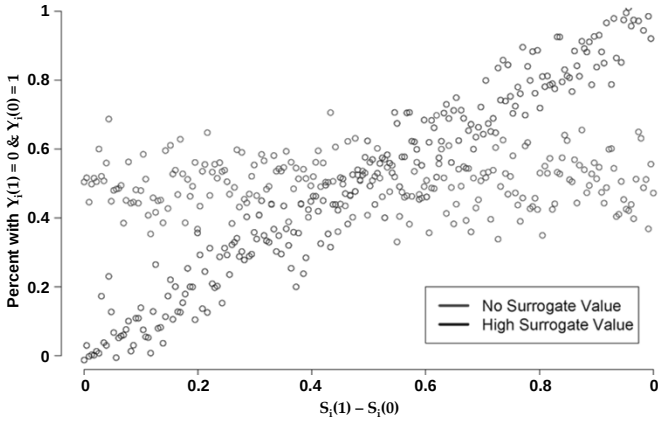
Statistical Center for HIV/AIDS Research & Prevention SCHARP

25

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Heuristic of Principal Surrogate Approach

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



Statistical Center for HIV/AIDS Research & Prevention SCHARP

26

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Assumptions

A1 Stable Unit Treatment Value Assumption (SUTVA):

$(S_i(1), S_i(0), Y_i(1), Y_i(0))$ is independent of the treatment assignments Z_j of other subjects

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

A2 Ignorable Treatment Assignments:

Z_i is independent of $(S_i(1), S_i(0), Y_i(1), Y_i(0))$

- **A2** holds for randomized blinded trials

A3 Equal individual clinical risk up to time τ that S is measured [$Y^{\tau}_i(1) = 1$ if and only if $Y^{\tau}_i(0) = 1$]

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

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

- $\mathbf{S_i(1) = S_i(0) \Rightarrow Y_i(1) = Y_i(0)}$

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

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

- **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 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., "without the vaccine-induced immune response, there is no protection"

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Definition of a Principal Surrogate
(Revised from 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 $\text{risk}_{(0)}(s_1, s_0)$] is sensitive to the values (s_1, s_0)

  i.e., the variables (S_1, S_0) modify vaccine efficacy
- **Thus,** a **principal surrogate** is defined to be a biomarker satisfying average causal necessity [need a vaccine-induced immune response for protection] and that is an effect modifier [clinical treatment efficacy varies markedly with levels of (S_1, S_0)]
- **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 more useful surrogate

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Principal Surrogate Value

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

  s_1 near s_0 ⇨ $\mathbf{risk}_{(1)}(s_1, s_0)$ is near $\mathbf{risk}_{(0)}(s_1, s_0)$

  – There are some $s_1 \neq s_0$ for which $\mathbf{risk}_{(1)}(s_1, s_0)$ is far from $\mathbf{risk}_{(0)}(s_1, s_0)$

- **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

  – $s_1 \neq s_0 \Rightarrow E[Y(1) \mid s_1, s_0] \neq E[Y(0) \mid s_1, s_0]$

  – **i.e., "A vaccine effect on the marker implies there is some protection"**

Statistical Center for HIV/AIDS Research & Prevention SCHARP

31

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Principal Surrogate Value

- Two heuristically desirable properties of a good surrogate:
  1. **Specificity:** VE = 0% implies S(1) $=^d$ S(0)
     - i.e., S(1) $\neq^d$ S(0) implies VE ≠ 0%
  2. **Sensitivity:** VE ≠ 0% implies S(1) $\neq^d$ S(0)
     - i.e., S(1) $=^d$ S(0) implies VE = 0%
- Case CB:
  – ACN ⇒ Sensitivity
  – ACN + Strong ACS ⇒ Specificity
- General Case:
  – ACN + Strong ACS ⇒ Specificity
  – ANC + Strong ACS ⇒ Sensitivity under 1 of 2 extra conditions
  [P(S(1) ≥ S(0))=1 or 'no harm for any subgroup' CEP(s_1, s_0) ≥ 0]

This highlights that ACN and Strong ACS are useful conditions to check

Statistical Center for HIV/AIDS Research & Prevention SCHARP

32

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

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

Causal Effect Predictiveness (CEP) 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:*

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

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

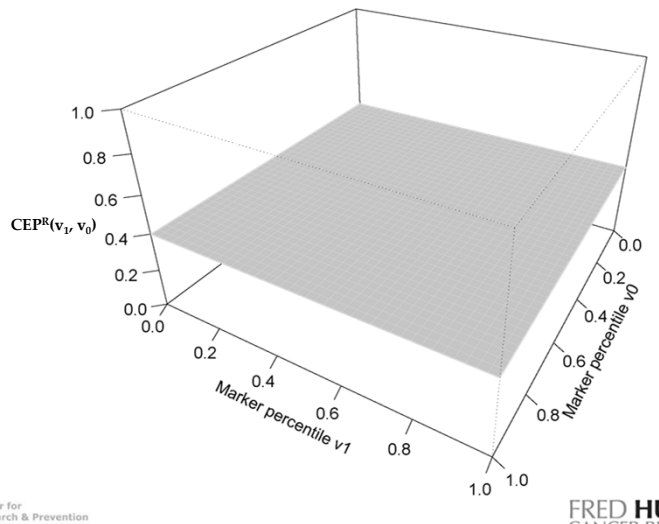
$$R_{(1)}(v_1, v_0) = Pr(Y(1) = 1 | S(1) = F^{-1}_{(1)}(v_1), S(0) = F^{-1}_{(1)}(v_0))$$

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

- CEP surface:

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

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

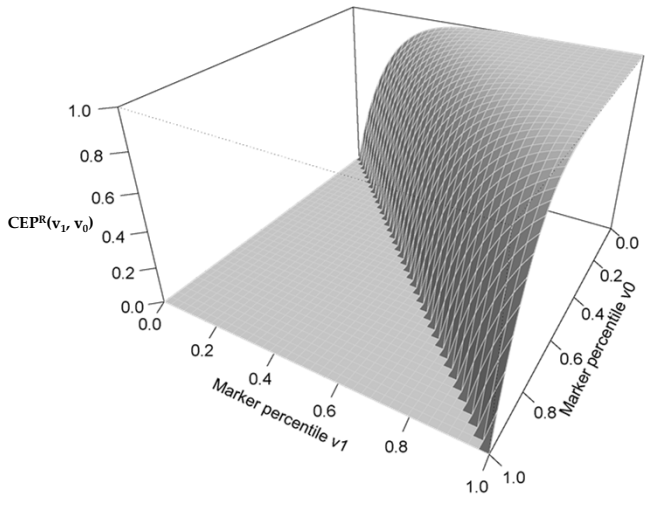


Statistical Center for HIV/AIDS Research & Prevention
SCHARP

35

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

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



Statistical Center for HIV/AIDS Research & Prevention
SCHARP

36

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

CEP Surface in Special Case
$S_i(0) = c$ for all i

- This special case typically occurs in vaccine trials where enrolled subjects are naïve to the pathogen
- In this case the CEP surface is a curve

  $CEP^{risk}(s_1, c)$ or $CEP^{R}(v_1, F_{(1)}(c))$
- A principal surrogate is a biomarker with

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

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

Marginal CEP Curve for the General Case

- Define

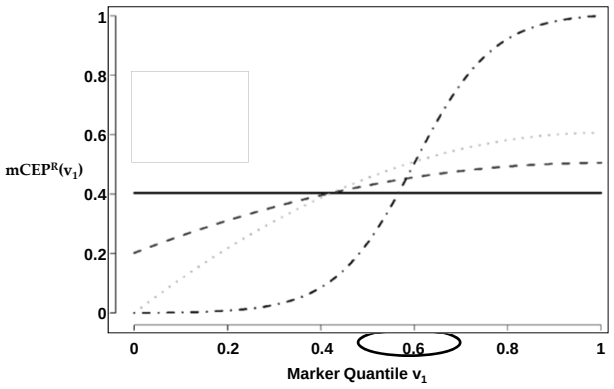
  $risk_{(1)}(s_1) = Pr(Y(1) = 1 | S(1) = s_1)$

  $risk_{(0)}(s_1) = Pr(Y(0) = 1 | S(1) = s_1)$
- Marginal CEP curve:

  $mCEP^{R}(s_1) = h(risk_{(1)}(s_1), risk_{(0)}(s_1))$
- Marginal CEP curve with percentile formulation:

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

Illustration of Marginal CEP Curve



Summary Measures of Surrogate Value

- Consider 1-sided setting where interest is in assessing if greater treatment 1 biomarker responses predict clinical benefit of treatment 1 [e.g., placebo-controlled trial]
- 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

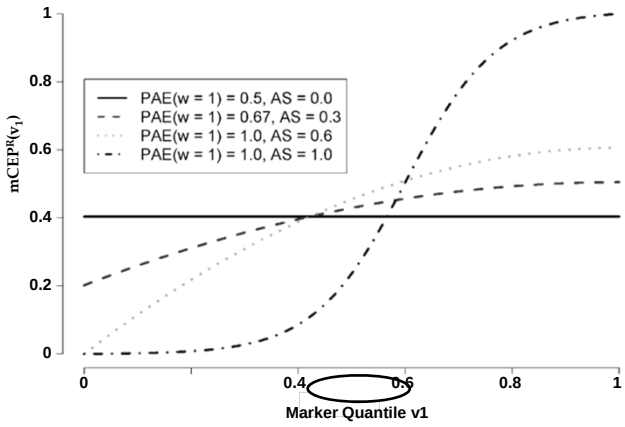
Summary Measures of Surrogate Value

- 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

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

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

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Challenge to Evaluating a Principal Surrogate: Missing Data

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

Statistical Center for HIV/AIDS Research & Prevention SCHARP

43

FRED HUTCHINSON CANCER RESEARCH CENTER VACCINE AND INFECTIOUS DISEASE DIVISION

Part 3:

Identifiability and Estimation of the CEP Surface and CEP Curve

Statistical Center for HIV/AIDS Research & Prevention SCHARP

44

FRED HUTCHINSON CANCER RESEARCH CENTER VACCINE AND INFECTIOUS DISEASE DIVISION

Dean Follmann's Augmented Vaccine Efficacy Trial Designs

- Follmann (2006, Biometrics) proposed augmented vaccine trial designs for discerning whether an immune response reliably predicts VE (i.e., a specific SoP)
- 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

Baseline Immunogenicity Predictor (BIP) Approach

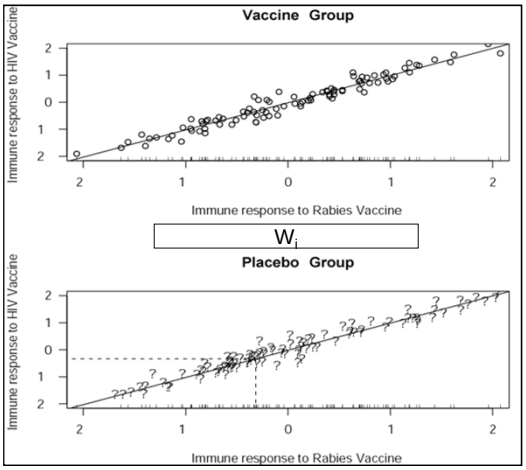
- More carefully, recall that S is only meaningfully measured in subjects who have not experienced the disease endpoint by time τ when S is measured – so the condition we really need is

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

- This requirement is the reason why we assume A3 [i.e., $Y^{\tau}_i(1) = 1$ if and only if $Y^{\tau}_i(0) = 1$] : A1-A3 imply (*)
- 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
- For notational simplicity, henceforth all conditional distributions implicitly condition on $Y^{\tau} = 0$

Baseline Immunogenicity Predictor (BIP) Approach*

$S_i(1)$



$S_i(1)$

Evaluate correlation of **W** and **S(1)** in vaccine group

Predict **S(1)** from vaccine group model and **W** in placebos

*Figure from Follmann (2006, Biometrics)

Build on Two-Phase Sampling Methods

- **Case-cohort or case-control sampling**
  - **(W, S(1))** measured in
    - All infected vaccines
    - Sample of uninfected vaccines
  - **W** measured in
    - All infected placebos
    - Sample of uninfected placebos
- **2-Phase designs** (E.g., Prentice, 1986, Biometrika; 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

IPW Case-Cohort Methods Do Not Apply:
Hence we use a Full Likelihood-Based Method

- None of the case-cohort methods described earlier in the workshop apply to this problem
- The reason is that they are all inverse probability weighted (IPW)-based methods, using partial likelihood 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
(Pepe and Fleming, 1991)

- Posit models for $\text{risk}_{(1)}(s_1,0;\beta)$ and $\text{risk}_{(0)}(s_1,0;\beta)$
- Vaccine arm:
  - $(\mathbf{W_i, S_i(1)})$ measured: Likld contribn $\mathbf{risk_{(1)}(S_i(1), 0; \beta)}$
  - $(\mathbf{W_i, S_i(1)})$ not measured: $\mathbf{\int risk_{(1)}(s_1, 0; \beta)\, dF(s_1)}$
- Placebo arm:
  - $\mathbf{W_i}$ measured: Likld contribn $\mathbf{\int risk_{(0)}(s_1, 0; \beta)\, dF^{S|W}(s_1 | W_i)}$
  - $\mathbf{W_i}$ 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:**
  - $F^{S|X,W}$ and $F^{W|X}$ have specified parametric distributions
  - 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 $risk_{(z)}$ (for z=0, 1)
  - $risk_{(z)}(s_1, 0, x, w; \beta_z) = g(\beta_{z0} + \beta_{z1} s_1 + \beta^T_{z2} x + \beta^T_{z3} w)$, g a known link
- **Example:**

  $F^{W|X}$ normal, $F^{S|X,W}$ censored normal with left-censoring below 0, A4-P holds with g = Φ, the standard normal cdf
- **No interactions assumption:** One of the components of $(\beta^T_{12}, \beta^T_{13})$ equals the corresponding component of $(\beta^T_{02}, \beta^T_{03})$ (untestable)

Modeling Approach 1
(Fully Parametric)

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

  $CEP^{risk}(s_1, 0, x, w) = (\beta_{10} - \beta_{00}) + (\beta_{11} - \beta_{01})s_1 + (\beta_{12} - \beta_{02})^T x + (\beta_{13} - \beta_{03})^T w$

- Under assumption of no interactions between Z and X nor between Z and W:

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

  = covariate-adjusted CEP-curve

Parametric Approach:
Interpretation of Parameters

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

- 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)

    $risk_{(z)}(j, 1, k; \beta) = \beta_{zj} + \beta'_k$ for j=1, …, J; k=1, …, 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 effect on 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 $\text{avg-risk}_{(z)}(j, 1) = (1/K) \Sigma_k risk_{(z)}(j, 1, k; \beta)$

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

Interpretation
(Fully Nonparametric)

- With $VE(j, 1) = 1 - \text{avg-risk}_{(1)}(j, 1) / \text{avg-risk}_{(0)}(j, 1)$:
  - **S** is a **principal surrogate if**

    $\mathbf{VE(1, 1) = 0}$ and $\mathbf{VE(j, 1) > 0}$ for all $j > 1$
- A biomarker with some value as a surrogate will have
  - $\mathbf{VE(1, 1)}$ near **0**
  - $\mathbf{VE(j, 1)} > 0$ for some $j > 1$
- The most useful surrogate will have $VE(j, 1)$ large for some $j > 1$

Modeling Approach 2
(Fully Nonparametric)

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

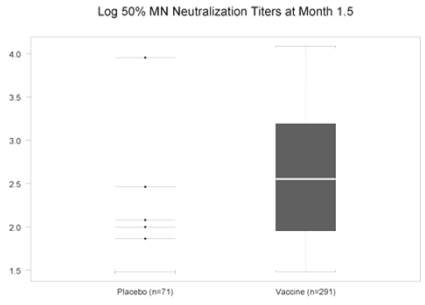
    $\text{Est. } \mu_z = (1/J) \Sigma_j \beta_{zj}$

Part 4:

Simulations

Simulation Plan
(Based on VaxGen Efficacy Trial)

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



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

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) = 0.067 and P(Y(0) = 1) = 0.134 (overall VE = 50%)
    - The biomarker has either (i) no or (ii) high surrogate value

Statistical Center for HIV/AIDS Research & Prevention SCHARP

61

FRED HUTCHINSON CANCER RESEARCH CENTER VACCINE AND INFECTIOUS DISEASE DIVISION

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

Statistical Center for HIV/AIDS Research & Prevention SCHARP

62

FRED HUTCHINSON CANCER RESEARCH CENTER VACCINE AND INFECTIOUS DISEASE DIVISION

Simulation Plan

- Step 3: Create case-cohort 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 VaxGen 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 $\widehat{CEP}^{risk}(j,1;\beta) = \log(\hat{\beta}_{1j}/\hat{\beta}_{0j})$ 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 VaxGen 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): $\text{CEP}^{\text{risk}}(s_1, 0) = \beta_{10} - \beta_{00} = -0.11$ [AS = 0; PAE(w) = 0.5]

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

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)

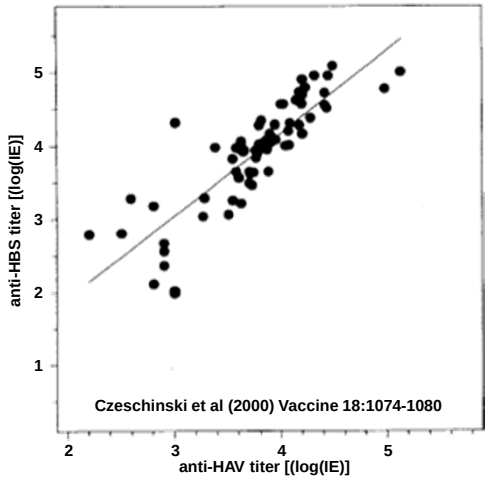
Conclusions of Simulation Study

- 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 available at the Biometrics website and at http://faculty.washington.edu/peterg/SISMID2013.html

Remarks on Power for Evaluating a Principal Surrogate Endpoint

- Crossing over more placebo subjects improves power of CPV and BIP + CPV designs
- There is no point of diminishing returns– steady improvement with more crossed over, out to complete cross-over
- 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 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)

Example of a Good Baseline Predictor for Which A4-P May Be Plausible: Antibody Levels to Hepatitis A and B Vaccination (n=75)



r = .85

No cross-reactivity: Supports plausibility of A4-P

Statistical Center for HIV/AIDS Research & Prevention SCHARP

71

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Some Avenues for Identifying Good BIPs

- Demographic factors
  - Age, gender, immune status
- Host 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 in preparation for efficacy trials

Statistical Center for HIV/AIDS Research & Prevention SCHARP

72

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Tetanus and Hepatitis B Vaccination in HVTN 097 (Planned 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®

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

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Part 5:

Discussion

Statistical Center for HIV/AIDS Research & Prevention SCHARP

FRED HUTCHINSON CANCER RESEARCH CENTER
VACCINE AND INFECTIOUS DISEASE DIVISION

Summary

- The CEP surface/marginal curve has a useful prediction interpretation for quantifying the surrogate value of a biomarker
  - May also be called a "principal causal effect" surface or curve, or simply the "vaccine efficacy" surface/curve
- The original/initial methods were developed for estimation and testing under baseline predictor and/or close-out placebo vaccination study designs
  - Binary and quantitative clinical endpoints (Follmann, 2006; Gilbert and Hudgens, 2008)

Elaborations of the Original Methods

- Huang and Gilbert (2011, Biometrics) used the same 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 biomarkers provide superior surrogate value compared to 1]
  - Developed a new summary measure of surrogate value for 1 or more immune biomarkers (standardized total gain)
- Huang, Gilbert, and Wolfson (2013, Biometrics) developed an improved 'pseudo-score' method incorporating BIP and/or CPV that is now the method of choice (Session 10)
- Similar methods with a time-to-event clinical endpoint have been developed
  - Qin et al. (2008, Annals of Applied Statistics)
    - Cox proportional hazards model with discrete failure time
  - Erin Gabriel (2012 Ph.D. dissertation)
    - Weibull model with continuous failure time
    - Allows for time-varying VE and time-varying surrogate value

Critical Assumptions for the Methods

- **Key assumptions of Gilbert and Hudgens (2008) and Huang and Gilbert (2011)**
  - **A3:** Equal individual clinical risk up to time τ that S is measured [$Y^{\tau}_i(1) = 1$ if and only if $Y^{\tau}_i(0) = 1$]
    - 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 $risk_{(z)}()$ functions, for z = 0, 1
    - The model for $risk_{(1)}()$ is fully testable
    - The model for $risk_{(0)}()$ 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 $risk_{(0)}()$ [discussed in Follmann, 2006, Biometrics]
  - Consistency of the MELE also depends on consistent estimation of the nuisance parameters- at least these assumptions are fully testable

Example Thought Process to Justify an A4-P Assumption

- A4-P: Structural models for $risk_{(z)}$ (for z=0, 1)

  $risk_{(z)}(s_1, 0, x, w; \beta_z) = g(\beta_{z0} + \beta_{z1} s_1 + \beta^T_{z2} x + \beta^T_{z3} w)$, g a known link
- No interactions assumption: One of the components of $(\beta^T_{12}, \beta^T_{13})$ equals the corresponding component of $(\beta^T_{02}, \beta^T_{03})$ (untestable)
- Example:

  $risk_{(1)}(s_1, 0, x, w; \beta_1) = \Phi(\beta_{10} + \beta_{11} s_1 + \beta_{12} x + \beta_3 w)$

  $risk_{(0)}(s_1, 0, x, w; \beta_1) = \Phi(\beta_{00} + \beta_{01} s_1 + \beta_{02} x + \beta_3 w)$

  This model allows baseline covariates X to effect Y differently in the vaccine and placebo groups; but assumes that, after accounting for X, W effects Y in the same way in the vaccine and placebo groups

Part 6:

R Tutorial

R Tutorial:
Application of Gilbert and Hudgens

- R code for implementing the nonparametric method of Gilbert and Hudgens introduced on slide 53:

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

  $\text{risk}_{(z)}(j, 1, k; \beta) = \beta_{zj} + \beta'_k$ for j=1, …, J; k=1, …, 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/SISMID2013.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 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 specific SoP?
- If so, what quality is the surrogate endpoint?