The Search for an Optimal Immunological Surrogate Endpoint in Randomized Vaccine Efficacy Trials

Peter Gilbert, Brenda Price, Mark van der Laan

Sanofi Pasteur, Swiftwater, PA, September 26, 2018

Outline

1. Surrogate Endpoint Frameworks
2. Optimal Surrogate Framework
3. Simulation Studies
4. Application to Dengue VE Trials
5. Discussion
Surrogate Endpoint Frameworks

Optimal Surrogate Framework

Simulation Studies

Application to Dengue Trials

Discussion
Preventive Vaccine Efficacy Trial

Primary Objective
Assess vaccine efficacy (VE) to prevent infection or disease with a pathogen

Secondary Objective
Assess immune response biomarkers measured after vaccination as “surrogate endpoints”
Randomized Vaccine Efficacy Trial Notation

A → S → Y

Fixed follow-up period of duration $\tau_1$

- $A =$ treatment (1=vaccine, 0=placebo or other control)
- $Y =$ clinical endpoint (1=event by $\tau_1$, 0=event-free at $\tau_1$)
- $S =$ candidate surrogate measured at time $\tau < \tau_1$

Vaccine Efficacy Parameter

$\text{VE} = 1 - \frac{P(Y = 1 | A = 1)}{P(Y = 1 | A = 0)}$

= Percent reduction in endpoint rate by vaccination vs. control group
Randomized Vaccine Efficacy Trial Notation

\[ A \longrightarrow S \longrightarrow Y \]

**Fixed follow-up period of duration \( \tau_1 \)**

- \( A = \) treatment (1= vaccine, 0= placebo or other control)
- \( Y = \) clinical endpoint (1= event by \( \tau_1 \), 0= event-free at \( \tau_1 \))
- \( S = \) candidate surrogate measured at time \( \tau < \tau_1 \)

**Vaccine Efficacy Parameter**

\[
VE = 1 - \frac{P(Y = 1|A = 1)}{P(Y = 1|A = 0)}
\]

= Percent reduction in endpoint rate by vaccination vs. control group
Distinct “Surrogate” Frameworks (Price et al. Suppl.)

1. **Prentice (1989) definition (Valid replacement endpoint)**
   - Reliable inferences or predictions of $VE$ based on $S$ without using $Y$
Distinct “Surrogate” Frameworks (Price et al. Suppl.)

1. Prentice (1989) definition (Valid replacement endpoint)
   - Reliable inferences or predictions of \( VE \) based on \( S \) without using \( Y \)

2. Controlled direct and indirect effects (Design new interventions)
   - Predict \( VE \) for a new vaccine that sets \( S \) to certain levels
Distinct “Surrogate” Frameworks (Price et al. Suppl.)

1. Prentice (1989) definition (Valid replacement endpoint)
   - Reliable inferences or predictions of VE based on S without using Y

2. Controlled direct and indirect effects (Design new interventions)
   - Predict VE for a new vaccine that sets S to certain levels

3. Natural direct and indirect effects (Mediation)
   - Insights into mechanisms/pathways of vaccine protection
Distinct “Surrogate” Frameworks (Price et al. Suppl.)

1. **Prentice (1989) definition (Valid replacement endpoint)**
   - Reliable inferences or predictions of $VE$ based on $S$ without using $Y$

2. **Controlled direct and indirect effects (Design new interventions)**
   - Predict $VE$ for a new vaccine that sets $S$ to certain levels

3. **Natural direct and indirect effects (Mediation)**
   - Insights into mechanisms/pathways of vaccine protection

4. **$VE$ curve (Response type effect modification)**
   - Inferences on how $VE$ varies over subgroups of vaccine recipients defined by their response $S$ to vaccination
Distinct “Surrogate” Frameworks (Price et al. Suppl.)

1. Prentice (1989) definition (Valid replacement endpoint)
   - Reliable inferences or predictions of VE based on S without using Y

2. Controlled direct and indirect effects (Design new interventions)
   - Predict VE for a new vaccine that sets S to certain levels

3. Natural direct and indirect effects (Mediation)
   - Insights into mechanisms/pathways of vaccine protection

4. VE curve (Response type effect modification)
   - Inferences on how VE varies over subgroups of vaccine recipients defined by their response S to vaccination

5. Meta-analysis
   - Associate causal effects on S with causal effects on Y, for inference on VE in new settings
In efficacy trials showing beneficial $VE > 0$, study selected immune response biomarkers:

1. As correlates of risk (CoRs) of the disease endpoint in the vaccine and control groups
   - Extensive methods, e.g., adjusted association (Buyse et al. 2000 *Biostatistics*)
Typical Evaluation in VE Trials of Immune Response Biomarkers as Potential Surrogate Endpoints

In efficacy trials showing beneficial $VE > 0$, study selected immune response biomarkers:

1. As correlates of risk (CoRs) of the disease endpoint in the vaccine and control groups
   - Extensive methods, e.g., adjusted association (Buyse et al. 2000 *Biostatistics*)

2. For how well they adhere to criteria that imply the Prentice definition of a valid surrogate endpoint
   - Proportion of treatment effect explained (PTE) (Freedman et al. 1992 *Stat Med*)
   - Proportion of treatment effect captured (Kobayashi and Kuroki 2014 *Stat Med*)
Typical Evaluation in VE Trials of Immune Response Biomarkers as Potential Surrogate Endpoints

In efficacy trials showing beneficial $VE > 0$, study selected immune response biomarkers:

1. As correlates of risk (CoRs) of the disease endpoint in the vaccine and control groups
   - Extensive methods, e.g., adjusted association (Buyse et al. 2000 *Biostatistics*)

2. For how well they adhere to criteria that imply the Prentice definition of a valid surrogate endpoint
   - Proportion of treatment effect explained (PTE) (Freedman et al. 1992 *Stat Med*)
   - Proportion of treatment effect captured (Kobayashi and Kuroki 2014 *Stat Med*)

3. As correlates of $VE$
   - Principal stratification modifier of $VE$ (Follmann 2006 *Biometrics*; Gilbert and Hudgens 2008 *Biometrics*)
The Approach Has Worked for Simple Univariate Surrogates for Many Great Vaccines

Many licensed vaccines have excellent $VE > 90\%$, and a single marker has been accepted by regulatory agencies as a surrogate endpoint (table from Norman Baylor, former FDA CBER director).

<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Test</th>
<th>Correlate of Protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diphtheria</td>
<td>Toxin Neutralization</td>
<td>0.01-0.1 IU/mL</td>
</tr>
<tr>
<td>Hepatitis A</td>
<td>ELISA</td>
<td>10 mIU/mL</td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>ELISA</td>
<td>10 mIU/mL</td>
</tr>
<tr>
<td>Hib Polysaccharides</td>
<td>ELISA</td>
<td>1 mcg/mL</td>
</tr>
<tr>
<td>Hib Conjugate</td>
<td>ELISA</td>
<td>0.15 mcg/mL</td>
</tr>
<tr>
<td>Influenza</td>
<td>HAI</td>
<td>1/40 dilution</td>
</tr>
<tr>
<td>Lyme</td>
<td>ELISA</td>
<td>1100 EIA U/mL</td>
</tr>
<tr>
<td>Measles</td>
<td>Microneutralization</td>
<td>120 mIU/mL</td>
</tr>
<tr>
<td>Pneumococcus</td>
<td>ELISA (Opsonophagocytosis)</td>
<td>0.20-0.35 mcg/mL (for children); 1/8 dilution</td>
</tr>
<tr>
<td>Polio</td>
<td>Serum Neutralization</td>
<td>1/4 - 1/8 dilution</td>
</tr>
<tr>
<td>Rabies</td>
<td>Serum Neutralization</td>
<td>0.5 IU/mL</td>
</tr>
<tr>
<td>Rubella</td>
<td>Immunoprecipitation</td>
<td>10-15 mIU/mL</td>
</tr>
<tr>
<td>Tetanus</td>
<td>Toxin Neutralization</td>
<td>0.1 IU/mL</td>
</tr>
<tr>
<td>Varicella</td>
<td>Serum Neutralization; gb ELISA</td>
<td>$\geq 1/64$ dilution $\geq 5$ IU/mL</td>
</tr>
</tbody>
</table>

Adapted from Plotkin S. Correlates of Vaccine Induced Immunity (Vaccines 2008:47)
However, In Our Era, More of a Multivariate Learning Paradigm May be Helpful

- Important pathogens for vaccine development (e.g., HIV, malaria, TB, influenza) have much greater genetic/antigen variability than pathogens for which there is a great vaccine
However, In Our Era, More of a Multivariate Learning Paradigm May be Helpful

- Important pathogens for vaccine development (e.g., HIV, malaria, TB, influenza) have much greater genetic/antigen variability than pathogens for which there is a great vaccine
- Vastly more immune response biomarkers are now being measured characterizing vaccine immunogenicity
  - Innate response systems vaccinology data (cellular subsets, transcriptomics, etc.) (high-dimensional)
However, In Our Era, More of a Multivariate Learning Paradigm May be Helpful

- Important pathogens for vaccine development (e.g., HIV, malaria, TB, influenza) have much greater genetic/antigen variability than pathogens for which there is a great vaccine
- Vastly more immune response biomarkers are now being measured characterizing vaccine immunogenicity
  - Innate response systems vaccinology data (cellular subsets, transcriptomics, etc.) (high-dimensional)

### Adaptive Response Data (Antibodies, T cells)

- Binding antibody (Isotype, Subclass, Frequency, Magnitude, Breadth, Specificity)
- Functional antibody (e.g., Neutralization, ADCC, Systems serology) (Frequency, Magnitude, Breadth, Specificity)
- CD4 and CD8 T cell responses (Frequency, Magnitude, Breadth, Specificity, Quality)
**Example: Search for Immunological Surrogates in the RV144 HIV-1 Vaccine Efficacy Trial**

**T Cell Correlates**
Cytokine response (IL-10, IL-13) from Env stimulated PBMC
Polyfunctional CD4+ T cell (CD40L, IL-2, IL-4, IFN-γ and TNF-α) and (CD40L, IL-2 and IL-4)

(Haynes et al. NEJM 2012; Lin et al. Nature Biotechnology 2015)

**Host Genetics and Antibodies**
IgG, IgG3, nAb, Avidity and FcγRIIC SNP
IgA/ HLA A*02 allele
IgA/ HLA II DQB1*06
IgG/ HLA II DPB1*13

(Li et al. JCI 2014; Gartland et al. JV 2014; Prentice et al. Sci. Trans Med. 2015)

**V2 Correlates**
V1V2 IgG, V1V2 IgG Breadth
V2 Linear AE hotspot
V1V2 IgG3


**IgA Correlates**
IgA Env Score
IgA A. OOMSA gp140 CF
IgA. A1 Congp140
IgA C1
IgA Non-Vaccine Strains
IgA/IgG ratio

(Haynes et al. NEJM 2012; Tomaras, Ferrari et al. PNAS 2013)

**Virus Sieve Analysis and Antibodies**
V2 Sieve (and V2 mAbs dependent on 169K)
Genetic distance from Vaccine strain /IgG and IgG3 V1V2 correlates


**Antibody Interaction Correlates**
Low IgA/ ADCC
Low IgA/ nAb
Low IgA/ IgG Env Avidity
IgG3/ ADCC
IgG3/IgG1

Principles of the “Estimated Optimal Surrogate” Approach to Developing an Immunological Surrogate

- Must be able to handle a high-dimensionality of immunological measurements, seeking to leverage all information in the data by modern computational machine learning
Principles of the “Estimated Optimal Surrogate” Approach to Developing an Immunological Surrogate

- Must be able to handle a high-dimensionality of immunological measurements, seeking to leverage all information in the data by modern computational machine learning

- Yet still embrace the virtue of simplicity
  - In the end seek a simple univariate surrogate that is a synthesis and encapsulation of all of the information
Principles of the “Estimated Optimal Surrogate” Approach to Developing an Immunological Surrogate

- Must be able to handle a high-dimensionality of immunological measurements, seeking to leverage all information in the data by modern computational machine learning

- Yet still embrace the virtue of simplicity
  - In the end seek a simple univariate surrogate that is a synthesis and encapsulation of all of the information

- Make this simple surrogate clinically interpretable in terms of vaccine efficacy
Principles of the “Estimated Optimal Surrogate” Approach to Developing an Immunological Surrogate

- Must be able to handle a high-dimensionality of immunological measurements, seeking to leverage all information in the data by modern computational machine learning.

- Yet still embrace the virtue of simplicity
  - In the end seek a simple univariate surrogate that is a synthesis and encapsulation of all of the information.

- Make this simple surrogate clinically interpretable in terms of vaccine efficacy.

- Set up the approach such that the excellent Prentice definition of a valid surrogate endpoint holds by construction.

\[ VE = 1 - \frac{P(Y = 1|A = 1)}{P(Y = 1|A = 0)} \]

**Definition**

S is a valid surrogate endpoint for Y if a valid test of

\[ H_0^Y : \text{No vaccine effect on } Y \text{ (i.e., } VE = 0) \]

is obtained by testing

\[ H_0^S : \text{No vaccine effect on } S \]
Illustration of Prentice Definition

Corresponds to VE=0

Corresponds to VE>0
The Surrogate Paradox

**Virtue of the Prentice definition: Guarantees the surrogate paradox cannot occur**

**Surrogate Paradox**

- Positive vaccine effect on $S$
  - i.e., immune responses higher in vaccine than control group
- $S$ and $Y$ are inversely correlated in both the vaccine and control groups
  - i.e., in each group a higher immune response is associated with a lower disease rate
- Yet $VE < 0$ (*a harmful vaccine!*)

Example of the Surrogate Paradox

**Sweden I Acellular Pertussis Trial of SKB and Aventis Pasteur vaccines vs. DT control arm \( (N \approx 10,000) \)**

- Immune response biomarkers \( S = \) Filamentous Haemagglutinin (FHA) and Pertussis Toxoid (PT) antibody responses higher for SKB than Aventis Pasteur vaccine
- Higher FHA and PT antibodies associated with lower pertussis disease rates

\* Gustafsson et al. (1996, NEJM); Fleming and Powers (2012, Stat Med)
Example of the Surrogate Paradox

Sweden I Acellular Pertussis Trial of SKB and Aventis Pasteur vaccines vs. DT control arm (N ≈ 10,000)*

- Immune response biomarkers $S =$ Filamentous Haemagglutinin (FHA) and Pertussis Toxoid (PT) antibody responses higher for SKB than Aventis Pasteur vaccine
- Higher FHA and PT antibodies associated with lower pertussis disease rates
- Yet estimated VE greater for the Aventis Pasteur vaccine: 85% (95% CI 81–89%) vs. 58% (95% CI 51–66%)

*Gustafsson et al. (1996, NEJM); Fleming and Powers (2012, Stat Med)
## Example of the Surrogate Paradox

### Sweden I Acellular Pertussis Trial of SKB and Aventis Pasteur vaccines vs. DT control arm ($N \approx 10,000$)*

- Immune response biomarkers $S =$ Filamentous Haemagglutinin (FHA) and Pertussis Toxoid (PT) antibody responses higher for SKB than Aventis Pasteur vaccine.
- Higher FHA and PT antibodies associated with lower pertussis disease rates.
- Yet estimated $VE$ greater for the Aventis Pasteur vaccine: 85% (95% CI 81–89%) vs. 58% (95% CI 51–66%).

**Possible explanation:** The Aventis Pasteur vaccine had additional antigens – Pertactin and Fimbriae types 2 and 3 – which stimulated additional immune responses contributing to protection not measured by the FHA and PT assays.

1 Surrogate Endpoint Frameworks

2 Optimal Surrogate Framework

3 Simulation Studies

4 Application to Dengue Trials

5 Discussion
Introduction to an Optimal Surrogate

Data from a VE Trial for Developing a Surrogate

- \( W = \) Baseline covariates
- \( A = \) Randomized treatment assignment
  \((1=\text{vaccine}, \ 0=\text{placebo})\)
- \( S = \) Immune response biomarkers measured by an intermediate time point \( \tau \) (e.g., 2 weeks post last vaccination)
- \( Y = \) Disease endpoint by the end of follow-up \( \tau_1 \) after \( \tau \)

**Goal:** Develop a most-promising surrogate endpoint for the disease endpoint so that future randomized studies can restrict themselves to only collecting the surrogate outcome
Optimal Surrogate $\equiv$ Valid Surrogate that Optimally Predicts $Y$

- Define an **optimal surrogate** as the function of $(W, A, S)$ that satisfies the Prentice definition and that optimally predicts $Y$
  - A true (unknown) parameter that is estimated
Define an **optimal surrogate** as the function of $(W, A, S)$ that satisfies the Prentice definition and that optimally predicts $Y$

- A true (unknown) parameter that is estimated

**Goal 1:** Obtain an efficient and robust estimate of the optimal surrogate based on the randomized efficacy trial
Optimal Surrogate = Valid Surrogate that Optimally Predicts \( Y \)

- Define an **optimal surrogate** as the function of \((W, A, S)\) that satisfies the Prentice definition and that optimally predicts \( Y \)
  - A true (unknown) parameter that is estimated

- **Goal 1:** Obtain an efficient and robust estimate of the optimal surrogate based on the randomized efficacy trial

- **Goal 2:** Use the estimated optimal surrogate built for Goal 1 in **future clinical trials** for estimation and testing of \( VE \) (treatment effect on \( Y \))
  - Tackles the **bridging objective** of inferring the causal treatment effect \( VE \) in a new trial without measuring \( Y \)
  - (also addressed by Pearl and Bareinboim, 2011, 2012)
This work is about the search for promising surrogates based on an efficacy trial(s) with \((W, A, S, Y)\) measured.

A promising surrogate is one that satisfies the Prentice definition and is optimally predictive of \(Y\) in this original trial(s).

This is a good starting point for building a surrogate that is promising for the ultimate objective of bridging – inference on \(VE\) in new settings based on \((W, A, S)\).
Applications of the Estimated Optimal Surrogate Approach to Immunological Surrogate Development

1. A given immune response biomarker is thought to provide a sufficiently valid surrogate endpoint, but it is unclear how to optimally define the readout.
   - E.g., the CYD14 and CYD15 dengue phase 3 VE trials studied PRNT\textsubscript{50}, a single estimated summary measure from a statistical model fit to a neutralization dilution series curve.
   - Is there a better surrogate based on a different feature of the curve? Would an alternative neutralization assay do better (e.g., Microneutralization Version 2)?
Applications of the Estimated Optimal Surrogate Approach to Immunological Surrogate Development

1. A given immune response biomarker is thought to provide a sufficiently valid surrogate endpoint, but it is unclear how to optimally define the readout
   - E.g., the CYD14 and CYD15 dengue phase 3 VE trials studied PRNT$_{50}$, a single estimated summary measure from a statistical model fit to a neutralization dilution series curve
   - Is there a better surrogate based on a different feature of the curve? Would an alternative neutralization assay do better (e.g., Microneutralization Version 2)?

2. Additional assays are applied measuring new immune response features (e.g., Fc effector function assays, T cell assays, innate immunity assays) and we ask whether an improved surrogate can be developed by adding one or more assays?
   - E.g., in RV144, the original anti-V2 antibody correlate of risk was improved by adding ADCC and CD4 T cell polyfunctionality (Haynes et al., 2012, *NEJM*; Lin et al., 2015, *Nat Biotech*)
3 At the outset of a correlates study a set of (possibly high-dimensional) immune response biomarkers are measured, and we wish to develop best surrogates based on this set.

- Currently planning such an analysis for the first TB vaccine infection endpoint efficacy trial (Nemes et al., 2018, *NEJM*).
3 At the outset of a correlates study a set of (possibly high-dimensional) immune response biomarkers are measured, and we wish to develop best surrogates based on this set. Currently planning such an analysis for the first TB vaccine infection endpoint efficacy trial (Nemes et al., 2018, *NEJM*).

4 Immune response assays are measured at multiple time points (e.g., baseline and post vaccinations, possibly longitudinally), and we wish to study whether a surrogate can be improved by including multiple time points. E.g., in the dengue trials, accounting for both baseline (pre-existing immunity) and post-vaccination readouts is evidently important (Moodie et al., 2018, *J Infect Dis*; Sridhar et al., 2018, *NEJM*)
Applications of the Estimated Optimal Surrogate Approach to Immunological Surrogate Development, Continued

3 At the outset of a correlates study a set of (possibly high-dimensional) immune response biomarkers are measured, and we wish to develop best surrogates based on this set
   - Currently planning such an analysis for the first TB vaccine infection endpoint efficacy trial (Nemes et al., 2018, *NEJM*)

4 Immune response assays are measured at multiple time points (e.g., baseline and post vaccinations, possibly longitudinally), and we wish to study whether a surrogate can be improved by including multiple time points
   - E.g., in the dengue trials, accounting for both baseline (pre-existing immunity) and post-vaccination readouts is evidently important (Moodie et al., 2018, *J Infect Dis*; Sridhar et al., 2018, *NEJM*)

For each application, a principled framework is needed for estimating optimal surrogates and for comparing the performance of different estimators
Statistical Formulation of an Optimal Surrogate

Data from a VE Trial for Developing a Surrogate

- $W = \text{Baseline covariates}$
- $A = \text{Randomized treatment assignment}$
  $(1=\text{vaccine}, 0=\text{placebo})$
- $S = \text{Immune response biomarkers measured by an}$
  $\text{intermediate time point } \tau$ (e.g., 2 weeks post last vaccination)
- $Y = \text{Disease endpoint by the end of follow-up } \tau_1 \text{ after } \tau$

Case-cohort or case-control sampling design, where $S$ (and perhaps components of $W$) is measured in a subset of study participants
Historically, evaluating surrogates has relied on correctly specified regression models linking disease risk to input covariates ($A$, $W$, $S$)

- E.g., logistic regression or Cox regression
- Mis-specified models leads to biased estimation and potentially misleading results about surrogate endpoints
A Nonparametric, Robust Approach

- Historically, evaluating surrogates has relied on correctly specified regression models linking disease risk to input covariates \((A, W, S)\)
  - E.g., logistic regression or Cox regression
  - Mis-specified models leads to biased estimation and potentially misleading results about surrogate endpoints

- This nonparametric approach avoids assumptions on the distribution of \(W\) or on the conditional distribution of \((S, Y)\) given \(A, W\), and thus is more robust
Any real-valued function \((W, A, S) \rightarrow \psi(W, A, S)\) is a \textbf{candidate surrogate}, representing a measurement one can collect by time \(\tau\) and depending on the unknown true observed data distribution \(P_0\).
Candidate Surrogate Outcomes (True Unknown Parameters)

- Any real-valued function \((W, A, S) \rightarrow \psi(W, A, S)\) is a **candidate surrogate**, representing a measurement one can collect by time \(\tau\) and depending on the unknown true observed data distribution \(P_0\)

- **Question:** How to define a good surrogate in terms of \(P_0\)?
Candidate Surrogate Outcomes (True Unknown Parameters)

- Any real-valued function \((W, A, S) \rightarrow \psi(W, A, S)\) is a **candidate surrogate**, representing a measurement one can collect by time \(\tau\) and depending on the unknown true observed data distribution \(P_0\).

**Question:** How to define a good surrogate in terms of \(P_0\)?

**Starting point:** Only consider \(S^\psi \equiv \psi(W, A, S)\) that are valid in the efficacy study, according to the Prentice definition:

\[
VE = 0 \iff \text{Vaccine effect on the mean of } S^\psi = 0
\]
Optimal Surrogate Outcome

- Criterion for ranking valid surrogates and defining a \( P_0 \)-optimal surrogate:
  - Mean squared error \( MSE(\psi) \)
  - Summarizes how close the outcome values \( Y_i \) are to the surrogate outcome values \( \psi(W_i, A_i, S_i) \)

- \( P_0 \)-optimal surrogate = the function \( \psi \) of \( (W, A, S) \) that minimizes \( MSE(\psi) \) subject to the Prentice definition constraint
Result 1

The minimizer of $\psi \rightarrow MSE(\psi)$ over all functions $(W, A, S) \rightarrow \psi(W, A, S)$ that satisfy the Prentice definition is the conditional disease risk:

$$\tilde{S}_0 = \psi_0(W, A, S) \equiv P_0(Y = 1 \mid W, A, S)$$
Result 1

The minimizer of $\psi \rightarrow MSE(\psi)$ over all functions $(W, A, S) \rightarrow \psi(W, A, S)$ that satisfy the Prentice definition is the conditional disease risk:

$$\tilde{S}_0 = \psi_0(W, A, S) \equiv P_0(Y = 1 \mid W, A, S)$$

Advantageous Implication:

- The vaccine effect on the optimal surrogate,

$$VE(\tilde{S}_0) = 1 - \frac{\text{Mean of } \tilde{S}_0 \text{ Vaccine Group}}{\text{Mean of } \tilde{S}_0 \text{ Placebo Group}}$$

has the same scale of interpretation as

$$VE = 1 - \frac{\text{Mean of } Y \text{ Vaccine Group}}{\text{Mean of } Y \text{ Placebo Group}}$$
Estimation of the $P_0$-optimal Surrogate

In practice, of course, the $P_0$-optimal surrogate $P_0(Y = 1 \mid W, A, S)$ is not available for use.
Estimation of the $P_0$-optimal Surrogate

In practice, of course, the $P_0$-optimal surrogate $P_0(Y = 1 \mid W, A, S)$ is not available for use.

It is estimated and the estimated regression function $\hat{P}_0(Y = 1 \mid W, A, S)$ is used as the surrogate.

- i.e., individual $i$ with covariates $(W_i, A_i, S_i)$ has surrogate endpoint value $\hat{P}_0(Y_i = 1 \mid |W_i, A_i, S_i)$.
Estimation of the $P_0$-optimal Surrogate

- **Objective:** Estimate the regression function
  
  $P_0(Y = 1 \mid W, A, S)$ – a standard prediction problem
Estimation of the $P_0$-optimal Surrogate

- **Objective**: Estimate the regression function $P_0(Y = 1 \mid W, A, S)$ – a standard prediction problem

- **Challenge**: A very large number of estimators are possible – How to achieve a best estimator?
  - i.e., how to optimally make the bias-variance tradeoff?
How to Best Estimate $P_0(Y = 1 \mid W, A, S)$?

Different regression methods tradeoff bias and variance in different ways

- **Nonparametric**: Empirical moment, kernel regression, neural networks, random forests
- **Semiparametric**: Generalized additive models, partially linear additive models
- **Parametric**: Logistic regression, spline regression
How to Best Estimate $P_0(Y = 1 | W, A, S)$?

- Different regression methods tradeoff bias and variance in different ways
  - **Nonparametric**: Empirical moment, kernel regression, neural networks, random forests
  - **Semiparametric**: Generalized additive models, partially linear additive models
  - **Parametric**: Logistic regression, spline regression

- For a given regression method, the tradeoff is governed by modeling choices and/or tuning parameters
  - Logistic regression with two immune response biomakers (include an interaction term?)
  - Uniform kernel estimator (large or small smoothing bandwidth?)
  - Regression tree (maximum depth one versus thirty?)

The best bias/variance tradeoff depends on the (unknown) true regression function
Super-Learner Estimator of $P_0(Y = 1 \mid |W, A, S)$

1. Specify a large library of regression methods/estimators for $P_0(Y = 1 \mid |W, A, S)$
2. Use a fair prediction performance criterion to compare all the estimators
3. Select the best estimator by this criterion
   - called the **Discrete Super Learner**
4. Also select the best combination estimator that is the best weighted average of all of the individual estimators
   - called the **Super Learner**
Fair Criterion: Cross-Validated Prediction Performance

Divide data into $V$ sets of size $\approx \frac{n}{V}$ (Here, $V = 10$)

Fold 1 = training sample $T_1 + \text{validation sample } V_1$

- Training sample is used to fit (“train”) the regressions
- Validation sample is used to estimate prediction performance (“validate”)
Fair Criterion: Cross-Validated Prediction Performance

Divide data into $V$ sets of size $\approx \frac{n}{V}$ (Here, $V = 10$)

Fold 1 = training sample $T_1$ + validation sample $V_1$

- Training sample is used to fit (“train”) the regressions
- Validation sample is used to estimate prediction performance (“validate”)

Several factors to consider when choosing $V$:

- Large $V$ = more data to fit regressions (helpful in small data sets or with high-dimensional covariates)
- Small $V$ = more data to evaluate prediction performance
- Large $V$ = greater computation time
Fair Criterion: Cross-Validated Prediction Performance

The validation set rotates until each set has been used as validation once.

<table>
<thead>
<tr>
<th>Fold 1</th>
<th>Fold 2</th>
<th>Fold 3</th>
<th>Fold 4</th>
<th>Fold 5</th>
<th>Fold 6</th>
<th>Fold 7</th>
<th>Fold 8</th>
<th>Fold 9</th>
<th>Fold 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>
Measure Cross-Validated Prediction Performance by Cross-Validated Risk

**Risk** = average loss of an estimator, where the loss scores how far away the prediction of \( Y \) made from the estimator \( \hat{P}_0(Y = 1|A, W, S) \) is from the true \( Y \)

- E.g., squared error loss \((Y_i - \hat{P}_0(Y_i = 1|W_i, A_i, S_i))^2\)
Measure Cross-Validated Prediction Performance by Cross-Validated Risk

**Risk** = **average loss** of an estimator, where the loss scores how far away the prediction of $Y$ made from the estimator $\hat{P}_0(Y = 1|A, W, S)$ is from the true $Y$

- E.g., squared error loss $(Y_i - \hat{P}_0(Y_i = 1|W_i, A_i, S_i))^2$

**Cross-Validated Risk**

1. Build the model from training set $\mathcal{T}_1$; estimate risk on validation set $\mathcal{V}_1$
2. Build the model from training set $\mathcal{T}_2$; estimate risk on validation set $\mathcal{V}_2$
3. 
4. 
5. 
6. 
7. 
8. 
9. 
10. Build the model from training set $\mathcal{T}_{10}$; estimate risk on validation set $\mathcal{V}_{10}$

**Cross-validated risk** = **average of the 10 validation set risks**
Super Learner Based on Cross-Validated Risk

- **Discrete Super Learner** is the estimator \( \hat{P}_0(Y = 1|W, A, S) \) with the smallest cross-validated risk

- **Super Learner** is the weighted average of all of the estimators \( \hat{P}_0(Y = 1|W, A, S) \) with the smallest cross-validated risk

  - Idea originated with “model stacking” of Wolpert (1992) and Breiman (1996)
  - Idea generalized and re-branded as “super learning” (van der Laan, Polley, and Hubbard, 2007)
Strong Practical Performance of Super-Learner*

Super-Learner with Cross-Validated Classification Accuracy Metrics*: A Framework for Comparing Estimated Optimal Surrogates and Seeking Parsimonious Surrogates

* Van der Laan, Hubbard, and Pajouh (2013)
Advantageous Properties of Super Learner

- **Oracle Property:** It has predictive performance risk *very close* to the oracle estimator that uses the true (unknown) best estimator
Advantageous Properties of Super Learner

- **Oracle Property:** It has predictive performance risk *very close* to the oracle estimator that uses the true (unknown) best estimator.

- **Flexibility:** The number of estimators is allowed to be very large – and including a large number of estimators in the library of learners aids achieving the oracle property.
Advantageous Properties of Super Learner

- **Oracle Property:** It has predictive performance risk *very close* to the oracle estimator that uses the true (unknown) best estimator.

- **Flexibility:** The number of estimators is allowed to be very large – and including a large number of estimators in the library of learners aids achieving the oracle property.

- Any given regression method can be used to construct multiple different estimators, e.g.:
  - Random forest with different tuning parameters
  - Generalized additive models with different knots and degrees
  - Logistic regression with interactions and stepwise selection
Advantageous Properties of Super Learner

- Traditional practice tries several models and checks model fit to select a model
  - This exploration practice without pre-specification invalidates inferences
Advantageous Properties of Super Learner

- Traditional practice tries several models and checks model fit to select a model
  - This exploration practice without pre-specification invalidates inferences

- In contrast, Super Learner is pre-specified
  - Eliminates the need to put all our eggs in a single estimation basket
  - Include in the library any model choice that could result from model checking
  - Oracle property ensures that Super Learner is good at choosing (approximately) the correct one
Advantageous Properties of Super Learner

- Traditional practice tries several models and checks model fit to select a model
  - This exploration practice without pre-specification invalidates inferences

- In contrast, Super Learner is pre-specified
  - Eliminates the need to put all our eggs in a single estimation basket
  - Include in the library any model choice that could result from model checking
  - Oracle property ensures that Super Learner is good at choosing (approximately) the correct one

- A major scientific activity is selection of the library of estimators
Super Learner-Based Estimated Optimal Surrogate

Estimated optimal surrogate (EOS): \( \hat{S}_0 = \hat{P}_0(Y \mid W, A, S) \)

Vaccine effect on the EOS \( \hat{S}_0 \):

\[
VE(\hat{S}_0) = 1 - \frac{\text{Mean of } \hat{S}_0 \text{ Vaccine Group}}{\text{Mean of } \hat{S}_0 \text{ Placebo Group}}
\]

Vaccine effect on the disease endpoint \( Y \):

\[
VE_0 = 1 - \frac{\text{Mean of } Y \text{ Vaccine Group}}{\text{Mean of } Y \text{ Placebo Group}}
\]

- Price, Gilbert, and van der Laan (2018) showed how to estimate \( VE(\hat{S}_0) \) with a confidence interval
- They showed that a TMLE-adjusted Super Learner estimator is an asymptotically efficient estimator of \( VE \)
  - A desirable property of a best surrogate built from an efficacy trial
Special Application of the EOS if Prentice’s (1989) “Full Mediation Condition” Holds

The key Prentice criterion for a surrogate endpoint to be valid is: **Within each subgroup defined by \((W, S)\), disease risk is the same in the vaccine and placebo groups**

\[
P(Y = 1|W = w, A = 1, S = s) = P(Y = 1|W = w, A = 0, S = s)
\]

- Typically fails, but may hold if the surrogate is tightly linked to a mechanism of protection, that operates the same for vaccine immunity vs. natural immunity
Example: Dunning et al. (2016, Clin Vacc Immun)

- **VE** trial of Sanofi’s Inactivated Influenza Vaccine High vs. Standard Dose in $\geq 65$ year-olds
- Correlates of protection analysis of antibodies (HAI, NAI, NT – neutralization test)
- Article concluded that combining assays improved surrogate quality

![Protection curves for the A/Victoria/361/2011 HAI assay using the circulating virus against A/H3N2 illness by three laboratory-confirmed influenza (LCI) case definitions (defn.), showing titers for 50% and 80% protection, with 95% CIs.](image_url)


- Could repeat with the EOS $\hat{P}_0(W, S)$ on the x-axis using HAI, NAI, NT
Goal 2: Bridging/Transportability

- **Goal 2:** Use the $P_0$-estimated optimal surrogate built from a previous efficacy trial(s) as the primary study endpoint in a future clinical trial, for inference on VE without measuring $Y$

  - Surrogate endpoint $\bar{S}_{0i} = \text{Super Learner } \hat{P}_0(Y_i = 1 \mid W_i, A_i, S_i)$
Goal 2: Bridging/Transportability

- **Goal 2**: Use the $P_0$-estimated optimal surrogate built from a previous efficacy trial(s) as the primary study endpoint in a future clinical trial, for inference on $VE$ without measuring $Y$
  - Surrogate endpoint $\bar{S}_{0i} = \text{Super Learner } \hat{P}_0(Y_i = 1 \mid W_i, A_i, S_i)$

- This bridging problem is hard given the implicit necessity of extrapolating beyond the empirical data

- We give (strong) conditions under which this bridging inference may be done in a valid way
Assumptions Under which the $P_0$-Optimal Surrogate Can be Used for Valid Estimation of $VE^*$ in the New Study

**Theorem 2 from Price et al. (2018)**

Consider a new randomized study with collected data $(W_i^*, A_i^*, S_i^*), i = 1, \cdots, n^*$

The Following Assumptions Guarantee Correct Bridging:

- **Equal Conditional Disease Risk:** Within each subgroup defined by $(W^*, A^*, S^*)$, disease risk is the same in the original and new studies
- **Contained Support:** All of the subgroups defined by $(W^*, A^*, S^*)$ are represented in the original study
- **Positivity:** All subgroups defined by $W^*$ are represented in both the vaccine and placebo groups $A^* = 1$ and $A^* = 0$
1 Surrogate Endpoint Frameworks

2 Optimal Surrogate Framework

3 Simulation Studies

4 Application to Dengue Trials

5 Discussion
Two Simulation Studies

- **Objective of First Study:** Simple illustration that the estimated optimal surrogate will always provide unbiased estimation of $VE_0 = P_0(Y_1 - Y_0)$ in the original trial, for any distribution of $(W, A, S, Y)$

- **Objective of Second Study:** Illustrate how well the estimated optimal surrogate built from one trial works for inference on $VE^* = EP(Y_1^* - Y_0^*)$ in a second trial, when Equal Conditional Disease Risk fails
Building upon an example published by VanderWeele (2014, *Biometrics*)

- Continuous outcome $Y$
- Treatment $A \in \{0, 1\}$
- 10 candidate surrogates $S^i$ ($S^i \in \{0, 1, 2\}$, $i = 1 \ldots 10$)

- $P(S^i_1 = 0, S^i_0 = 0) = P(S^i_1 = 1, S^i_0 = 1) = P(S^i_1 = 2, S^i_0 = 2) = 0.1$,
- $P(S^i_1 = 1, S^i_0 = 0) = 0.5$,
- $P(S^i_1 = 1, S^i_0 = 2) = 0.2$

$$Y = \sum_{i=1}^{3} [0.1 \times i \times I_{S^i_1 \neq 1} + I_{S^i_1 \neq 2}] + \epsilon_Y, \ \epsilon_Y \sim N(0, 0.1^2)$$
The Surrogate Paradox Occurs

1. $S \rightarrow Y$
   POSITIVE relationship between surrogates and outcome
   $Y = \sum_{i=1}^{3} [0.1 \times i \times I_{S_i=1} + I_{S_i=2}] + \epsilon_Y$, $\epsilon_Y \sim N(0, 0.1^2)$;

2. $A \rightarrow S$
   POSITIVE treatment effect on surrogates
   $E[S_i^1 - S_i^0] = 0.3$;

3. $A \rightarrow Y$
   NEGATIVE overall treatment effect
   $E[Y_1 - Y_0] = -0.18$
Simulation 1: Compare the TMLE-SL Estimator of \( VE_0 \) to a Standard Estimator

**Standard estimator of \( VE_0 \):** Simple regression estimator after selection of the surrogate based on the Proportion of the Treatment Effect Captured (PCS) by the candidate Surrogate*

- For each \( S^i \), estimate PCS nonparametrically

\[
PCS = \frac{CP^2}{CP^2 + NCP^2}
\]

(true PCS = 0.87, 0.2, 0.002 for \( i = 1, 2, 3 \); PCS = 0 for \( i = 4, \ldots, 10 \))

- Select “best surrogate”: \( S^{PCS_{opt}} = S^i \) with the greatest \( \hat{PCS} \)

- Estimate \( VE_0 \) by the difference \((a = 1 \) minus \( a = 0\)) in average predicted \( Y \)'s \( \hat{P}(Y_i = 1 \mid S^{PCS_{opt}}_i, A_i = a) \)

Simulation 1: Estimation under the Surrogate Paradox
($n = 2000$ Subjects; 200 Simulated Data Sets)

- **Surrogate Paradox**
  - Occurs for 95% of PCS* method estimates: $\hat{VE}_0 > 0$ (vs. truth $VE_0 = -0.18$)
  - Does not occur with TMLE-SL method: $\hat{VE}_0 < 0$

*Proportion of treatment effect captured (PCS) (Kobayashi and Kuroki, 2014, Stat Med)
Simulation 2: Transportability When Equal Conditional Disease Risks Fails

- Both the PCS and the TMLE-SL methods are biased.
- Surrogate Paradox
  - Occurs for 95% of PCS method estimates: \( \hat{VE}^* > 0 \) (vs. truth \( VE^* = -0.10 \))
  - Does not occur with the TMLE-SL method: \( \hat{VE}^* < 0 \)
Conclusion from Simulation 2

- Demonstrates that the Equal Conditional Disease Risks assumption is necessary for valid inference about $VE^*$ in a new setting.

- When Equal Conditional Disease Risks is majorly violated, the estimated optimal surrogate can still preserve some accuracy in bridging the clinical treatment effect to a new setting.
1. Surrogate Endpoint Frameworks

2. Optimal Surrogate Framework

3. Simulation Studies

4. Application to Dengue Trials

5. Discussion
Dengue Phase 3 Trial Example

- Two randomized, double-blinded, placebo-controlled, multicenter, Phase 3 trials of a recombinant, live, attenuated, tetravalent (4 serotypes) dengue vaccine (CYD-TDV)
  - **CYD14**: Asia-Pacific region, 2–14 year-olds (Capeding et al., 2014, *The Lancet*)
  - **CYD15**: Latin America, 9–16 year-olds (Villar et al., 2015, *NEJM*)

**Trial Designs**

- 2:1 randomization to vaccine:placebo
- Immunizations at months 0, 6, 12
- Primary follow-up from Month 13 to Month 25 (active phase of follow-up)
- Primary endpoint: Symptomatic, virologically confirmed dengue (VCD)
Results on Vaccine Efficacy to Prevent VCD from Month 13 to 25 (Proportional Hazards Models)

**CYD14:** $\hat{VE} = 56.5\%$ (95% CI 43.8–66.4)

$N = 10,275, n = 244$ endpoints $Y = 1$

**CYD15:** $\hat{VE} = 64.7\%$ (95% CI 58.7–69.8)

$N^* = 20,869, n^* = 415$ endpoints $Y^* = 1$
Correlates of Risk and Correlates of VE Study in CYD14 and CYD15 (Moodie et al., 2018, JID)

Neutralizing Antibody Correlates Analysis of Tetravalent Dengue Vaccine Efficacy Trials in Asia and Latin America

Zoe Moodie, Michal Juraska, Ying Huang, Yingying Zhuang, Youyi Fong, Lindsay N. Carpp, Steven G. Self, Laurent Chambonneau, Robert Small, Nicholas Jackson, Fernando Noriega, and Peter B. Gilbert

Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; Department of Biostatistics, University of Washington, Seattle; Sanofi Pasteur, Marcy-L’Etoile, France; Sanofi Pasteur, Swiftwater, Pennsylvania; Sanofi Pasteur, Lyon, France

Background. In the CYD14 and CYD15 Phase 3 trials of the CYD-TDV dengue vaccine, estimated vaccine efficacy (VE) against symptomatic, virologically confirmed dengue (VCD) occurring between months 13 and 25 was 56.5% and 60.8%, respectively.

Methods. Neutralizing antibody titers to the 4 dengue serotypes in the CYD-TDV vaccine insert were measured at month 13 in a randomly sampled immunogenicity subcohort and in all VCD cases through month 25 (2848 vaccine, 1574 placebo) and studied for their association with VCD and with the level of VE to prevent VCD.

Results. For each trial and serotype, vaccinees with higher month 13 titer to the serotype had significantly lower risk of VCD with that serotype (hazard ratios, 0.19–0.43 per 10-fold increase). Moreover, for each trial, vaccinees with higher month 13 average titer to the 4 serotypes had significantly higher VE against VCD of any serotype ($P < .001$).

Conclusions. Neutralizing antibody titers postdose 3 correlate with CYD-TDV VE to prevent dengue. High titers associate with high VE for all serotypes, baseline serostatus groups, age groups, and both trials. However, lowest titers do not fully correspond to zero VE, indicating that other factors influence VE.

Keywords: case cohort; immune correlate of protection; neutralizing antibodies; surrogate endpoint; vaccine efficacy trial.
Month 13 PRNT$_{50}$ and Microneutralization V2 neutralization levels measured from a case-cohort sample

- **Cases:** All symptomatic VCD cases between Month 13 and 25 (n=244 CYD14; n=415 CYD15)

- **Controls:** All in the immunogenicity subset free of the VCD endpoint at Month 25 (n=1879 CYD14; n=1884 CYD15)
Application of the Estimated Optimal Surrogate Approach to CYD14 and CYD15

Example in Price et al. (2018)

- Treat CYD14 as the current trial; CYD15 as the future trial

Notation and Variables

- $W =$ Baseline covariates: age, sex, country-specific fractions of VCD endpoints of each specific serotype
- $A =$ Vaccination status (1=vaccine; 0=placebo)
- $S =$ Month 13 PRNT$_{50}$ and Microneutralization Version 2 neutralization titers to the 4 vaccine strains (serotypes 1–4), average, min, max
- $Y =$ Disease outcome (1=VCD endpoint between Month 13 and 25; 0 = no VCD endpoint by Month 25)
Month 13 PRNT\textsubscript{50} Titer Data \(S\) by Levels of \(W\) (Sex, Age) and \(A\) (Vaccine, Placebo): CYD14

![Month 13 PRNT\textsubscript{50} Titer Data](image)
Month 13 PRNT$_{50}$ Average Titers*: CYD14 and CYD15

*Average nAb titer = Geometric mean PRNT50 to the 4 dengue viruses in the vaccine construct
Month 13 PRNT$_{50}$ Average Titers a Correlate of Risk in CYD14 (Moodie et al. 2017)

Risk of DENV–Any Decreases with Month 13 Average nAb Titer in CYD14

**Vaccine**

- Low: ≤ 58
- Med: 58–266
- High: > 266

**Placebo**

- Low: ≤ 58
- Med: 58–266
- High: > 266

---

**Cumulative DENV–Any Rate**

- **Vaccine**
  - Low
  - Medium
  - High

- **Placebo**
  - Low
  - Medium
  - High

**p.val.HR**

- Vaccine: < 0.001
- Placebo: 0.041

**No. at risk**

- **Vaccine**
  - Low: 236, 229, 226, 224, 221, 206, 78
  - Med: 618, 609, 598, 593, 582, 570, 246
  - High: 536, 535, 531, 528, 524, 521, 257

- **Placebo**
  - Low: 407, 395, 384, 376, 362, 344, 143
  - High: 135, 134, 130, 129, 127, 123, 50

**Cumulative No. of DENV–Any Endpoints**

- **Vaccine**
  - Low: 0, 7, 10, 12, 15, 30, 39
  - Med: 0, 9, 19, 24, 35, 47, 60
  - High: 0, 1, 5, 6, 10, 13, 16

- **Placebo**
  - Low: 0, 12, 23, 31, 45, 63, 80
  - Med: 0, 4, 8, 10, 15, 24, 34
  - High: 0, 1, 5, 6, 8, 12, 15
Month 13 PRNT$_{50}$ Average Titers a Correlate of Risk in CYD15 (Moodie et al. 2017)

**Vaccine**

<table>
<thead>
<tr>
<th>PRNT$_{50}$ Categories:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low: ≤ 135</td>
</tr>
<tr>
<td>Med: 135–631</td>
</tr>
<tr>
<td>High: &gt; 631</td>
</tr>
</tbody>
</table>

**Placebo**

<table>
<thead>
<tr>
<th>PRNT$_{50}$ Categories:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low: ≤ 135</td>
</tr>
<tr>
<td>Med: 135–631</td>
</tr>
<tr>
<td>High: &gt; 631</td>
</tr>
</tbody>
</table>

---

**Figure Details**

- **PRNT$_{50}$**:
  - Low: ≤ 135
  - Med: 135–631
  - High: > 631

- **Cumulative DENV–Any Endpoint Rates**:
  - Vaccine:
    - Low: 347, 328, 306, 292, 276, 64
    - Med: 566, 553, 536, 526, 517, 146
    - High: 517, 511, 510, 508, 506, 123
  - Placebo:
    - Low: 471, 434, 395, 361, 334, 312, 68
    - Med: 289, 283, 272, 263, 257, 252, 76
    - High: 81, 81, 80, 79, 79, 11

- **Months Since Month 13 Visit**:
  - 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
First Application: Inference on $VE_0$ in CYD14

1. Obtain the TMLE-adjusted EOS from the CYD14 data $(W_i, A_i, S_i, Y_i), \ i = 1, \cdots, n$
   - Surrogate endpoint $\hat{S}_{0i} = \hat{P}_0(Y_i = 1 \mid W_i, A_i, S_i)$
First Application: Inference on $VE_0$ in CYD14

1. Obtain the TMLE-adjusted EOS from the CYD14 data $(W_i, A_i, S_i, Y_i), \ i = 1, \ldots, n$
   - Surrogate endpoint $\hat{S}_{0i} = \hat{P}_0(Y_i = 1 \mid W_i, A_i, S_i)$

2. Based on this EOS, calculate point and confidence interval estimates of

   $$VE(\hat{S}_0) = 1 - \frac{\text{Mean of } \hat{S}_0 \ \text{Vaccine Group}}{\text{Mean of } \hat{S}_0 \ \text{Placebo Group}}$$

   and of the numerator and denominator above
First Application: Inference on \( VE_0 \) in CYD14

1. Obtain the TMLE-adjusted EOS from the CYD14 data \((W_i, A_i, S_i, Y_i), \ i = 1, \ldots, n\)
   - Surrogate endpoint \( \hat{S}_{0,i} = \hat{P}_0(Y_i = 1 \mid W_i, A_i, S_i) \)

2. Based on this EOS, calculate point and confidence interval estimates of
   \[
   VE(\hat{S}_0) = 1 - \frac{\text{Mean of } \hat{S}_0 \text{ Vaccine Group}}{\text{Mean of } \hat{S}_0 \text{ Placebo Group}}
   \]
   and of the numerator and denominator above

3. Compare these results to direct estimates of
   \[
   VE_0 = 1 - \frac{\text{Overall Disease Rate in the CYD14 Vaccine Group}}{\text{Overall Disease Rate in the CYD14 Placebo Group}}
   \]
   and of the numerator and denominator, based on the CYD14 data \((W_i, A_i, Y_i), \ i = 1, \ldots, n\)
Second Application: Estimation of $VE^*$ in CYD15 Based on the Surrogate Built from CYD14

Calculate the $\hat{S}_i^* = \hat{P}(Y_i^* = 1 \mid W_i^*, A_i^*, S_i^*)$ surrogate endpoint values for CYD15 participants, $i = 1, \cdots, n^*$
Second Application: Estimation of $\text{VE}^*$ in CYD15 Based on the Surrogate Built from CYD14

1. Calculate the $\hat{S}_i^* = \hat{P}(Y_i^* = 1 | W_i^*, A_i^*, S_i^*)$ surrogate endpoint values for CYD15 participants, $i = 1, \cdots, n^*$

2. Use these to obtain point and confidence interval estimates of the CYD15 vaccine effect on the EOS and on the CYD15 Vaccine Group and Placebo Group means of $\hat{S}^*$
   - Assume the three assumptions needed for valid bridging
Second Application: Estimation of $VE^*$ in CYD15 Based on the Surrogate Built from CYD14

1. Calculate the $\hat{S}^*_i = \hat{P}(Y^*_i = 1 | W^*_i, A^*_i, S^*_i)$ surrogate endpoint values for CYD15 participants, $i = 1, \cdots, n^*$

2. Use these to obtain point and confidence interval estimates of the CYD15 vaccine effect on the EOS and on the CYD15 Vaccine Group and Placebo Group means of $\hat{S}^*$
   - Assume the three assumptions needed for valid bridging

3. Compare these results to direct estimates of

$$VE^* = 1 - \frac{\text{Overall Disease Rate in the CYD15 Vaccine Group}}{\text{Overall Disease Rate in the CYD15 Placebo Group}}$$

and of the numerator and denominator, based on the CYD15 data $(W^*_i, A^*_i, Y^*_i), i = 1, \cdots, n^*$
Table 1

Input variables, screens, and learner types used in the super-learner for the CYD14 dengue vaccine efficacy trial (35 total statistical algorithms for estimating $\psi_0 = E_0(Y|W, A, S)$ defined by screens crossed with learner types)

<table>
<thead>
<tr>
<th>Input variables</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>Baseline demographics age (range 2–14 years), sex, empirical frequencies of the 4 serotypes in placebo group failure events by country of the participant</td>
</tr>
<tr>
<td>$S$</td>
<td>Month 13 seropositivity to each of the 4 serotypes in the CYD-TDV vaccine, and average, minimum, and maximum of the 4 titers for both PRNT$_{50}$ and Microneutralization Version 2 (V2) assays</td>
</tr>
</tbody>
</table>

Screens

| screen.glmnet | Include variables with non-zero coefficients in a standard implementation of SL.glmnet (i.e., lasso) |
| screen.univar.logistic.x | Univariate logistic regression p-value < 0.10 using “x” most univariately significant terms. |
| screen.corX.x | Disallow pairs of quantitative variables with $R^2 > \text{"x"}$ |
| screen.PRNT | Disallow Microneutralization V2 titer variables |
| screen.MNv2 | Disallow PRNT$_{50}$ titer variables |

Learner types

| SL.mean | $E_0(Y|W, A = a, S) = \beta_a$ for $a \in \{0, 1\}$ |
| SL.glm | Logistic regression with all input variables |
| SL.step | Best logistic regression model by AIC from a step-wise search |
| SL.bayesglm | Logistic regression utilizing Cauchy Bayesian priors on model parameters |
| SL.polymars | Multivariate adaptive polynomial spline regression |
| Discrete SL | van der Laan, Polley, and Hubbard (2007) |
| Super Learner (SL) | van der Laan, Polley, and Hubbard (2007) |

Note: $^a$ All learners were fit separately for each treatment group $A = a$ for $a \in \{0, 1\}$ as described in Section 6.1. This is explicitly stated here for SL.mean.
Comparison of Prediction Performance Across Estimators (CV-AUCs\(^*\)) : CYD14

*Cross-validated area under the ROC curve (Van der Laan, Hubbard, and Pajouh, 2013)
Adding Month 13 Neutralization Markers Improves Prediction Over Baseline Demographics Only: CYD14

Cross-Validated ROC Curves*

Models including neutralization markers

Models based on Demo only

*van der Laan, Hubbard, and Pajouh (2013)

D1: Demographics
D2: Demo + MNv2
D3: Demo + PRNT
D4: Demo + MNv2 + PRNT
Using Both Assays Improves Performance in the Relevant Range of False Positive Rates: CYD14

Zooming in on the lower-left of the figure

Both MNv2 and PRNT

Vaccine Model (CV-AUC)
- D1 Best: SL_gam (0.61)
- D1 SuperLearner: (0.60)
- D2 Best: SuperLearner (0.82)
- D3 Best: SuperLearner (0.79)
- D4 Best: SuperLearner (0.84)

Cross-validated True Positive Rate

False positive rate (FP = No VCD but predicts VCD)
Predicted Risks of DENV-Any for the Vaccine Group on Held-Out Data: CYD14

Figure 11: Vaccine Estimated Probability of DENV-Any for Best Performing Models.
Best Performing Models for Estimating \( P_0(Y \mid W, A, S) \): CYD14

Table 2

Best performing models for estimating \( \psi_0 = E_0(Y \mid W, A, S) \) for the vaccine and placebo groups of the CYD14 trial. For both the vaccine and placebo groups the model with the lowest CV-MSE was a logistic regression (glm) using variables selected from the screen.MNv2 in Table 1.

<table>
<thead>
<tr>
<th>Model term</th>
<th>Coefficient</th>
<th>Odds ratio</th>
<th>2-Sided P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.09</td>
<td>2.96</td>
<td>0.26</td>
</tr>
<tr>
<td>AGE.9.11</td>
<td>-0.09</td>
<td>0.91</td>
<td>0.74</td>
</tr>
<tr>
<td>AGE.12.14</td>
<td>-2.46</td>
<td>0.09</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>MALE</td>
<td>-0.36</td>
<td>0.70</td>
<td>0.09</td>
</tr>
<tr>
<td>M13.MNv2.S1</td>
<td>-3.02</td>
<td>0.03</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>M13.MNv2.S2</td>
<td>0.77</td>
<td>2.16</td>
<td>0.02</td>
</tr>
<tr>
<td>M13.MNv2.S3</td>
<td>1.41</td>
<td>4.09</td>
<td>0.04</td>
</tr>
<tr>
<td>M13.MNv2.S4</td>
<td>-0.12</td>
<td>0.89</td>
<td>0.81</td>
</tr>
<tr>
<td>M13.MNv2.Ave</td>
<td>3.45</td>
<td>31.53</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>M13.MNv2.Min</td>
<td>-3.53</td>
<td>0.05</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>M13.MNv2.Max</td>
<td>-0.59</td>
<td>0.55</td>
<td>0.28</td>
</tr>
<tr>
<td>Sero2.frequency(a)</td>
<td>-0.91</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Sero3.frequency</td>
<td>-0.57</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Sero4.frequency</td>
<td>-0.38</td>
<td>0.02</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>1.97</td>
<td>7.16</td>
<td>0.01</td>
</tr>
<tr>
<td>AGE.9.11</td>
<td>0.84</td>
<td>2.32</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>AGE.12.14</td>
<td>-0.17</td>
<td>0.85</td>
<td>0.55</td>
</tr>
<tr>
<td>MALE</td>
<td>0.04</td>
<td>1.04</td>
<td>0.82</td>
</tr>
<tr>
<td>M13.MNv2.S1</td>
<td>-1.10</td>
<td>0.33</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>M13.MNv2.S2</td>
<td>0.25</td>
<td>1.29</td>
<td>0.34</td>
</tr>
<tr>
<td>M13.MNv2.S3</td>
<td>0.56</td>
<td>1.76</td>
<td>0.19</td>
</tr>
<tr>
<td>M13.MNv2.S4</td>
<td>0.06</td>
<td>1.06</td>
<td>0.84</td>
</tr>
<tr>
<td>M13.MNv2.Ave</td>
<td>1.01</td>
<td>2.75</td>
<td>0.43</td>
</tr>
<tr>
<td>M13.MNv2.Min</td>
<td>-2.62</td>
<td>0.07</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>M13.MNv2.Max</td>
<td>-0.25</td>
<td>0.78</td>
<td>0.51</td>
</tr>
<tr>
<td>Sero2.frequency(a)</td>
<td>-0.72</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Sero3.frequency</td>
<td>-0.54</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Sero4.frequency</td>
<td>-0.46</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Notes: \(a\) The reference age category is 2–8 year olds; \(b\) M13.MNv2.S1 is the binary indicator of a Month 13 positive response to serotype 1 using the MNv2 assay, with positive response defined by MNv2 serotype neutralization titer \(\geq 10\). M13.MNv2.S2-M13.MNv2.S4 are defined similarly; \(c\) M13.MNv2.Ave, M13.MNv2.Min, and M13.MNv2.Max coefficients are per one \(\log_{10}\) increase in neutralization titer value; \(d\) Serotype frequency variable coefficients are per 0.10 increase in the estimated serotype frequency of a participant’s country.
Estimated Optimal Surrogate (EOS) TMLEs of Target Parameters: CYD14

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TMLE Based on EOS</th>
<th>TMLE Based on ((W, A, Y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_0(Y = 1</td>
<td>\text{Vac}))</td>
<td>0.017 (0.016–0.019)</td>
</tr>
<tr>
<td>(P_0(Y = 1</td>
<td>\text{Plc}))</td>
<td>0.039 (0.036–0.042)</td>
</tr>
<tr>
<td>(VE_0 = 1 - \frac{P_0(Y=1</td>
<td>\text{Vac})}{P_0(Y=1</td>
<td>\text{Plc})})</td>
</tr>
</tbody>
</table>

- The point estimate results have to be similar by construction!
Using the Estimated Optimal Surrogate (EOS) in CYD15

How well do the EOS values $\hat{S}_i^*$ predict $Y_i^*$ in CYD15?

(b) CYD15 Reverse CDFs

![Graph showing reverse cumulative distribution functions for different groups in CYD15.](graph.png)
How Well Does the Surrogate-Based Estimator Estimate $VE^*$ in CYD15?

Table: Estimation in CYD15 based on the EOS built in CYD14 (not using outcome data $Y^*$ in CYD15) vs. TMLE estimation using ($W^*, A^*, Y^*$) in CYD15

<table>
<thead>
<tr>
<th></th>
<th>TMLEs of Surrogate Parameters$^1$</th>
<th>TMLEs of Clinical Parameters$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $\hat{S}^*$ Vac</td>
<td>0.020 (0.017–0.022)</td>
<td>$P(Y^* = 1</td>
</tr>
<tr>
<td>Mean $\hat{S}^*$ Plc</td>
<td>0.057 (0.049–0.065)</td>
<td>$P(Y^* = 1</td>
</tr>
<tr>
<td>$VE$ on $\hat{S}^*$</td>
<td>66% (58–72)</td>
<td>$VE^*$ 61% (51–69)</td>
</tr>
</tbody>
</table>

$^1$Based on $(W_i^*, A_i^*, \hat{S}_i^*(W_i^*, A_i^*, S_i^*))$

$^2$Based on $(W_i^*, A_i^*, Y_i^*)$ [use the actual clinical data]
1. Surrogate Endpoint Frameworks

2. Optimal Surrogate Framework

3. Simulation Studies

4. Application to Dengue Trials

5. Discussion
VanderWeele (2013, *Biometrics*) and discussants Joffe (2013) and Pearl (2013) suggest that a minimal requirement for an intermediate endpoint to be a useful surrogate endpoint is that it avoids the surrogate paradox.

VanderWeele (2013) shows that commonly used methods for surrogate endpoint evaluation generally do not guarantee avoiding this paradox.
Start at the Right Place

- VanderWeele (2013, *Biometrics*) and discussants Joffe (2013) and Pearl (2013) suggest that a minimal requirement for an intermediate endpoint to be a useful surrogate endpoint is that it avoids the surrogate paradox.

- VanderWeele (2013) shows that commonly used methods for surrogate endpoint evaluation generally do not guarantee avoiding this paradox.

- The optimal surrogate approach starts at this minimal requirement, defining the optimal surrogate in a way guaranteed to satisfy the Prentice definition of a valid surrogate.

  - Responds to Pearl's (2013) question: "If we take the negation of the “surrogate paradox” as a criterion for “good” surrogate, why cannot we create a new, formal definition of “surrogacy” that will automatically avoid the paradox?..."
The proposed approach uses predicted clinical endpoint values as the surrogate, implying that the mean surrogate treatment effect has the same interpretation as the mean clinical treatment effect.

An obvious approach for maximally tying the surrogate to the clinical endpoint.

Yet typically surrogate endpoints are biomarkers on their own scale, which is often different from the clinical endpoint scale.
Nonparametric Supervised Learning Approach

- Using super-learner + TMLE seeks to avoid dubious assumptions and use all of the information in the data.

- Broad application to clinical fields where multiple biomarkers are measured that could contribute to a surrogate endpoint, and the objective is supervised learning of most promising surrogate endpoints that may depend on baseline covariates as well as post-vaccination response response endpoints.
Nonparametric Supervised Learning Approach

- Using super-learner + TMLE seeks to avoid dubious assumptions and use all of the information in the data.

- Broad application to clinical fields where multiple biomarkers are measured that could contribute to a surrogate endpoint, and the objective is supervised learning of most promising surrogate endpoints that may depend on baseline covariates as well as post-vaccination response endpoints.

- This framework also applies for exploratory analyses of observational studies to generate promising candidate surrogates, with all of the results holding under the additional assumption that all confounders \( W \) of treatment assignment are measured and included in the super-learner.
Some Challenges Posed to the Framework

- The estimated optimal surrogate (EOS) may be based on a complicated combination of models that is hard to interpret.

- Hence the importance of building multiple EOSs from different input variable sets ranging from single-variable to all-variable models, where cross-validation criteria allow principled selection of a most parsimonious EOS with acceptable performance.
The estimated optimal surrogate (EOS) may be based on a complicated combination of models that is hard to interpret.

Hence the importance of building multiple EOSs from different input variable sets ranging from single-variable to all-variable models, where cross-validation criteria allow principled selection of a most parsimonious EOS with acceptable performance.

Will it be complicated for other researchers to use an EOS?

Broad use may require a research paradigm embracing open research that posts to the web a calculator that inputs \((W, A, S)\) and outputs the EOS value.
Elaborations

- Missing data on \((W, A, S, Y)\)
  - E.g., case-cohort or case-control sampling of \(S\)
  - Happenstance missing data

- Some participants experience \(Y\) before \(S\) is measured at \(\tau\)

- Right-censoring of \(Y\) (failure time endpoint), competing risks outcomes

- Tailoring the super-learner to contextual features [sample size, event rate, dimensionality of \((W, S)\)]

- Confidence intervals about the clinical treatment effect
  \[
  VE^* = 1 - \frac{P(Y^* = 1|A = 1)}{P(Y^* = 1|A = 0)}
  \]
  in a new setting accounting for the error in estimating the optimal surrogate
Acknowledgements

- SanofiPasteur colleagues for collaboration and sharing the data

- Participants and study personnel of the CYD14 and CYD15 dengue Phase 3 trials

- NIH NIAID support for the grant “Statistical Methods in HIV Vaccine Efficacy Trials”