Module 8: Evaluating Vaccine Efficacy

Session 7

Dean Follmann, Peter Gilbert, **Erin Gabriel**, Michael Sachs

July 25, 2017

Background

Ideal Causal Comparison

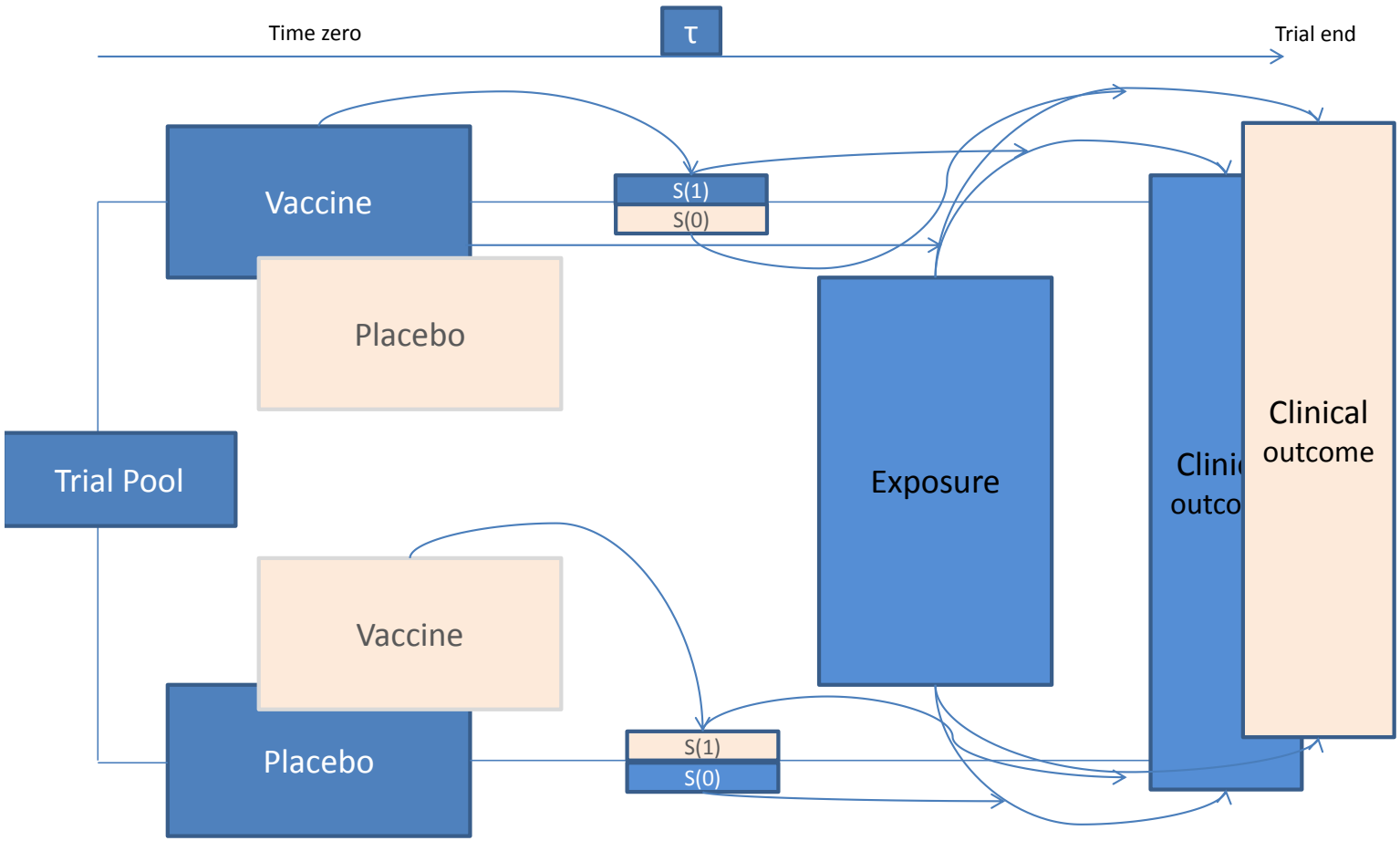


Figure: image

Risk

Joint:

$$\begin{aligned} risk_1(s_1, s_0) = &\quad Pr\{Y(1) = 1 | S(1) = s_1, S(0) = s_0\}, \\ risk_0(s_1, s_0) = &\quad Pr\{Y(0) = 1 | S(1) = s_1, S(0) = s_0\}. \end{aligned}$$

Marginal :

$$\begin{aligned} risk_1(s_1) = &\quad Pr\{Y(1) = 1 | S(1) = s_1\}, \\ risk_0(s_1) = &\quad Pr\{Y(1) = 1 | S(1) = s_1\}. \end{aligned}$$

causal effect predictiveness (CEP) function $h(x, y) = 0$ iff $x = y$
$VE(s_1, s_0) = 1 - risk_1(s_1, s_0)/risk_0(s_1, s_0)$ and
$VE(s_1) = 1 - risk_1(s_1)/risk_0(s_1)$

Specific Correlate of Protection based on VE modification

A Specific Correlate of Protection (CoP) is a biomarker that predicts vaccine efficacy in same setting as the evaluation trial:

- Average Causal Necessity (ACN) $VE(s_1) = 0$ where $S(1) =$ CB level of S(0), or $VE(s_1, s_0) = 0$ where $S(1) = S(0)$ [Frangakis and Rubin, 2002]
- Average Causal Sufficiency (ACS) $VE(s_1, s_0) > 0$ when $s_1 > s_0$ CB level of S(0) [Gilbert and Hudgens, 2008]
- Large variation in VE over $(S(1), S(0))$, Wide effect Modification (WEM) [Gilbert et al., 2011aWolfson and Gilbert [2010]]

Specific Correlate of Protection more generally

Rather than using the VE function, consider the CEP $h(x, y) = log(x/y)$ or $x - y$:

- ACN $h(risk_1(s_1), risk_0(s_1)) = 0$ where $S(1) =$ CB level of S(0), or $h(risk_1(s_1), risk_0(s_1)) = 0$
- ACS $h(risk_1(s_1), risk_0(s_1)) < 0$ for some $s_1 >> s_0$ or $s_1 >>$ the CB level of S(0)
- WEM, Large variation in CEP over $(S(1), S(0))$

Criteria Ranking

Several works have suggested that WEM is the primary criteria for a CoP:

- WEM alone provides a target for vaccine improvement.
- ACS or ACN can hold alone for a useless CoP.
- ACS or ACN alone are not sufficient to have value as a CoP.

- WEM plus ACS is sufficient
- ACS plus ACN is sufficient, as this implies WEM
- ACN and ACS with $VE > 0$ when $s_1 - s_0 > 0$, is the strongest minimal criteria and implies a consistent surrogate.

Correlate Quality Example

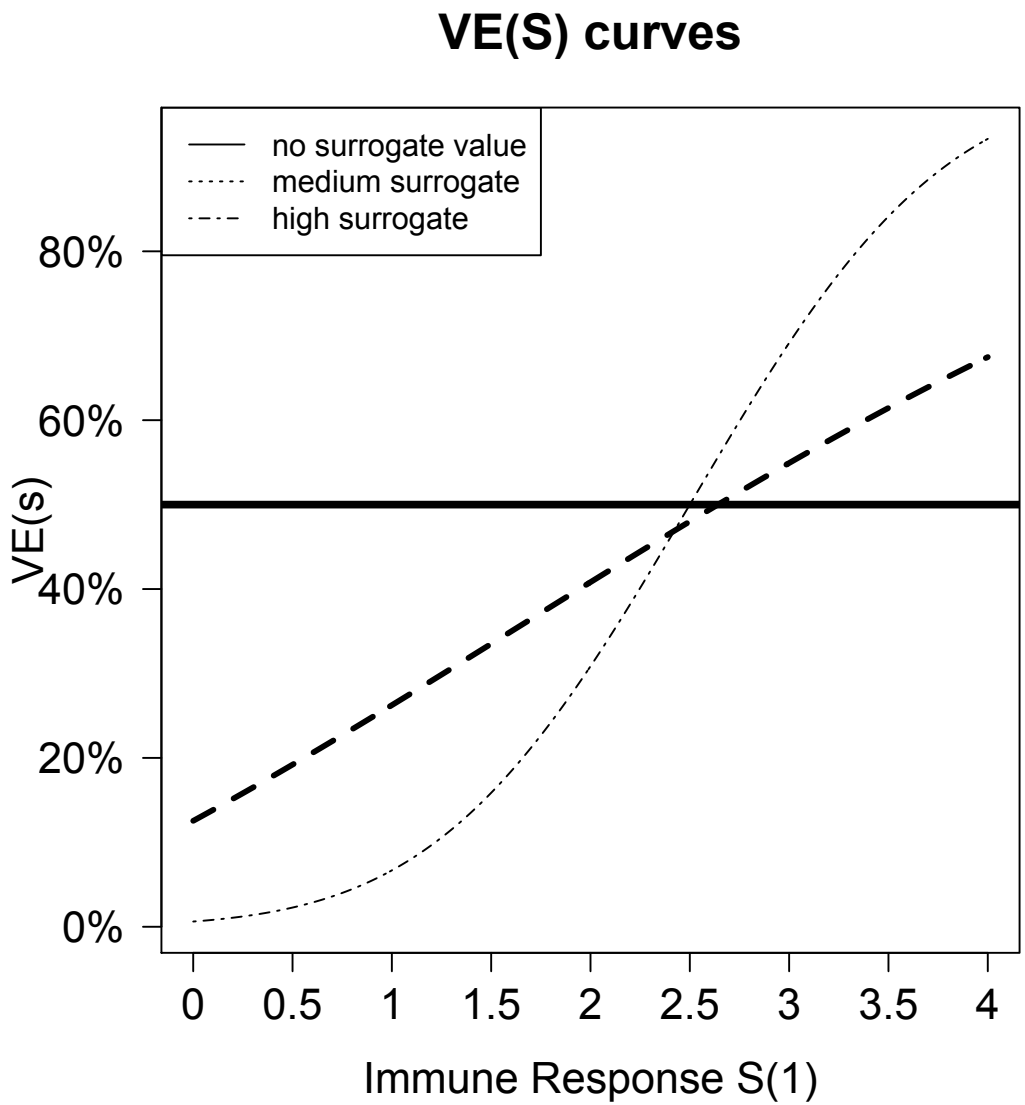


Figure: image

The problems:

1. Interpretation of the estimates as desired, i.e. identify the desired estimands
2. Systematic missing data
3. Estimation of and inference on the risk estimands for evaluation of biomarkers as CoP
4. Implementation of this estimation and inference

Assumptions: identification of the desired estimands

No missing data

If we could observe all outcomes in all subjects:
$\{Y_i, T_i, C_i, S_i, \mathbf{W}, \mathbf{X}, |Z_i = 1\}$ and $\{Y_i, T_i, C_i, S_i, \mathbf{W}, \mathbf{X}, |Z_i = 0\}$
Where:

- Y is the observed outcome; $Y = I_{T<C}$
- T is the time an event outcome would occur, observed or not
- C is the time on trial or prior to an even driven outcome
- S in the intermediate outcome measured before or at time τ
- **W, X** are the baseline characteristics for subjects
- Z is the observed randomization assignment

We could use standard methods of estimation, but we would still have problem 1. In order to link the observed S to S(1), even if we observe $S|Z = 1$, requires assumptions.

Assumptions for Identification: Set 1

- A1: Stable unit treatment value assumption (SUTVA) and consistency
- A2: Ignorable treatment assignment

A3 may not be needed in all cases. Under A1-A2 alone what we observe in the vaccine arm, are the potential outcomes of interest $\{Y(1), S(1), \ldots\}$

Assumptions for Identification: Example 1

For example: For a clinical outcome of HPV wart recurrence by 1 year and the CoP of interest HPV DNA detection by 6 weeks, post vaccination in those known to be infected prior to vaccination

- 6 weeks post in the vaccine arm is S(1); S(0) is the placebo arm under A1&A2
- 1 year post in the vaccine arm is Y(1); Y(0) in the placebo arm under A1&A2
- There is no need for A3, if there is recurrence within 6 weeks, there is DNA detection within 6 weeks. So, S(1) and S(0) are still observable.
- A different assumption may be needed that vaccination had no impact on risk of death before 6 weeks. However, this is a much more plausible assumption than the vaccine having no impact on the desired clinical outcome

Assumptions for Identification: Example 2

A3 is needed when Y(1) can occur before the measurement of S(1), and alter S(1) in some way
For example: HIV vaccine, Y infection status at 1 year, S immune response at 6 weeks post vaccination

- 6 weeks post in the vaccine arm is S(1); S(0) is the placebo arm under A1&A2
- 1 year in the vaccinated the observed infection status is Y(1); Y(0) in the placebo arm under A1&A2
- When infection occurs prior to 6 weeks, S is undefined, and therefore that subject must be removed.
- The assumption that vaccination had no impact on risk of death before 6 weeks is likely still needed, however, often ignored.

Assumptions for Identification: Set 2

- A3: Equal individual clinical risk up to time τ, $T(1) < \tau$ if and only if $T(0) < \tau$
- A4: Case CB, $S(0) = Q$, some constant Q for all subjects

Assumptions for Identification

Under assumption A1-A3 alone we can identify the marginal risk estimand in the vaccine arm:

- $\text{risk}_1(s_1)$

Under assumption A1-A4 we can identify the joint and marginal risk estimand in the vaccine arm:

- $\text{risk}_1(s_1) = \text{risk}_1(s_1, s_0)$

The A1-A4 assumptions allow for identification of risk estimand in vaccinated subjects alone (CoR analysis):

We are still missing all the S(1) for all placebo/control subjects.

Systematic missing data

BIV, BIP, CPV

Follmann [2006] introduced two trial augmentations for observing or imputing the missing S(1) values, assuming constant biomarker:

- Baseline irrelevant vaccination, (BIV) vaccinating all subjects with a different vaccine prior to randomization and using this response to predict S(1) for the placebo arm
- Baseline immunogenicity predictor, (BIP) measuring baseline variables that are predictive of S(1)
- Close-out placebo vaccination, (CPV) at the close of the trial vaccinate those subjects that have not dropped-out or had an observed event

BIV, BIP assumptions

BIV is a special form of BIP:

- BIV are assumed to be independent of outcome conditional on S(1). For the BIV response W, we can assume $Y \perp W | S(1)$,
- While a BIP, W, should be considered for inclusion in the risk model.

When BIP or BIV alone are used, risk model testing is very limited.

Assumptions CPV

Assumptions for CPV:

- Individual time constancy of the immune response distribution, $S(1) = S^c(1)$ almost surely.
- No infections in the uninfected placebo group during the close-out period, $Pr\{Y(0)^c = 0 | Y(0) = 0\} = 0$.

Where $S^C(1)$ is the measurement taken τ time after closeout vaccination and Y^C is the indicator of observed event during the closeout period. The second assumption only needed in event-driven settings same as A3.

Extensions to BIP and CPV

Gabriel and Follmann [2016] introduce several augmented trial designs that extend Follmann [2006]:

- Baseline measurement of the candidate correlate (BSM)
- Close out vaccination, or treatment, of all placebo or control subjects (CCT)
- Run-in vaccination of all subjects
- Step-wedge and Cross-over trials

BSM, BIP, CCT Augmentations

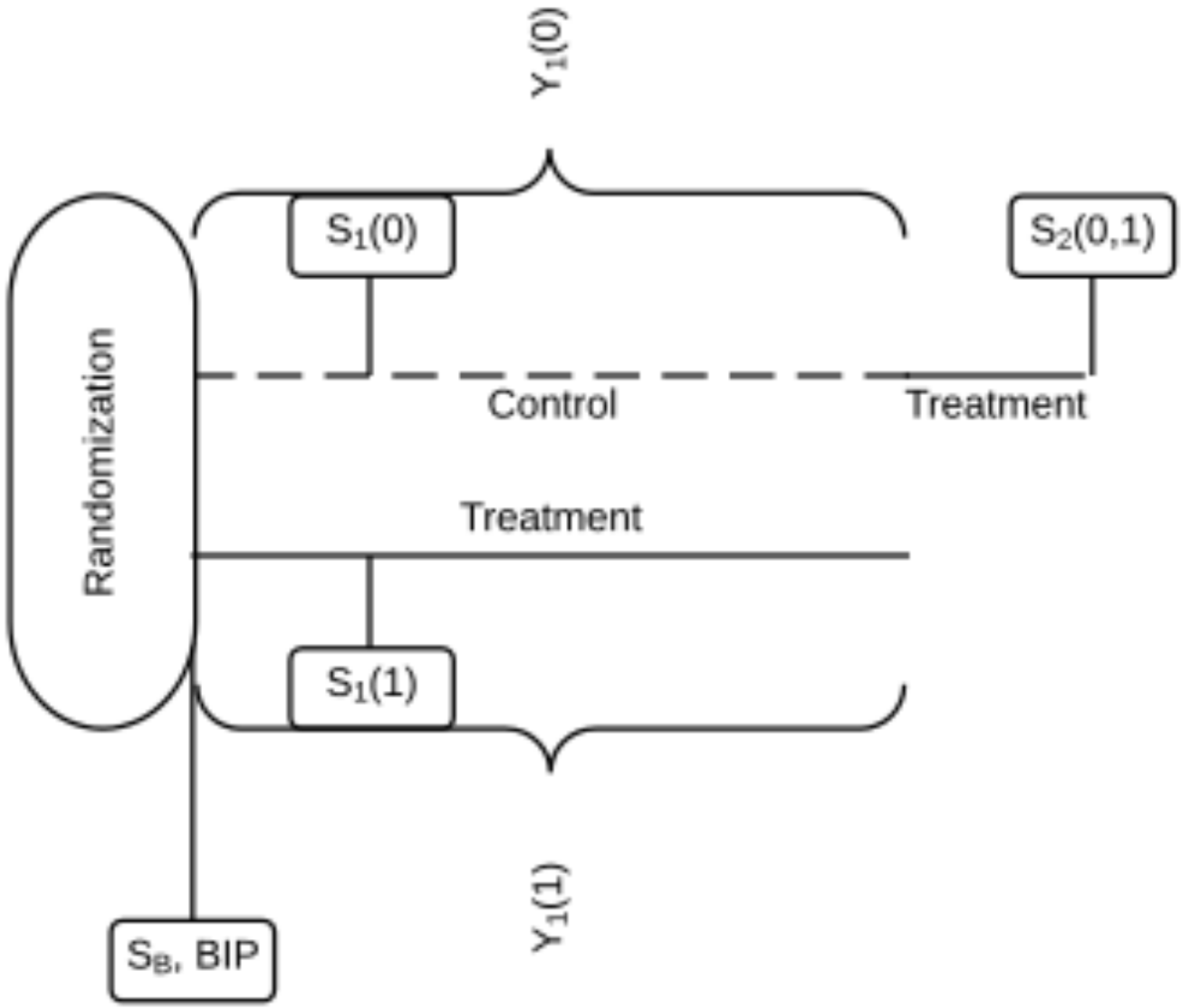


Figure: image

Assumptions BSM

Assumptions for BSM to induce CB:

- Individual time constancy of the intermediate response from baseline to time τ after randomization under control, $S_B = S_1(0)$ almost surely.

When this assumption holds, for the candidate correlate $S = S_\tau - S_B$, $S(0) = 0$ for all subjects, i.e. Case CB, assumption A4 will hold. The immune response to malaria antigen during the dry season, would be an example of a BSM measurement that should not change.

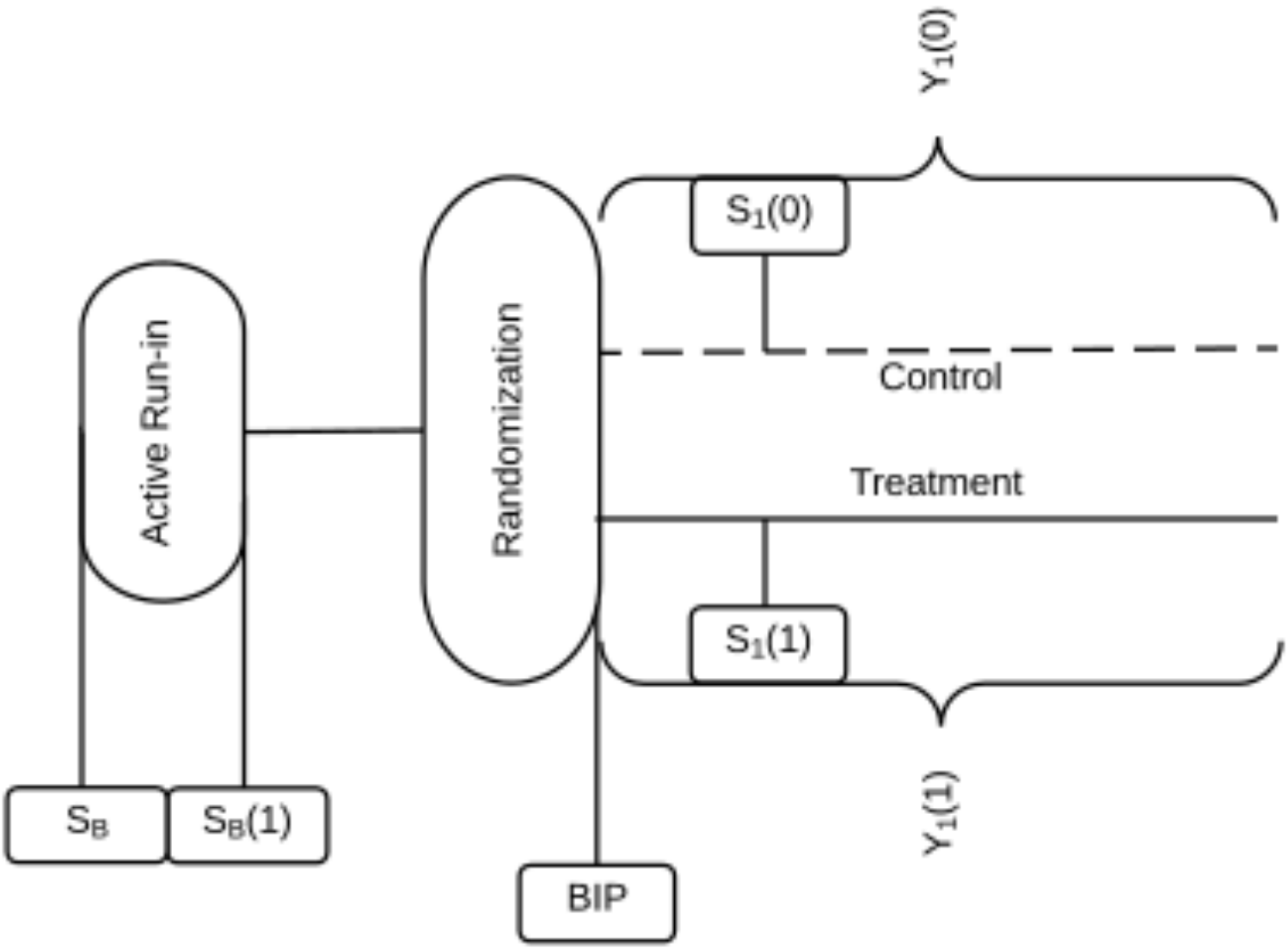
CCT

Assumptions for CCT:

- Individual time constancy of the intermediate response at time τ post treatment under previous control $S_t(0,1) = S_1(1)$ almost surely.

Where $S_t(0,1) = S^C$ is the measurement taken τ time after closeout vaccination of those subjects previously on control for 1 period. A immuno-therapy vaccine in cancer is an example of a setting where CCT could be used, as all subjects have cancer, crossing over all placebo subjects over at the end of the trial is possible.

Run-in


Active Run-in
Randomization
S_B
$S_B(1)$
BIP
$S_1(0)$
Control
Treatment
$S_1(1)$
$Y_1(0)$
$Y_1(1)$

Assumptions Run-in

Assumptions for Run-in:

- Distributional time constancy of clinical outcome after one period of active treatment within all subgroups defined by the intermediate response, $(Y_1(z)|S(1) = s_1, S(0) = s_0) =^d (Y_2(1, z)|S(1) = s_1, S(0) = s_0)$ for all s_1 and s_0 and $z \in \{0, 1\}$
- Individual time constancy of the intermediate response any treatment regardless of previous treatment $S_B(1) = S_1(1)$ and $S_B = S_1(0)$ almost surely.

Although this might seem implausible, I am currently working in malaria vaccines where this is very plausible.

Run-in Example

Malaria vaccine in a seasonal driven malaria endemic area:

- Enrolling, measuring S_B and vaccinating all subjects at the end of the dry season during year 1 and measuring $S_B(1)$ τ after vaccination
- As current vaccine has short lived immune response and efficacy, allowing for a wet and dry season for wash out will adequately reduce the immune responses and efficacy. As subjects are exposed every year, a vaccination should not increase immune responses in a way that a previous year's worth of repeated exposure does not.
- Randomize and vaccinate subjects at the end of the dry season, measure S(1) and S(0) τ time after vaccination.
- When assumptions hold, all the intermediate counterfactual measurements are obtained
- We can directly test the Run-in assumption 2, and when it does not hold we can perfectly model the use of $S_B(1)$ as a BIP

Run-in Example

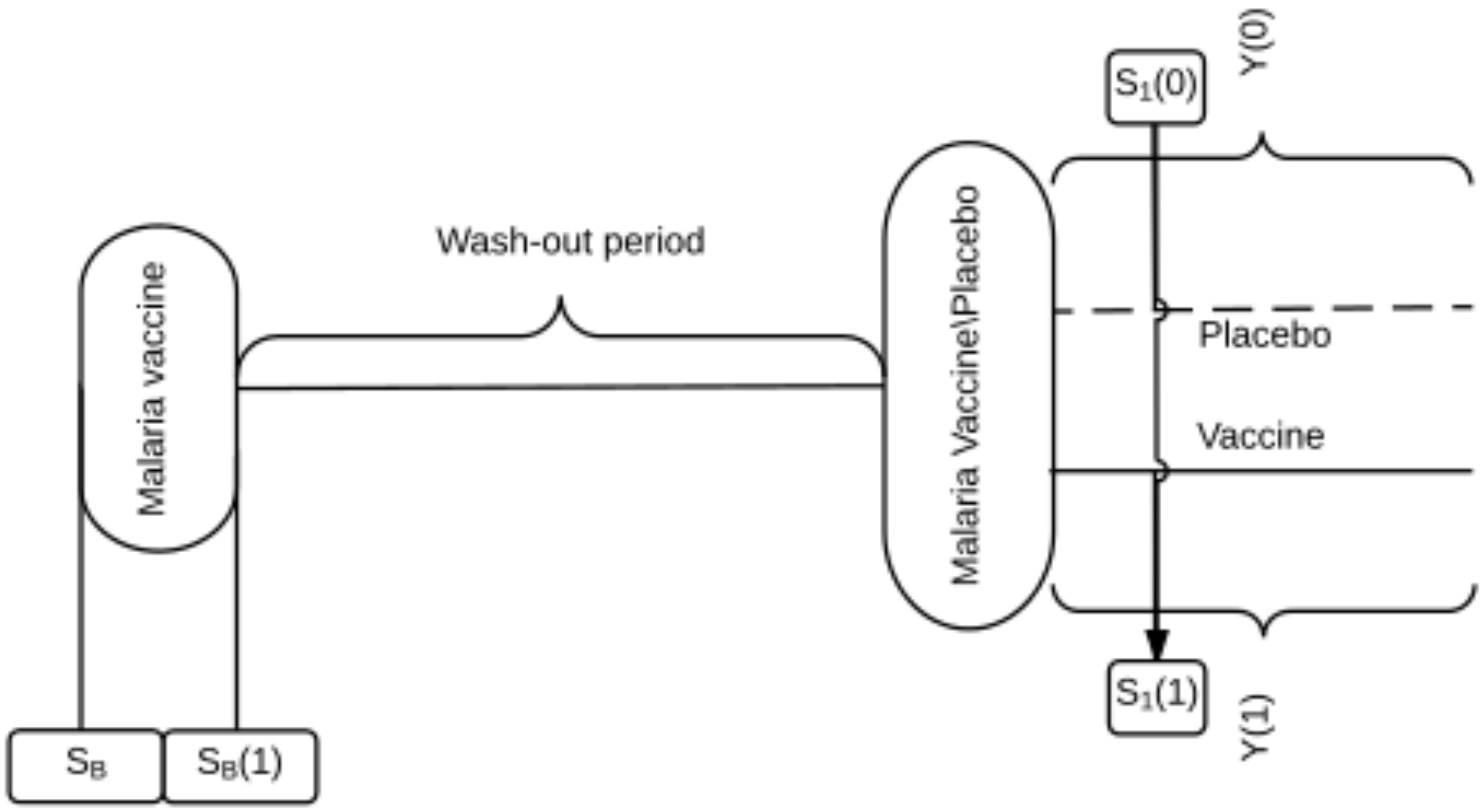


Figure: image

Estimation and Inference

General Risk Definition

Joint:

$$\begin{aligned} risk_1(s_1, s_0) = &\quad g[F_{s_0,s_1}\{Y(1)|S(1) = s_1, S(0) = s_0\}], \\ risk_0(s_1, s_0) = &\quad g[F_{s_0,s_1}\{Y(0)|S(1) = s_1, S(0) = s_0\}]. \end{aligned}$$

Marginal :

$$\begin{aligned} risk_1(s_1) = &\quad g[F_{s_1}\{Y(1)|S(1) = s_1\}], \\ risk_0(s_1) = &\quad g[F_{s_1}\{Y(0)|S(1) = s_1\}]. \end{aligned}$$

Time-dependent Risk Definitions

Many Ways to Define Risk in this Setting:

$$risk_z^{CDF}(t|s_1) \equiv \quad Pr(T(z) \leq t | S(1) = s_1, T(1) > \tau, T(0) > \tau)$$

or based on the hazard function,

$$risk_z^{HZ}(t|s_1) = \quad \frac{\int_t 1 - risk_z^{CDF}(t|s_1)}{1 - risk_z^{CDF}(t|s_1)}.$$

$risk_z^{HZ}(t|s_1)$ **conditions on being at risk at time t**

$$VE(t|s_1) = 1 - \frac{risk_1(t|s_1)}{risk_0(t|s_1)}$$

Constant Shape Weibull model for risk

$$\lambda = 1/exp(\beta_0 + \beta_1 * z + \beta_2 * s_1 + \beta_3 * s_1 * z)$$

$$risk_z^{CDF}(t|s_1)$$

$$= \quad 1 - exp(-(t * \lambda)^a)$$

$$risk_z^{HZ}(t|s_1)$$

$$= \quad \frac{a}{\lambda} \times \left(\frac{t}{\lambda}\right)^{(a-1)}.$$

with $risk_z^{HZ}(t|s_1)$ and a CEP of log(RR), the CEP is time-free. However, contrasts of $risk_z^{CDF}(t|s_1)$ are always causal, whereas contrasts in the Hazard are not in some cases.

With assumption of this model and A3, time must start at τ in the model

Exponential model for risk

$$risk_z^{CDF}(t|s_1) = \quad 1 - exp(-(t * 1/\lambda))$$

$$risk_z^{HZ}(s_1) = \quad 1/\lambda.$$

Then for the CEP of log(RR):

$$log(risk_1(s_1)/risk_0(s_1)) = \quad (-\beta_1 - \beta_3 * s_1)$$

- ACN $\beta_1 = 0$
- ACS $\beta_3 \geq 0$
- WEM $\beta_3 \neq 0$

Count outcomes

$$risk_z(s_1) =$$

$$E[Y|S(1) = s_1, Z = z] = \quad exp(\beta_0 + \beta_1 * z + \beta_2 * s_1 + \beta_3 * s_1 * z)$$

Then for the CEP of log(RR):

$$log(risk_1(s_1)/risk_0(s_1)) = \quad (\beta_1 + \beta_3 * s_1)$$

- ACN $\beta_1 = 0$
- ACS $\beta_3 \leq 0$
- WEM $\beta_3 \neq 0$

Continuous outcomes

In the package we use the CDF for risk
$risk_z(s_1) =$

$$E[Y|S(1) = s_1, Z = z] = \quad \beta_0 + \beta_1 * z + \beta_2 * s_1 + \beta_3 * s_1 * z$$

Then for the CEP of risk difference:

$$risk_1(s_1) - risk_0(s_1) = \quad (\beta_1 + \beta_3 * s_1)$$

- ACN $\beta_1 = 0$
- ACS $\beta_3 \leq 0$
- WEM $\beta_3 \neq 0$

Standard ML when Full CCT or Run-in is performed

When CCT, or Run-in is performed we have all the missing S(1) or S(1) and S(0) measurements:

- Standard methods can be used for estimation ***Assumptions are still needed to link these estimates to the estimands**
- Model and assumption testing can be carried out
- When models or assumptions fail, EML (or measurement error modeling) can be used to correct for the bias and much of the efficiency loss [Gabriel and Follmann, 2016]

Full likelihood ML with BIP alone

Observed Likelihood:

$$\mathrm{L}(\beta, \gamma, \nu) \equiv \prod_i f(T_i | Z_i, S_i(1), W_i, Y_i, \delta_i; \gamma, \beta)$$

$$f(T_i | Z_i, S_i(1), W_i, Y_i, \delta_i; \gamma\, \beta) = \{g_z(t | s_1, w, y; \gamma, \beta,)\}^{\delta}$$

$$\times \quad \left\{ \int g_z(t | s, w, y, \gamma, \beta) dF_{S(1)|W}(s) \right\}^{(1-\delta)}$$

Where δ indicates $S(1)$ is observed, and assuming a model for $S(1)|W$

Full likelihood ML with CPV alone or CPV+BIP

Observed Likelihood is the same less W:

$$\mathrm{L}(\beta, \gamma, \nu) \equiv \prod_i f(T_i | Z_i, S_i(1), X_i, Y_i, \delta_i; \gamma, \beta)$$

$$f(T_i | Z_i, S_i(1), X_i, Y_i, \delta_i; \gamma\, \beta) = \{g_z(t | s_1, x, y; \gamma, \beta,)\}^{\delta}$$

$$\times \quad \left\{ \int g_z(t | s, x, y, \gamma, \beta) dF_{S(1)}(s) \right\}^{(1-\delta)}$$

Where δ indicates $S(1)$ is observed, and assuming a model for $S(1)$

Estimated Maximum Likelihood

Integrate the observed likelihood over the estimated distribution of $S(1)|W$

$$\mathsf{L}(\beta, \gamma, \widehat{\nu}) \equiv \prod_i \widehat{f}(T_i|Z_i, S_i(1), W_i, Y_i, \delta_i; \gamma, \beta)$$

$$\widehat{f}(T_i|Z_i, S_i(1), W_i, Y_i, \delta_i; \gamma\, \beta) = \{g_z(t|s_1, w, y; \gamma, \beta,)\}^{\delta}$$

$$\times \quad \left\{\int g_z(t|s, w, y, \gamma, \beta) d\widehat{F}_{S(1)|W}(s)\right\}^{(1-\delta)}$$

Assumed parametric form of $S(1)|W$ can be tailored to the data [Pepe and Fleming, 1991]

Parametric EML

Assumed parametric model for $S(1)|W$

- Follmann [2006] linear normal, linear in W
- Gilbert and Hudgens [2008] censored normal, linear in W

Package allows more models.

Semi-parametric EML

The Model for outcome remains the same, the model for the correlate is given by:

$$F_{S(1)|W} \sim \quad F[\{s_1 - \mu(w)\}/\sigma(w)] = F(\varsigma),$$

$$\sum_{k=1}^{n_V} \frac{w_k(s_{(1,k)} - \gamma' w_k)}{\sigma(w_k)^2} = 0$$

$$\sum_{k=1}^{n_V} \frac{w_k\{(s_{(1,k)} - \gamma' w_k)^2 - \sigma(w_k)^2\}}{\sigma(w_k)^2} = 0$$

$\mu(w) = \gamma' W$ and $\ln(\sigma(w)) = \eta' W$
[Huang and Gilbert, 2011, Heagerty and Pepe, 1999]

Creating new S(1)

$$S^*_{i,k}(1) = \hat{\gamma}' w_i + \exp(\hat{\eta}' w_i)\varsigma_k$$

There are k total imputations used for each missing $S(1)$ value. We can then use these imputed values to estimate $\left\{ \int g_z(t_i|s, w_i, q_i, y_i, \gamma, \beta) dF_{S(1)|W}(s) \right\}$ by the empirical integral:

$$\left(\frac{1}{n_V}\right) \sum_k^{n_V} g_z(t_i|S^*_{i,k}(1), w_i, q_i, y_i, \gamma, \beta).$$

This gives us a general estimated log likelihood of:

$$\mathrm{l}(\beta, \gamma, \hat{\nu}) = \sum_i log(g_z(t_i|s_{i,1}, w_i, q_i, y_i; \gamma\, \beta,)) * \delta_i$$

$$+ \quad \sum_i \left\{ \left(\frac{1}{n_V}\right) \sum_k^{n_V} log(g_z(t_i|S^*_{i,k}(1), w_i, q_i, y_i, \gamma, \beta)) \right\} * (\delta_i - 1)$$

Non-parametric EML

Gilbert and Hudgens [2008] introduced a non-parametric EML

- Categorical S(1) and W, and binary outcome Y
- For those Subjects missing S(1), sum the likelihood contribution over the S(1) for those subjects with the same W
- Because S(1) is categorical the model for $Y|S(1)$ can also be non-parametric

Can categorize S(1), however makes evaluation of ACN harder.

EML Method: Assumptions for unbiased/consistent estimation

- The assumed model for $S(1)|W$ is correct and consistently estimated
- The risk model for $Y|S(1), Z$ is correct
- Under BIP alone, no interaction between the BIP W and S associated with outcome, No $S(1) * W$ or $S(1) * W * Z$
- Under BIP alone linear model for $S(1)|W$ no interaction between the BIP W and Z associated with outcome except through S, No $Z * W$

EML Method: Asymptotic Properties

- risk model parameters are asymptotically normal
- consistent for risk model parameters
- Given the zero probability of observing S(1) in the placebo arm, asymptotic unbiasedness and asymptotic variance unknown and highly dependent on the model.

Strange CPV+BIP outcomes - more CPV measurements taken when a good BIP is used, BIP alone is more efficient [Gilbert et al., 2011b]

Sub-sampling Paradox

CPV is only ever observed for those subjects without an event.

- The estimation of S(1)|W assumes a validation set that includes subjects with events in the vaccine arm
- Using the CPV measurements as direct imputations in the conditional risk, uses a different validation set for the risk estimation than was assumed for estimation of S(1)|W, even under assumption 2 as it depends on Y.

Huang et al. [2012] discuss this paradox and provide an estimation method that solves the problem.

Pseudoscore Method

Biased Observation of $F(s|Z,W)$, which we do when we have CPV for Z =0

$$F(s|Z,W) = \frac{P(S \leq s|W,Z,\delta=1)P(\delta=1|Z,W)}{P(\delta|S=s,Z,W)}$$
$$\equiv \frac{F^*(s|Z,W)P(\delta=1|Z,W)}{P(\delta|S=s,Z,W)}.$$

$F^*(s|Z=z,W=w) = F(s|W=w,Z=z,\delta=1)$
an empirical estimate of $F^*(S(1)|Z,W)$,

$$F_N(s_1|z,w) = \frac{\sum_i I_{[S\leq s_1,Z=z,W=w,\delta=1]}}{\sum_i I_{[Z=z,W=w,\delta=1]}}$$

where δ is the indicator of observing S(1) and
$\phi(t,Z,W) = P(\delta=1|T=t,Z,W=w)$, positive expected probability of selection into second phase with respect to outcome.

Pseudoscore Equation For binary outcome

Pseudoscore Equation

$$S_{Ps}(\beta,\gamma;F_N,\phi)=\sum_{i\epsilon v} S_{\beta,\gamma}(Y_i|S(1)_i,Z_i,W_i)+$$

$$\sum_{j\epsilon\overline{v}}\sum_{i\epsilon v}\frac{S_{\beta.\gamma}(Y_j|S(1)_i,Z_j,W_j)h^{\phi}_{z_j}(Y_j|S(1)_i,W_j,\beta,\gamma)I_{[z_j=Z_i,w_j=W_i]}}{\sum_{l\epsilon v}h^{\phi}_{z_j}(Y_j|S(1)_l,W_j,\beta,\gamma)I_{[z_j=Z_l,w_j=W_l]}}$$

$$h^{\phi}_{z}(y|s_1,z,w,\beta,\gamma)=\frac{g_z(y|s_1,w,\beta,\gamma)}{q^{\phi}_{z}(s_1,w,\beta,\gamma)}.$$

$$q^{\phi}_{z}(S(1),W\beta,\gamma)=\int\phi(y,z,W)g_z(y|S(1),W,\beta,\gamma)dy.$$

Pseudoscore Equation For Time-to-event

Pseudoscore Equation

$$S_{Ps}(\beta,\gamma;F_N,\phi)=\sum_{i\epsilon v}S_{\beta,\gamma}(T_i|S(1)_i,Z_i,W_i,Y_i)+$$

$$\sum_{j\epsilon\overline{v}}\sum_{i\epsilon v}\frac{S_{\beta.\gamma}(T_j|S(1)_i,Z_j,W_j,Y_j)h^{\phi}_{z_j}(T_j|S(1)_i,W_j,Y_j,\beta,\gamma)I_{[z_j=Z_i,W_j=W_i]}}{\sum_{l\epsilon v}h^{\phi}_{z_j}(T_j|S(1)_l,W_j,Y_j,\beta,\gamma)I_{[z_j=Z_l,W_j=W_l]}}$$

$$h^{\phi}_{z}(t|s_1,z,w,y,\beta,\gamma)=\frac{g_z(t|s_1,w,y,\beta,\gamma)}{q^{\phi}_{z}(s_1,w,\beta,\gamma)}.$$

$$q^{\phi}_{z}(S(1),W,\beta,\gamma)=\int\phi(t,z,W)g_z(t|S(1),W,Y,\beta,\gamma)dt.$$

Pseudoscore Method: Assumptions

- $\int_t \phi(t, z, W)dt > 0$ for all ϕ in the neighborhood of the true ϕ_0.
- $g(t|s, w, y; \beta, \gamma) > 0$ for almost all observed data in the neighborhood of the true β_0 and γ_0. Strictly positive value given the assumption of the parametric model for outcome T.
- $P(\delta = 1|T, S, Z, W) = P(\delta = 1|T, Z, W) = \phi(T, Z, W)$, S is missing at random, (MAR).

Assumption 1 as stated requires CPV.

Pseudoscore Method: Asymptotic Properties

Under regularity conditions 4.1-4.3 of Theorem 4.1 of Chatterjee [1999] and van der Vaart and Wellner [1996]

- a. The pseudoscore estimating equations $S_{Ps}(\beta, \gamma; F_N, \hat{\phi}) = 0$ have a unique, consistent sequence of solutions, $\{\hat{\theta}_N^{Ps}\}_{N \geq 1}$, and
- 2.

$$\sqrt{N}(\hat{\theta}_N^{Ps} - \theta_0) = -\Psi_\theta^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} g^0(T_i | S_i, W_i, Y_i, \delta_i) + o_p(1);$$

where $g^0(T, S, W, Y, \delta) =$
$\delta\{S_{0,\beta_0,\gamma_0}(t|s, w, y) + a(s, w)\} + (1 - \delta)S_{0,\beta_0,\gamma_0;F_0}(t|w, y)$
where the subscript 0 indicates that both the model and the parameters in the model are the truth, and

- c. If $Var_0(g^0(T|S, W, Y, \delta)) < \infty$, then $\sqrt{N}(\hat{\theta}_N^{Ps} - \theta_0) \rightarrow_d N(0, \Omega)$, where Ω is defined by the sandwich formula,

$$\Omega = [\Psi_\theta(\theta_0, F_0^*)]^{-1} Var^0(g_0(T, S, W, Y, \delta))[\Psi_\theta^t(\theta_0, F_0^*)]^{-1}$$

BIP alone Pseudoscore for Binary outcome

Huang et al. [2012] develop a Pseudoscore method for a Binary outcome for BIP alone settings, relaxing PS assumption 1, using the randomization assumption:

- $\phi(t, W) > 0$ for all ϕ in the neighborhood of the true ϕ_0, .

$$S_{Ps}(\beta, \gamma; F_N, \phi) = \sum_{i \epsilon v} S_{\beta,\gamma}(Y_i | S(1)_i, Z_i, W_i) +$$

$$\sum_{j \epsilon \overline{v}} \sum_{i \epsilon v} \frac{S_{\beta,\gamma}(Y_j | S(1)_i, Z_j, W_j) h^{\phi}_{z_j}(Y_j | S(1)_i, W_j, \beta, \gamma) I_{[w_j = W_i]}}{\sum_{l \epsilon v} h^{\phi}_{z_j}(Y_j | S(1)_l, W_j, \beta, \gamma) I_{[w_j = W_l]}}$$

$$h^{\phi}_z(y | s_1, W, \beta, \gamma) = \frac{g_z(y | s_1, w, \beta, \gamma)}{q^{\phi}_z(s_1, W, \beta, \gamma)}.$$

$$q^{\phi}_z(S(1), W, \beta, \gamma) = \int \phi(y, W) g_z(y | S(1), W, \beta, \gamma) dy.$$

Evidence of ACN

ACN in general cannot be directly tested, but the CEP curve and CI of it at $S(1) = 0$ can be used as supportive evidence of ACN.

- If the main effect term for vaccine in the parametric model does not reject the null of equaling zero
- If CI of the CEP when $S(1) = 0$ or $S(1) = S(0)$ covers 0 and is narrow
- If CEP $= 0$ at $S(1) = 0$

these are all evidence in support of ACN.

Testing for WEM and ACS

Under most models we are interested in testing for WEM with a test of the interaction term $S(1) * Z$

- Although for some CEP and structural risk models there is potential variation in the CEP even when there is interaction between vaccination and S(1) in the risk models, this is not generally considered wide effect modification. $CEP* \neq CEP$ even with $\beta_3 = 0$
- Null hypothesis under all parametric models and CEP for WEM is the testing of the interaction term from the risk model being different than zero. $\beta_3 \neq 0$
- ACS has the same null as WEM for the parametric models, the alternative is now just one sided in the direction of lower risk on the vaccine arm. $\beta_3 \geq 0$
- ACS can also be evaluated by plotting the CEP

Inference

For the EML estimation procedure
Bootstrap - Although Pepe and Fleming [1991] did provide a

closed form variance, this requires a non-zero probability of observing S(1) in all subjects.

- There is possible a closed form variance for EML even in our setting, however, it may be too complicated to be worth using when bootstrap works.
- The closed form variance will not solve the efficiency paradox.

Variance Formula Pseudoscore

The closed form variance is given by:

$$\Omega = [\Psi_\theta(\theta_0, F_0^*)]^{-1} var_0 g^0(Y|S(1), W, Y, Z, \delta)[\Psi_\theta(\theta_0, F_0^*)]^{-1}$$

Thus, $VE(s_1) - VE(s_1^{'})$ has asymptotic variance given by:

$$\Omega[VE_{\theta_0}^{'}(y|s_1)]^2 + \Omega[VE_{\theta_0}^{'}(y|s_1^{'})]^2 - 2\Omega[VE_{\theta_0}^{'}(y|s_1)VE_{\theta_0}^{'}(y|s_1^{'})].$$

Although there is a closed form variance for Pseudoscore estimated models, **Currently only bootstrap is implemented in the package**

Comparison of Methods

- CCT or Cross-over designs allow for more efficient estimation and more robust assumption testing
- Full ML most efficient when all models are correct, in BIP alone settings. Computationally taxing particularly with complex models.
- Parametric EML almost as efficient as full ML when all models are correct, in BIP alone settings. Computationally faster than Full ML.

Comparison of Methods: continued

- Semi-parametric EML & Pseudoscore more robust due to flexible modeling of $S(1)|W$ distribution, less efficient when assumed parametric models are correct.
- Parametric or Semi-parametric EML allow for continuous W, Pseudoscore methods allowing for continuous W have been developed in other lines of research, but not for CoP evaluation, yet.
- Power declines with ρ_{SW} rapidly under EML. $\rho_{SW} \geq 0.5$ needed for unbiased estimation with EML under BIP alone
- Power declines with ρ_{SW} more slowly under Pseudoscore, lower ρ_{SW} needed for unbiased estimation.

What settings have these methods been developed for?

Augmentation	S(1)	W	Outcome	Method(s)
BIP alone	Categorical	Categorical	Binary	Any EML Pseudoscore
BIP alone	Any	Categorical	Any	Semi or parametric EML, Pseudoscore
BIP alone	continuous	continuous	Any	Semi or parametric EML
BIP +CPV	Any	Categorical	Any	Pseudoscore
BIP +CPV	Any	continuous	Any	Semi or parametric EML
CCT	Any	Any	Any	glm
sub-sampling of W	Any	Any	Any	EML

Risk Difference Based Summary Statistics: binary setting

In the Binary setting:
Let $\rho_z = Pr(Y(z) = 1)$ and $R(v) = F^{-1}(CEP(s(1)))$ the quantile curve for the risk difference CEP. The area sandwiched between $R(v)$ and $\rho_0 - \rho_1$ can be used to compare candidate CoPs.

$$\begin{aligned} TG = & \int_0^1 |R(v) - (\rho_0 - \rho_1)| dv, \\ STG = & \; TG(t)/[2(\rho_0 - \rho_1)\{1 - \rho_0 + \rho_1\}] \end{aligned}$$

[Gilbert and Hudgens, 2008Huang and Gilbert [2011]]

As shown in Huang and Gilbert [2011] the STG is proportional to the sum of the maximal sensitivity and specificity.

Risk Difference Based Summary Statistics: time-to-event setting

Similarly in the time-to-event setting: Let $\rho_z(t) = Risk^{CDF}(t)$ and $R^t(v) = F^{-1}(CEP(t|s(1)))$

$$TG(t) = \int_0^1 |R^t(v) - (\rho_0(t) - \rho_1(t))|dv,$$
$$STG(t) = TG(t)/[2(\rho_0(t) - \rho_1(t))\{1 - \rho_0(t) + \rho_1(t)\}]$$

This and other summary statistics for time-to-event clinical outcome CoP comparison outlined in Gabriel et al. [2015]

Integrated STG

For a time independent summary one can $\int_t dF(t)$, Gabriel et al. [2015] used a KM estimate of the marginal time distribution:

$$\begin{aligned} \widetilde{TG} = & \int_t \int_0^1 |R^t(v) - (\rho_0(t) - \rho_1(t))| dv dF(t), \\ \widetilde{STG} = & \widetilde{TG} / [2(\rho_0(t) - \rho_1(t))\{1 - \rho_0(t) + \rho_1(t)\}] \end{aligned}$$

STG and STG(t) comparisons

As was shown in Huang and Gilbert [2011] and Gabriel et al. [2015], one can use the null:

$$STG_k - STG_l = 0$$

to test for quality differences between two correlates, l and k.

Both STG must bounded away from zero in order from this to have good properties. A CoP must have some WEM, then comparisons between CoPs can be made via the difference in the summary statistics $STG_k - STG_l$ or $STG_k^t - STG_l^t$ or $\widetilde{STG}_k - \widetilde{STG}_l$ by Wald test based on bootstrap standard errors.

ZEST Example

ZEST Merck Protocol 022 Phase III Vaccine Trial:

Estimated VE = 69.8%, 95% CI 54.1% to 80.6%

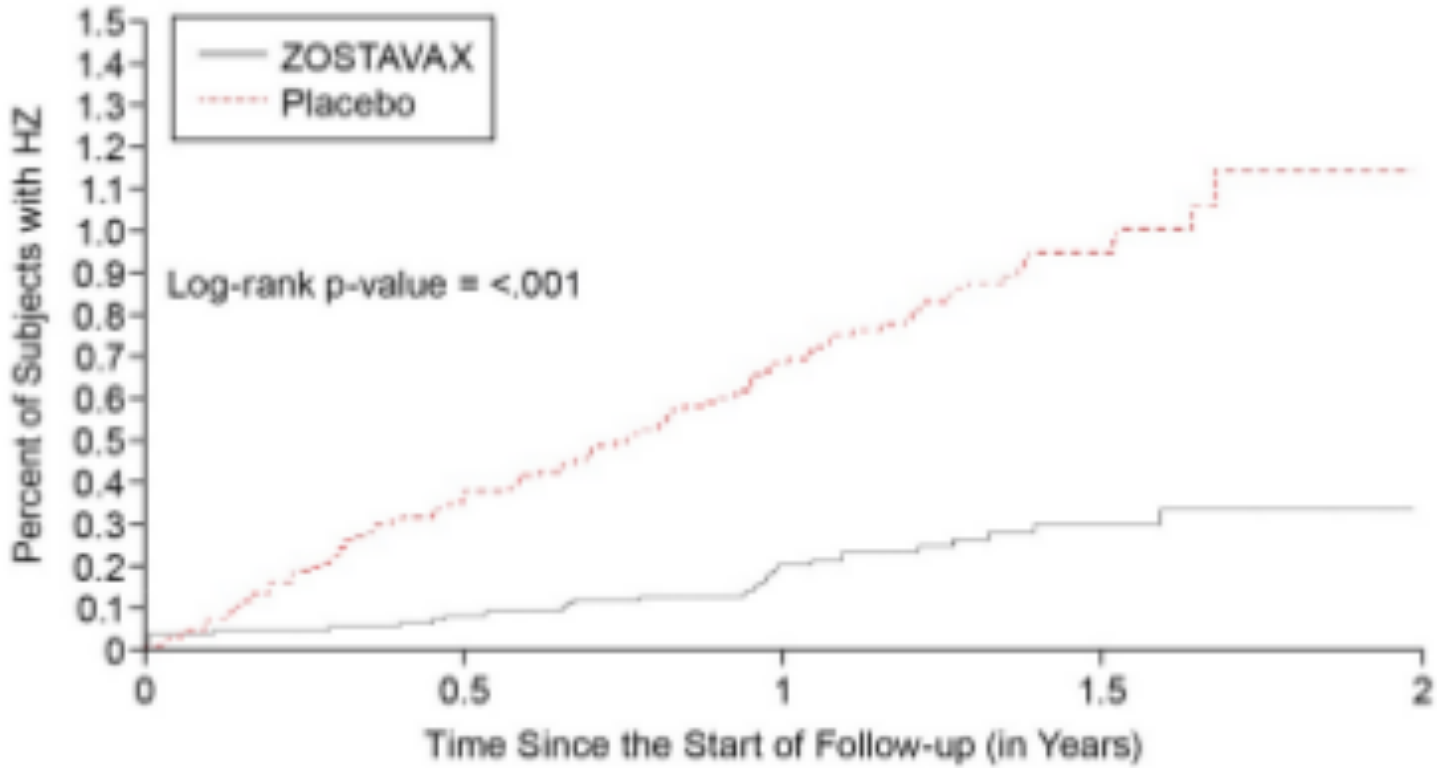


Figure: image

- Phase III ZEST trial: $N = 22,439$ 50-59 year-olds randomized in 1:1 allocation to attenuated zoster vaccine (ZV, Zostavaxő; Merck Sharp & Dohme Corp.) or placebo and followed for 1-2 years for Herpes Zoster (HZ) Sutradhar et al. [2009]

ZEST Example BIP Only

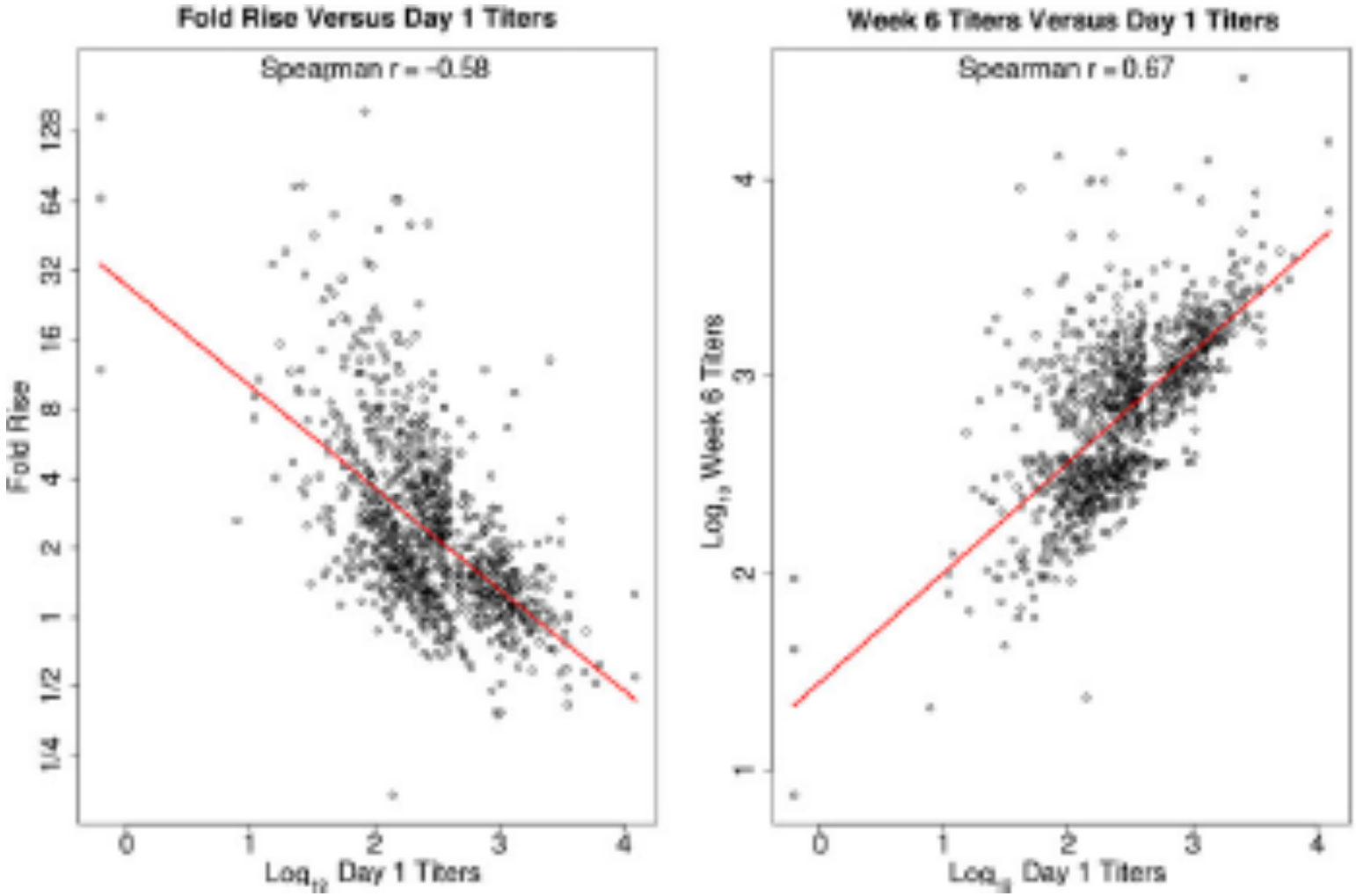


Figure: image

Four Statistical Methods Applied

- Method A: Parametric estimated maximum likelihood (EML) binary clinical endpoint treating the S(1) as continuous Gilbert and Hudgens [2008]
- Method B: Parametric EML time-to-event clinical endpoint treating the S(1) as continuous and allowing for time-variation in VE and surrogate quality and accounting for censoring Gabriel and Gilbert [2014]
- Method C: Non-parametric EML binary clinical endpoint treating the S(1) as categorical Gilbert and Hudgens [2008]
- Method D: Semi-parametric pseudo-score binary clinical endpoint treating the S(1) as continuous Huang and Gilbert [2011]

The results from all methods agree

ZEST Example Fold-Rise

Results of from the different EML methods:

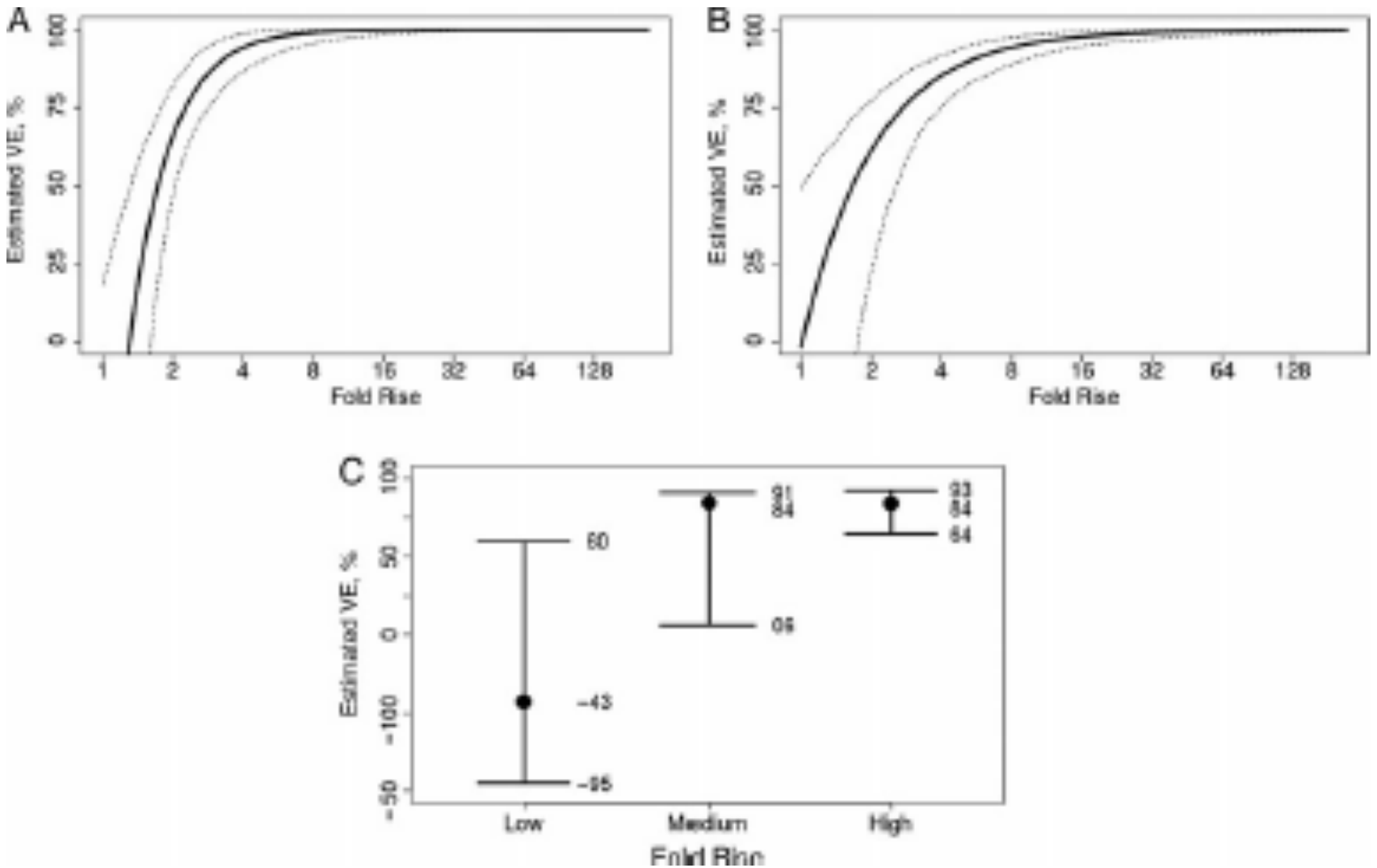


Figure: image

ZEST Example Fold-Rise

Pseudoscore:

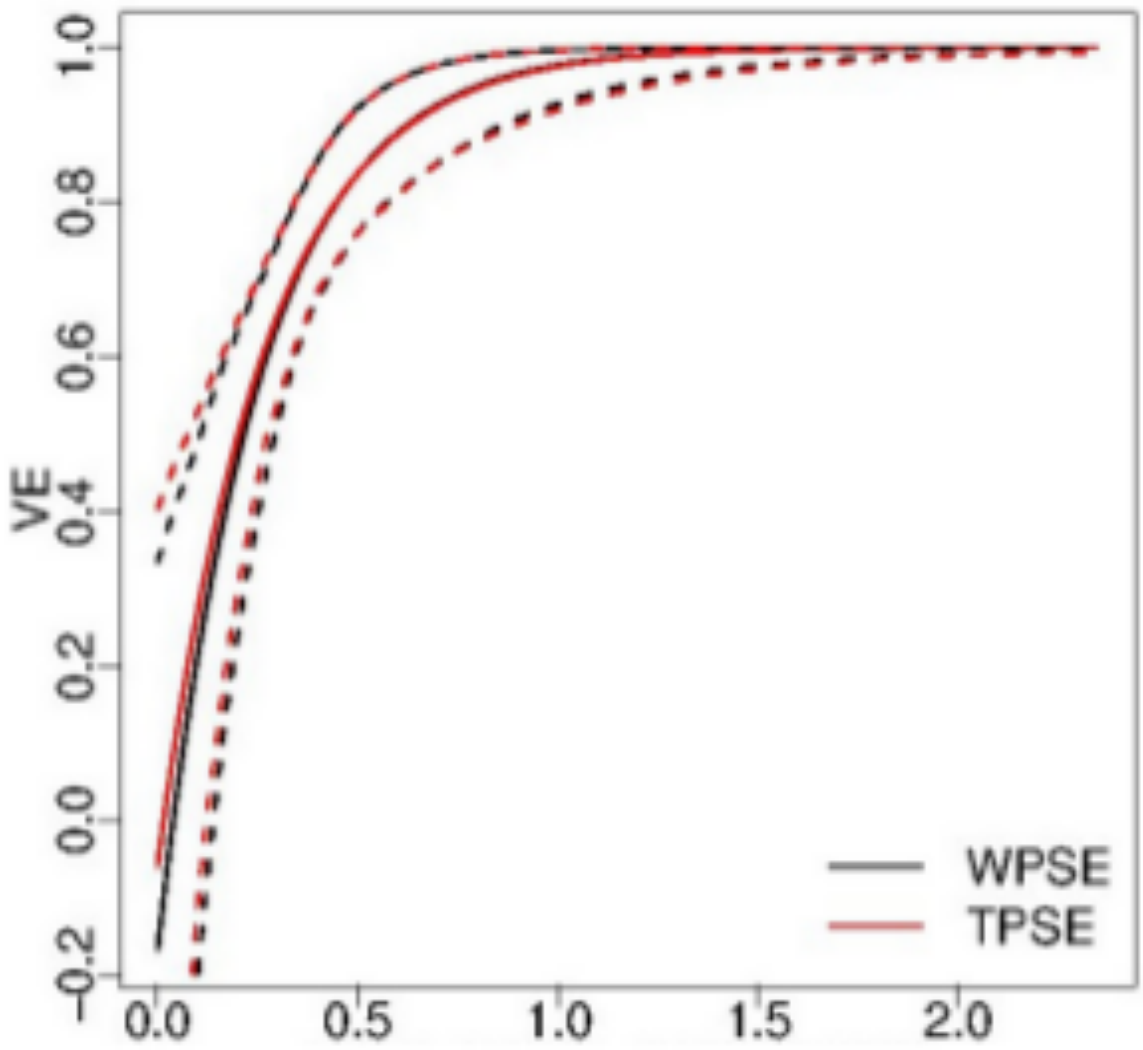


Figure: image

VE Curve: Titer Difference, Continued

The estimated $VE(s_1)$ curves support that titer difference is an excellent CoP

- Method A: P-values < 0.001 for variation in $VE(s_1)$
- Method B: No evidence that $VE(t|s_1)$ varied with time ($p = 0.78$), proportional hazards version of the Weibull model used, p-values < 0.001 for variation of $VE(s_1)$ in s_1
- Methods C: P-values < 0.001 for variation in $VE(s_1)$
- Methods D: Tests not performed

ZEST Example Titer level

Results of from the different EML methods:

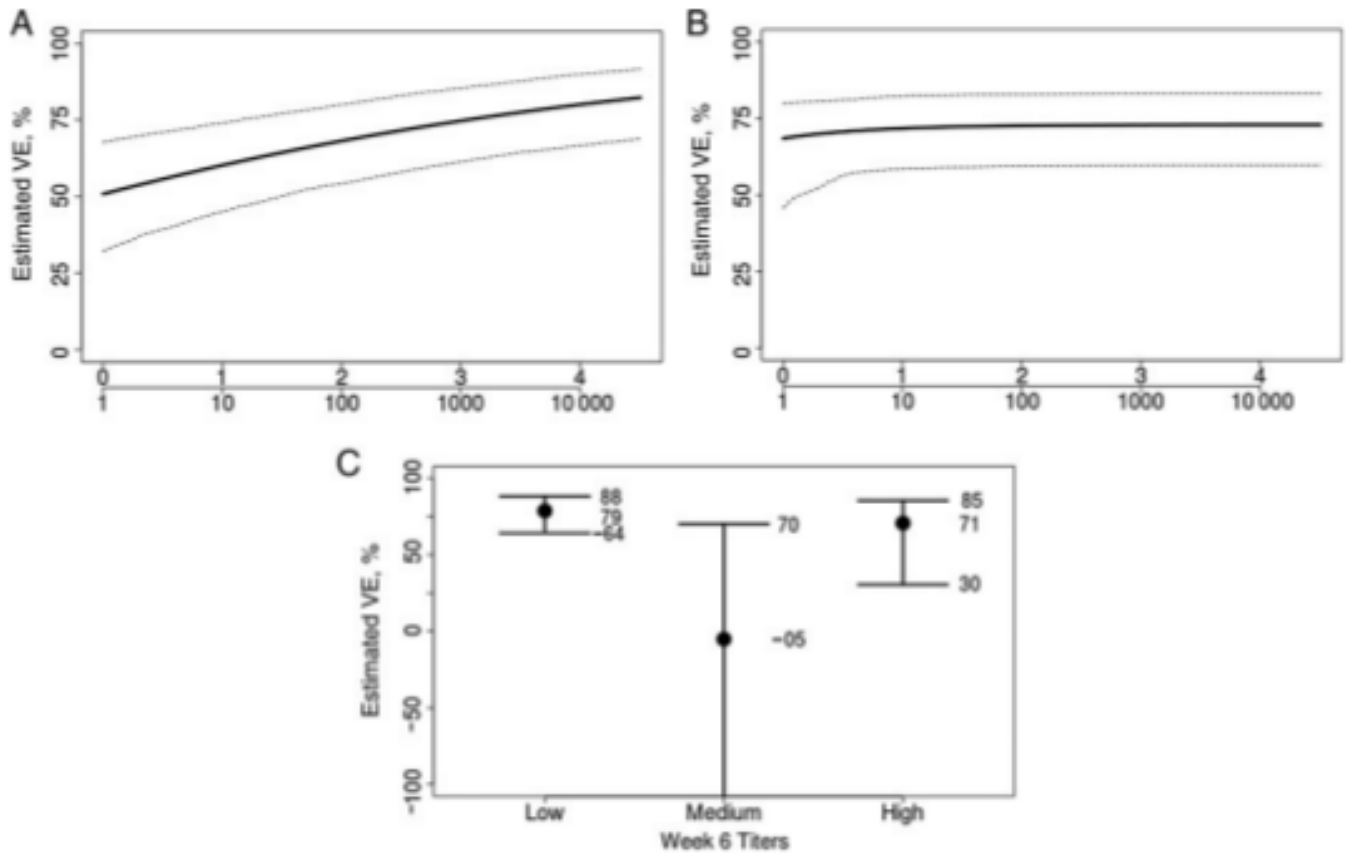


Figure: image

VE Curve: Week 6 Titers, Continued

The estimated $VE(s_1)$ curves varied only slightly over the range of titers, supporting that Week 6 titer is a poor CoP

- Method A: P-value $= 0.91$ for variation in $VE(s_1)$
- Method B: No evidence for time-varying $VE(s_1)$ ($p = 0.55$), proportional hazards version of the Weibull model used, p-value=0.98 for variation in $VE(s_1)$
- Method C: P-value $= 0.82$ for variation in $VE(s_1)$
- Methods D : Tests not performed

Comparison Based on STG

- Test based on difference in standardized total gain Huang and Gilbert [2011] supports titer difference as a superior CoP ($p = 0.045$)
- Only binary test performed
- This test was not really needed, given that there was no evidence of WEM for 6-week titers

Implementation

Implementation of Methods Comparison

- Continuous BIP, in BIP alone designs, only EML or full ML methods have been developed or implemented
- Categorical BIP alone or CPV+BIP designs, Pseudoscore methods have been developed and well as EML and ML, only methods for binary clinical outcome has been published or implemented
- Only Bootstrap variance has been implemented for any of the EML or Pseudoscore methods, although Pseudoscore methods have closed form asymptotic variance
- Although 2-phase sampling methods have been developed in the literature, only methods where W is measured in all subjects, even with sub-sampling of S(1) in the vaccine arm have been implemented in the software

Currently implemented in software

- All three estimation methods with BIP are implemented for binary outcomes
- Saturated Weibull model allows for the characterization of time-varying effects, but can have poor convergence, not implemented in software yet
- Continuous Time-to-event, other outcomes, semi-parametric and fully parametric EML only
- VE and risk-difference curve plotting from any of the estimation methods, with confidence bands
- Tests of WEM based on coefficients from any of the estimation methods
- STG and time-dependent STG estimation implemented

Coming soon in software

- Pseudoscore for Continuous Time-to-event and other outcomes
- Tests for STG and time-dependent STG differences for comparison of candidate CoP
- Integrated STG
- Expanded models for $S(1)|W$

References I

Nilanjan Chatterjee. *A Pseudoscore Estimator for Regression Problems With Two-Phase Sampling*. Doctor of philosophy dissertation, University of Washington; Department of Statistics, Chair: Wellner, Jon and Norman Breslow, 1999.

D Follmann. Augmented designs to assess immune response in vaccine trials. *Biometrics*, 62(4):1161–1169, 2006.

CE Frangakis and DB Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

Erin E Gabriel and Dean Follmann. Augmented trial designs for evaluation of principal surrogates. *Biostatistics*, 17(3):453–467, 2016.

Erin E Gabriel and Peter B Gilbert. Evaluating principal surrogate endpoints with time-to-event data accounting for time-varying treatment efficacy. *Biostatistics*, 15(2):251–265, 2014.

References II

Erin E Gabriel, Michael C Sachs, and Peter B Gilbert. Comparing and combining biomarkers as principle surrogates for time-to-event clinical endpoints. *Statistics in medicine*, 34(3): 381–395, 2015.

PB Gilbert and MG Hudgens. Evaluating candidate principal surrogate endpoints. *Biometrics*, 64(4):1146–1154, 2008.

PB Gilbert, D Grove, E Gabriel, Y Huang, G Gray, SM Hammer, SP Buchbinder, J Kublin, L Corey, and SG Self. A sequential phase 2b trial design for evaluating vaccine efficacy and immune correlates for multiple hiv vaccine regimens. *Statistical Communications in Infectious Diseases*, 3(1), 2011a.

References III

Peter B Gilbert, Douglas Grove, Erin Gabriel, Ying Huang, Glenda Gray, Scott M Hammer, Susan P Buchbinder, James Kublin, Lawrence Corey, and Steven G Self. A sequential phase 2b trial design for evaluating vaccine efficacy and immune correlates for multiple hiv vaccine regimens. *Statistical communications in infectious diseases*, 3(1), 2011b.

PJ Heagerty and MS Pepe. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in us children. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(4):533–551, 1999.

Y Huang and PB Gilbert. Comparing biomarkers as principal surrogate endpoints. *Biometrics*, 2011.

Y. Huang, P. B. Gilbert, and J. Wolfson. Design and estimation for evaluating principal surrogate markers in vaccine trials. *Biometrics*, Under Review(Submitted), 2012.

References IV

MS Pepe and TR Fleming. A nonparametric method for dealing with mismeasured covariate data. *Journal of the American Statistical Association*, 86(413):108–113, 1991.

Santosh C Sutradhar, William W B Wang, Katia Schlienger, Jon E Stek, Jin Xu, Ivan S F Chan, and Jeffrey L Silber. Comparison of the levels of immunogenicity and safety of zostavax in adults 50 to 59 years old and in adults 60 years old or older. *Clin Vaccine Immunol*, 16(5):646–652, May 2009.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.

J Wolfson and PB Gilbert. Statistical identifiability and the surrogate endpoint problem, with application to vaccine trials. *Biometrics*, 66(4):1153–1161, 2010.