Power/Sample Size Calculations for Assessing Correlates of Risk in Clinical Efficacy Trials (Gilbert, Janes, Huang, 2016, *Stat Med*)

Peter Gilbert

Sanofi Pasteur Swiftwater PA

September 24–26, 2018

Outline

Set-Up and Objectives

Assume a randomized vaccine vs. placebo/control vaccine efficacy trial

- **Primary objective:** Assess vaccine efficacy (VE) against an infection or disease endpoint over some follow-up period
- **Secondary objective:** Assess within the vaccine group an immune response biomarker measured at time τ post-enrollment as a correlate of risk (CoR) of the primary study endpoint
  - Assess by case-cohort/case control/two-phase regression analysis, as previously discussed
  - E.g., Cox or logistic regression

Selected Literature on CoR Power Calculations

- Examples of available methods for CoR power calculations within a study group:
  1. Cai J, Zeng D. Sample size/power calculation for case-cohort studies. *Biometrics* 2004; **60**:1015–1024. (Case-cohort)
  2. Dupont WD, Plummer Jr WD. Power and sample size calculations: a review and computer program. *Controlled Clinical Trials* 1990; **11**:116–128. (Case-control)
  3. García-Closas M, Lubin JH. Power and sample size calculations in case control studies of gene-environment interactions: comments on different approaches. *American Journal of Epidemiology* 1999; **149**:689–692. (Two-phase)
  4. Haneuse S, Saegusa T, Lumley T. osDesign: an R package for the analysis, evaluation, and design of two-phase and case-control studies. *Journal of Statistical Software* 2011; **43**:11. (Two-phase)

Power Calculations: Issue 1

- The available approaches typically do not account for the level of overall *VE* and the level of *VE* in biomarker response subgroups, precluding interpretation of the results in terms of correlates of *VE*
- Gilbert, Janes, and Huang (2016, *Stat Med*) developed an approach that accounts for this issue
- Relevant because if the power calculations are based solely on the biomarker-outcome association in the vaccine group, then one could design a case-control study to, say, have 90% power to detect a biomarker-outcome odds ratio of 0.50, but not realize that this power is achieved under a tacit assumption that the endpoint rate is higher in the vaccine arm than the control arm for the subgroup with lowest biomarker responses
  - **Overly optimistic power calculations**

Power Calculations: Issue 1

- By specifying overall *VE* and biomarker-specific *VE* as input parameters, our approach makes transparent in the power calculations the link between the CoR effect size in the vaccine arm and the corresponding difference in biomarker-specific VE
- The biomarker-specific *VE* is the same parameter used in Juraska, Huang, and Gilbert (under review)

Power Calculations: Issue 2

- Our approach also accounts for the component of inter-individual variability of the biomarker that is not biologically relevant
  - E.g., due to technical measurement error of the immunological assay
- Important because the degree of measurement error of the biomarker heavily influences power of the CoR analysis – thus must account for measurement error to obtain accurate power calculations
- Our approach shows how power varies with the user-inputted estimated fraction of the biomarker's variance that is potentially biologically relevant for protection
  - Helps in determining which assays/biomarkers to study as CoRs

Scope of the Power Calculations

- Our approach can be used for a general binary clinical endpoint model with case-cohort, case-control, or two-phase sampling of the biomarker
  - Without replacement or Bernoulli sampling
- We illustrate the approach with the Breslow and Holubkov (1997, *JRSS-B*) logistic regression model and case-control without replacement sampling
- For rare event studies, simulations and applications show that power for the logistic regression model tends to be very similar to that for a Cox regression model
- The power calculations are for a univariate marker that may be censored normal, trichotomous, or dichotomous/binary

Clarifying Our Objective

- Often, the measurement error literature considers the assessment of an **underlying true** biomarker as a CoR
  - Leverage a validation set and/or replicates to correct for bias from measurement error
- Not our objective– we study the association of the **measured/observed** biomarker as a CoR
  - This is what is needed for developing a surrogate endpoint or an obsverable effect modifier
  - The true-biomarker analyses may have objective to gain more insights into potential biological mechanisms of protection – beyond our scope

Set-up and Notation

Z = Indicator of assignment to vaccine (vs. placebo or other control)

W = Baseline covariates

S = Immune response biomarker measured at a fixed time τ post-randomization (continuous, trichotomous, or dichotomous)

T = Time from enrollment until the study endpoint

- Participants are followed for occurrence of the primary clinical study endpoint through time τ_{max}

Set-up and Notation

$Y = I[T \leq \tau_{max}]$ = Indicator of binary outcome of interest

$Y^\tau = I[T \leq \tau]$ = Indicator of binary outcome by time τ

V^τ = Indicator a subject attends the visit at τ

- Subjects observed to be at-risk at τ (that could potentially have immune response biomarkers measured) are those with

$$(1 - Y^\tau)V^\tau = 1$$

Set-up and Notation

$R =$ Indicator that S is measured

$\Delta =$ Indicator that Y is observed, i.e., $\Delta = 0$ if the subject drops out before time τ_{max} and before experiencing the event, and $\Delta = 1$ otherwise

$L = (R(z), R(z)S(z), Y^{\tau}(z), V^{\tau}(z), \Delta(z), \Delta(z)Y(z))$
$=$ Potential outcomes if assigned treatment $z = 0, 1$, where $S(z)$ is defined if and only if $Y^{\tau}(z) = 0$, such that $S(z) = *$ if $Y^{\tau}(z) = 1$

$O \equiv (Z, W, R, RS, Y^{\tau}, V^{\tau}, \Delta, \Delta Y) =$ Observed data for a subject

Set-up and Notation

- The CoR power calculations are based on the N vaccine recipients observed to be at-risk at τ (those with $Z(1 - Y^\tau)V^\tau = 1$), and assess whether $P(Y = 1|S = s_1, Z = 1, Y^\tau = 0)$ varies in s_1
- The CoR power calculations do not need the potential outcomes formulation, as they are based solely on the observable random variables O
  - The potential outcomes are used (only) to define biomarker-specific VE and hence provide a way to relate CoR effect sizes to VE effect sizes

Set-up and Notation

- We assume the vaccine has no effect on the study endpoint before the biomarker sampling time τ:

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

- This assumption is useful by ensuring that the VE parameters measure causal effects of vaccination, and by linking the CoR and correlate of VE parameter types:

$$\begin{aligned} & P(Y = 1 | S = s_1, Z = 1, Y^{\tau} = 0) \\ = \quad & P(Y(1) = 1 | S(1) = s_1, Y^{\tau}(1) = Y^{\tau}(0) = 0) \end{aligned}$$

$$VE(s_1) \equiv 1 - \frac{P(Y(1) = 1 | S(1) = s_1, Y^{\tau}(1) = Y^{\tau}(0) = 0)}{P(Y(0) = 1 | S(1) = s_1, Y^{\tau}(1) = Y^{\tau}(0) = 0)}$$

Set-up and Notation

$$
\begin{aligned}
VE(s_1) &\equiv 1 - \frac{P(Y(1)=1|S(1)=s_1, Y^\tau(1)=Y^\tau(0)=0)}{P(Y(0)=1|S(1)=s_1, Y^\tau(1)=Y^\tau(0)=0)} \\
&= 1 - \frac{P(Y=1|S=s_1, Z=1, Y^\tau=0)}{P(Y(0)=1|S(1)=s_1, Y^\tau(1)=Y^\tau(0)=0)}
\end{aligned}
$$

- Henceforth all unconditional and conditional probabilities of $Y=1$ and $Y(z)=1$ tacitly condition on $Y^\tau(1)=Y^\tau(0)=0$

VE Parameters: Trichotomous Biomarker

- We suppose that each of the N vaccine recipients is in one of three latent/unknown biomarker response subgroups X

  "lower protected" $(X = 0)$, "medium protected" $(X = 1)$, "higher protected" $(X = 2)$ with
  $P_x^{lat} = P(X = x|Z = 1)$ the prevalence of $X = x$

- Define the x-specific outcome risks as

$$risk_z^{lat}(x) \equiv P(Y(z) = 1|X = x) \text{ for } x = 0, 1, 2 \text{ and } z = 0, 1$$

  Thus

$$VE^{lat}(x) = 1 - RR_x^{lat} = 1 - \frac{risk_1^{lat}(x)}{risk_0^{lat}(x)}$$

  for $x = 0, 1, 2$

VE Parameters: Trichotomous Biomarker

- Risks and *VE*'s for subgroups defined by $S(1)$ or by $(X, S(1))$:

$$\begin{aligned} risk_z(s_1) &\equiv P(Y(z) = 1 | S(1) = s_1) \\ risk_z^{lat}(x, s_1) &\equiv P(Y(z) = 1 | X = x, S(1) = s_1) \end{aligned}$$

for $x = 0, 1, 2, s_1 = 0, 1, 2$ and $z = 0, 1$, and

$$\begin{aligned} VE(s_1) &\equiv 1 - RR(s_1) = 1 - \frac{risk_1(s_1)}{risk_0(s_1)} \\ VE^{lat}(x, s_1) &\equiv 1 - RR^{lat}(x, s_1) = 1 - \frac{risk_1^{lat}(x, s_1)}{risk_0^{lat}(x, s_1)} \end{aligned}$$

VE Parameters: Trichotomous Biomarker

- The observed biomarker response $s_1 = 0$ represents a "low" response and $s_1 = 2$ a higher response, with $s_1 = 1$ an intermediate response
  - E.g., $s_1 = 0$ could be negative/non-response/below LLOQ and $s_1 = 2$ a response above a pre-specified putative correlate of protection threshold
- If S were measured without error, then $X = S$ such that $VE(s_1) = VE^{lat}(x)$ and the latent variable formulation would not be needed
  - We use it to allow measurement error to create differences in $VE(s_1)$ versus $VE^{lat}(x, s_1)$, with greater differences for noisier biomarkers

Accounting for Measurement Error in the Biomarker

Protection-related sensitivity/specificity and false positive/negative parameters:

$$Sens \equiv P(S = 2|X = 2), \quad Spec \equiv P(S = 0|X = 0),$$
$$FP^0 \equiv P(S = 2|X = 0), \quad FN^2 \equiv P(S = 0|X = 2),$$
$$FP^1 \equiv P(S = 0|X = 1), \quad FN^1 \equiv P(S = 0|X = 1)$$

Define $P_0 = P(S = 0|Z = 1)$ and $P_2 = P(S = 2|Z = 1)$

$$P_0 = Spec * P_0^{lat} + FN^1 * P_1^{lat} + FN^2 * P_2^{lat},$$
$$P_2 = Sens * P_2^{lat} + FP^1 * P_1^{lat} + FP^0 * P_0^{lat}$$

The perfectly measured (noise-free) biomarker has $Sens = Spec = 1$ and $FP^0 = FN^1 = FP^1 = FN^1 = 0$, implying $P_0 = P_0^{lat}$ and $P_2 = P_2^{lat}$

- (I.e., the proportions of vaccine recipients with $S = 0, 1, 2$ biomarker responses equal the proportions with $X = 0, 1, 2$ levels of protection, and these subgroups are identical)

Two Approaches to Trichotomous Marker Power Calculations

- Approach 1 inputs ($Sens$, $Spec$, FP^0, FN^2, FP^1, FN^1)
- Approach 2 uses a measurement error model for a normally distributed continuous-readout biomarker S^* and defines the values of S by

  $S = 0$ if $S^* \leq \theta_0$, $S = 2$ if $S^* > \theta_2$, and $S = 1$ otherwise,

  with θ_0 and θ_2 two constants with $\theta_0 < \theta_2$ (that are determined by specification of σ^2_{obs}, ρ, P_0, P_2, P_0^{lat}, P_2^{lat})
  - **Classical measurement error model:** Assume a normally distributed latent 'true' biomarker X^*, and link S^* to X^* by the model:

    $$S^* = X^* + e, \quad X^* \sim N(0, \sigma^2_{tr}), \quad e \sim N(0, \sigma^2_e), \tag{1}$$

    with X^* independent of e, implying
    $S^* \sim N(0, \sigma^2_{obs})$ with $\sigma^2_{obs} = \sigma^2_{tr} + \sigma^2_e$

Approach 2 to Trichotomous Marker Power Calculations

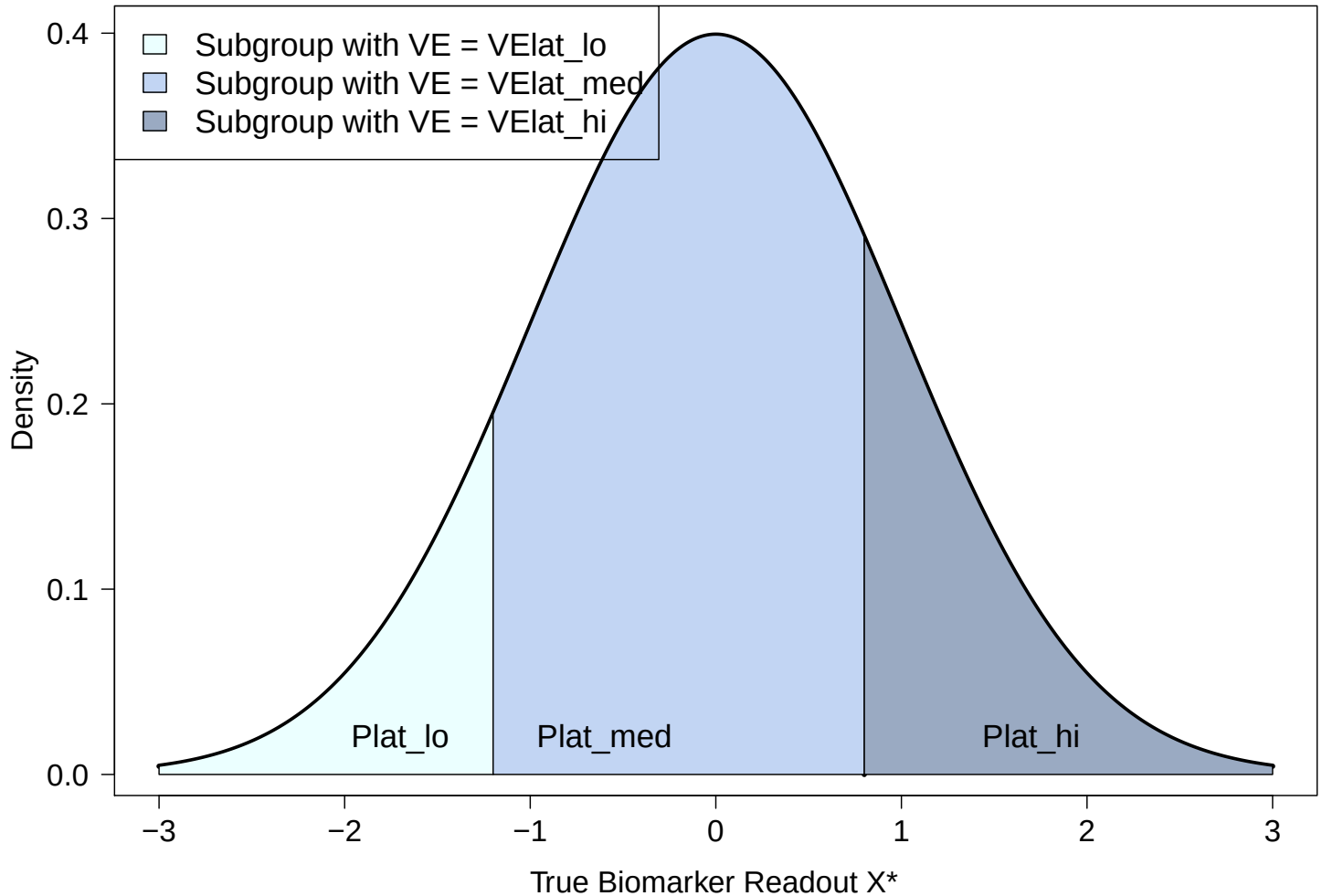
- In the classical measurement error model

$$\rho \equiv 1 - \sigma_e^2/\sigma_{obs}^2$$

  is the fraction of the variability of S^* that is potentially biologically relevant for protection, and is specified to reflect the quality of the biomarker

- The 'true' trichotomous biomarker X is defined by two percentiles of X^* that are determined mathematically by the model and the two percentiles θ_0 and θ_2

Illustration of the Approach 2 Set-Up



Special Case of a Binary Biomarker

- A trichotomous marker may generally be more useful, because it is hard to find a single threshold that majorly discriminates risk
  - With two thresholds, one is for low risk and the other is for high risk
- Nevertheless the power code applies to a binary biomarker
  - Set $P_1^{lat} = P_1 = 0$, in which case only the *Sens* and *Spec* parameters are needed for the calculations [because $FN^2 = 1 - Sens$ and $FP^0 = 1 - Spec$]
  - The R code handles the dichotomous biomarker as a special case

VE Parameters and Model: Continuous Biomarker

- Similar formulation, where now the latent subgroups are defined by the true unobservable biomarker X^* in the classical measurement error model (1)

$$VE^{lat}(x^*) \equiv 1 - \frac{risk_1^{lat}(x^*)}{risk_0^{lat}(x^*)}, \quad VE(s_1) \equiv 1 - \frac{risk_1(s_1)}{risk_0(s_1)},$$

with

$$risk_z^{lat}(x^*) \equiv P(Y(z) = 1 | X^*(1) = x^*)$$

and

$$risk_z(s_1) \equiv P(Y(z) = 1 | S^*(1) = s_1)$$

for x^* and s_1 varying over the continuous support of $X^*(1)$ and $S^*(1)$, respectively

VE Parameters and Model: Continuous Biomarker

- Specify a fraction $P^{lat}_{lowestVE}$ of subjects with the lowest $X^*(1)$ values $\leq \nu$ to all have the same specified lowest level of efficacy VE_{lowest}:

$$VE_{lowest} \equiv VE^{lat}(X^*(1) \leq \nu) = 1 - \frac{risk^{lat}_1(\nu)}{risk^{lat}_0(\nu)} \tag{2}$$

  E.g., Set VE_{lowest} to 0, and interpret $P^{lat}_{lowestVE}$ as the fraction of subjects without a positive vaccine-induced immune response, reflecting an assumption that non-take = zero protection

- The constant $\nu = \sqrt{\rho}\sigma_{obs}\Phi^{-1}(P^{lat}_{lowestVE})$, where $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cdf

CoR Models: Continuous Biomarker

For $x^* \leq \nu$, $risk_1^{lat}(x^*)$ is modeled as a constant:

$$risk_1^{lat}(x^*) = (1 - VE_{lowest})risk_0^{lat}(\nu) \qquad \text{for} \quad x^* \leq \nu, \tag{3}$$

and, for $x^* > \nu$, $risk_1^{lat}(x^*)$ is modeled via a logistic regression model

$$logit(risk_1^{lat}(x^*)) = \alpha^{lat} + \beta^{lat} x^* \qquad \text{for} \quad x^* > \nu \tag{4}$$

Using model (3)–(4) that specifies a lowest value of VE is useful because the alternative simpler model that would specify (4) for all x^* would force $VE(x)$ to be negative for the lowest values of x^*

CoR Models: Continuous Biomarker

- Model (3)–(4) combined with (1) and the assumption $risk_0^{lat}(x) = risk_0$ (as stated below) implies that

$$VE = 1 - \frac{1}{risk_0}\left[P_{lowestVE}^{lat}\, risk_1^{lat}(\nu) + \int_{\nu}^{\infty} logit^{-1}(\alpha^{lat} + \beta^{lat} x^*)\phi(x^*/\sqrt{\rho}\sigma_{obs})dx^*\right]$$

  where $\phi(\cdot)$ is the standard normal pdf

- This formula is used for implementing the power calculations (Michal Juraska)

CoR Hypotheses and Parameters of Interest

- **Objective:** To assess an immune response biomarker at τ in at-risk vaccine recipients at τ as a CoR of the study endpoint
- Trichotomous S: Test

$$H_0 : risk_1(s_1 = 2) = risk_1(s_1 = 1) = risk_1(s_1 = 0) \quad \text{vs.}$$
$$H_1 : risk_1(s_1 = 2) \leq risk_1(s_1 = 1) \leq risk_1(s_1 = 0)$$

with '$<$' for at least one of the two inequalities in H_1

Continuous S^*: Test

$$H_0 : risk_1(s_1) \quad \text{is constant in} \quad s_1 \quad \text{vs.}$$
$$H_1 : risk_1(s_1) \leq risk_1(s_1') \quad \text{for all} \quad s_1' < s_1$$

with '$<$' for some $s_1' < s_1$

Correlate of Risk (CoR) Hypotheses and Estimands of Interest

- While for data analysis 2-sided tests would typically be used, the power calculations are clearer to interpret by testing for the 1-sided alternative H_1 of lower clinical risk in vaccine recipients with increasing s_1
- The code uses 1-sided Wald tests

Methods of Analysis with Bernoulli and Without Replacement Sampling

- Two main approaches to selecting the subset of subjects for whom to measure the biomarkers

  Prospective case-cohort: Select a simple or stratified random sample from all randomized vaccine recipients, and augment the sample with all study endpoint cases that were not randomly sampled;

  Retrospective case-control or 2-phase sampling: Conditional on final case status and possibly a discrete stratification covariate measured in all subjects, select afixed number of vaccine recipients (or random sample) from each case status $\times$ covariate stratum

- The power calculations consider both approaches

Identifiability Assumptions

Recall

$$L = (R(z), R(z)S(z), Y^{\tau}(z), V^{\tau}(z), \Delta(z), \Delta(z)Y(z))$$

and

$$O = (Z, W, R, RS, Y^{\tau}, V^{\tau}, \Delta, \Delta Y)$$

Assumptions:

- iid random variables (L_i, X_i^*, X_i) and (O_i, X_i^*, X_i) for $i = 1, \ldots, N$
- SUTVA (Consistency + No interference)
- Ignorable treatment assignment $(Z \perp L|W)$
- Equal early clinical risk $(P(Y^{\tau}(1) = Y^{\tau}(0)) = 1)$

Identifiability Assumptions

Assumptions, Continued:

- Random censoring ($Y(z) \perp \Delta(z)$ for $z = 0, 1$)
- S is missing at random (MAR): R depends only on the observed data O
- After accounting for the latent category (and any baseline covariates W included in the CoR analysis) the measured biomarker in vaccine recipients does not affect risk, i.e.,

$$risk_1^{lat}(x^*, s_1) \equiv P(Y(1) = 1 | X^*(1) = x^*, S^*(1) = s_1) = risk_1^{lat}(x^*)$$

for all s_1 and x^*, and similarly for risk as a function of trichotomous X and S
(so-called "surrogate assumption")

Identifiability Assumptions

Assumptions, Continued:

- Scenario/assumption for power calculations:

$$risk_0^{lat}(x^*, s_1) = risk_0(s_1) = risk_0$$

  for all s_1 and x^* and similarly for risk as a function of trichotomous X and S
- $risk_0(x^*, s_1)$ and $risk_0(s_1)$ are not identifiable (because $S(1)$ is a counterfactual random variable for subjects assigned $Z = 0$), and power calculations could be conducted under many scenarios for these functions
  - The special case is very helpful for power calculations because $risk_0$ can be specified based on the observed or projected incidence in the trial
- Because the CoR data analysis itself would control for known baseline prognostic factors W, the scenario in which the power calculations are accurate is

$$risk_0^{lat}(x^*, s_1) = risk_0(s_1) = risk_0$$

  after conditioning on W

CoR Effect Sizes RR_t and RR_c as a Function of Vaccine Efficacies

- Analysis of the vaccine group data provides inference on the relative risks

$$RR_t \equiv \frac{risk_1(2)}{risk_1(0)}$$

for a trichotomous biomarker and

$$RR_c \equiv \frac{risk_1(s_1)}{risk_1(s_1 - 1)}$$

for a continuous biomarker

- RR_t and RR_c are the user-specified "CoR effect sizes" of the power calculations

CoR Effect Sizes RR_t and RR_c as a Function of VEs

- Inference on RR_t and RR_c makes inference on the population of all vaccine recipients at-risk for the study endpoint at τ
- RR_t and RR_c are identified from the assumptions and the observed data measured from the subset of vaccine recipients with $R = 1$
- Therefore the power calculations for testing H_0 can be based on the set of vaccine recipients with S (or S^*) measured at τ

CoR Effect Sizes RR_t and RR_c as a Function of VEs

- For a trichotomous biomarker, RR_t is linked to the latent VE parameters via:

$$\begin{aligned} RR_t &= \frac{risk_1(2)}{risk_1(0)} \\ &= \frac{\sum_{x=0}^{2} RR_x^{lat} P(X = x|S = 2)}{\sum_{x=0}^{2} RR_x^{lat} P(X = x|S = 0)} \end{aligned}$$

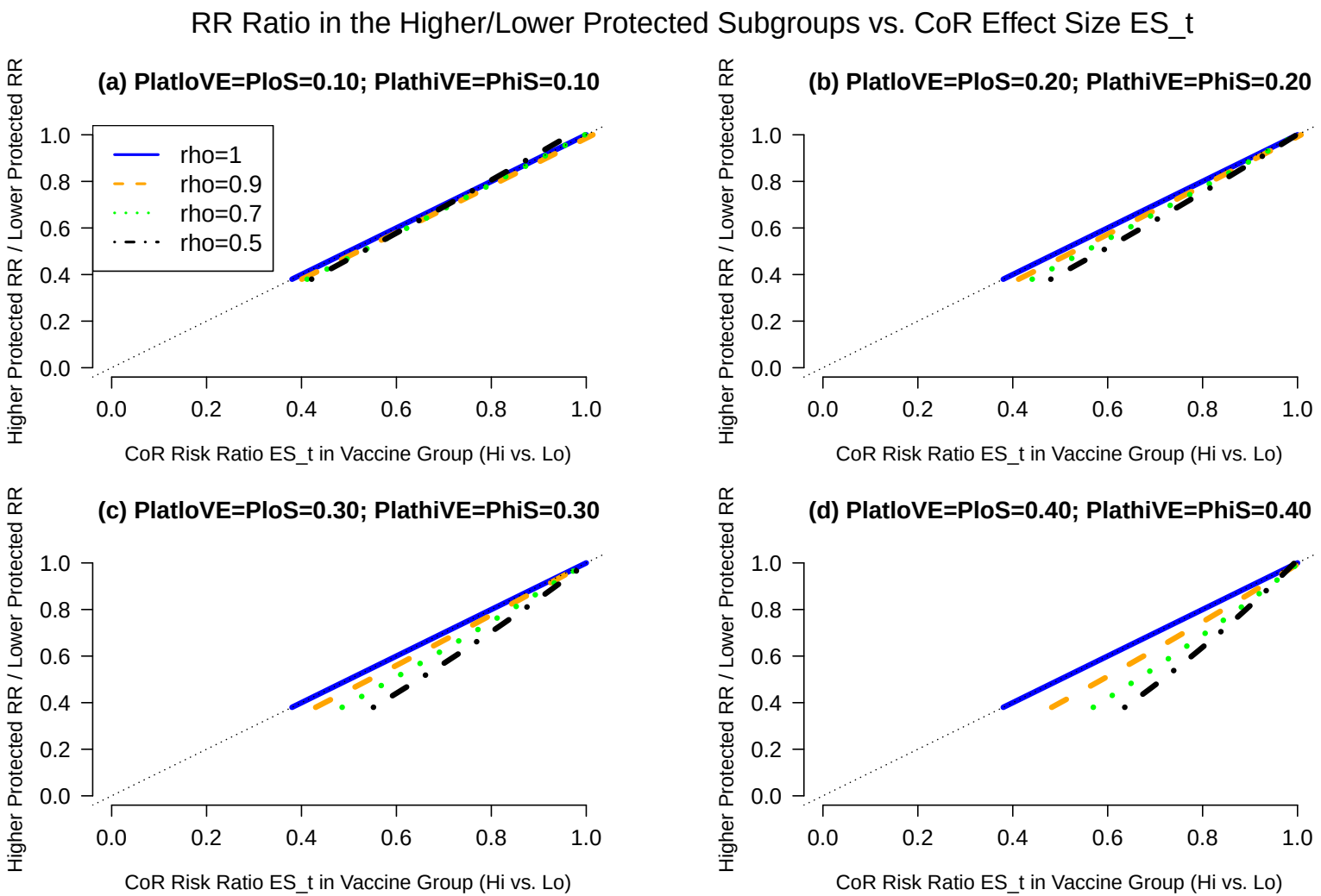
  This formula makes the estimable RR_t interpretable in terms of a gradient in VEs, where

$$RR_t = RR_2^{lat} / RR_0^{lat}$$

  for a noise-free biomarker with
  $1 - Sens = 1 - Spec = FP^0 = FP^1 = FN^2 = FN^1 = 0$

- Otherwise, if $\rho < 1$, then RR_t is closer to 1.0 than RR_2^{lat} / RR_0^{lat}

Interpretation of RR_t ($RR_t = ES_t$ in the Figure)



Overall VE = 0.31

VE_lower varies from 0.31 to 0 as VE_higher varies from 0.31 to 0.62

CoR Effect Sizes RR_t and RR_c as a Function of Vaccine Efficacies

- For a continuous biomarker S^* following the classic measurement error model (1), RR_c is linked to the latent VE parameters via an equation that depends on s_1
- Because $RR_c = \frac{risk_1(s_1)}{risk_1(s_1-1)}$ depends on s_1, it is not particularly useful to index power calculations by RR_c
- Instead, we interpret RR_c as the effect size for a noise-free biomarker ($\rho = 1$)
- Under the logistic model, RR_c is the relative risk per standard deviation increase in X^* in the region above ν, where we use the approximation of a relative risk by an odds ratio

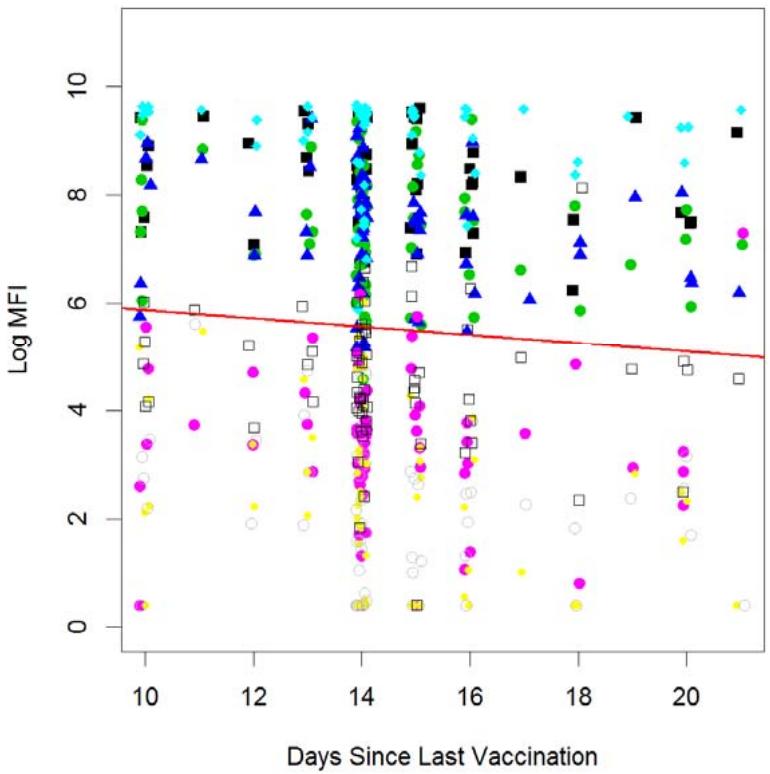
Estimates of ρ for BAMA [Tomaras Lab]: Week 26 Responses from HVTN 096 and RV144

Isotype	Antigen	ρ*
IgG	A244 gp 120 gDneg/293F/mon	0.97
	gp70_B.CaseA2 V1V2169K	0.97
	gp70_B.CaseA_V1_V2	0.95
	AE.A244 V1V2 Tags/293F	0.91
IgG3	A244 gp 120 gDneg/293F/mon	0.99
	gp70_B.CaseA2 V1V2169K	0.91
	gp70_B.CaseA_V1_V2	0.94
	AE.A244 V1V2 Tags/293F	0.98

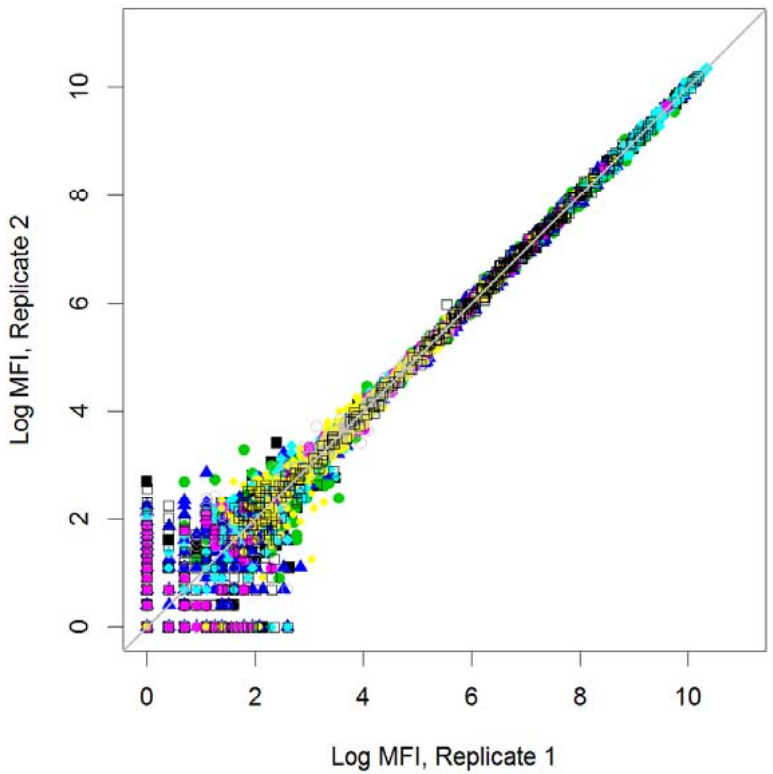
$$*\rho = 1 - \frac{\text{Variance from within vaccinee replicates + Variance from days since Mo 6 vaccination}}{\text{Total inter-vaccinee variance of response}}$$

Two Sources of Protection-Irrelevant Variability in BAMA Week 26 Responses

Variability due to variation in days since last vaccination [HVTN 096]*



Variability across replicates for the same vaccinee [RV144]



*Each point is for an individual vaccine recipient
Each vaccine type × isotype × antigen is plotted using a different symbol.

Summary

- The power calculation methods Gilbert, Janes, and Huang (2016) apply for assessing in the vaccine group of a VE trial a fixed time biomarker (continuous normal, trichotomous, or binary) as a CoR of subsequent occurrence of a study endpoint
- Focused on the two issues of interpreting results relative to VE and biomarker measurement error
  - Indexing the power calculations by the degree of measurement error is useful for selecting assays/biomarkers with adequate power for inclusion in CoR studies
- While the methods are for CoR power calculations, they also apply for correlate of VE power calculations under the strong assumption that outcome risk in placebo recipients is independent of immune response if vaccinated, after conditioning on baseline covariates W
  - (i.e., $risk_0^{lat}(x^*, s_1) = risk_0(s_1) = risk_0$ conditional on W)

Utility of Calculations for a Trichotomous Biomarker

- Dividing vaccine recipients into three biomarker-response subgroups has broad utility
  - **Example application 1:** $S = 0$ is response below LLOQ (negative); $S = 2$ is response above a specified threshold *thresh*; and the CoR analysis could study a series of trichotomous biomarkers varying *thresh*
- In our systems vaccinology era (transcriptomics, metabolomics, etc.), vaccinated subgroups of interest may be defined by signatures derived from high-dimensional data analysis (e.g., by hierarchical clustering)
  - **Example application 2:** $S = 0$ is a signature of putative non-protection; $S = 2$ is a signature of putative protection

Limitations of the Methods and Code

- In practice CoR analysis should adjust for baseline pathogen exposure variables, yet the current code does not consider covariate adjustment
- The current code only considers the scenario/assumption that $risk_0^{lat}(x^*, s_1) = risk_0(s_1) = risk_0$
  - While it may be the most important single scenario to study, it easily could fail
- The method and code restrict to a univariate biomarker
  - In practice multivariate biomarker analysis is at least as interesting

Partnership of Statistical and Laboratory Science

The approach emphasizes study of how power depends on the signal-to-noise ratio of an immune response biomarker

- To be used effectively, partnership with lab scientists is needed to estimate ρ (or at least upper bound it)
- The approach only considered a few measurement error models such as the classical additive measurement error model – in practice it is recommmended to work with laboratory scientists to build a maximally accurate model