Statistical Methods for Evaluating Correlates of Risk

Peter Gilbert

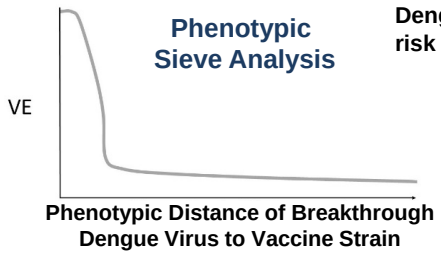
Sanofi Pasteur Swiftwater PA

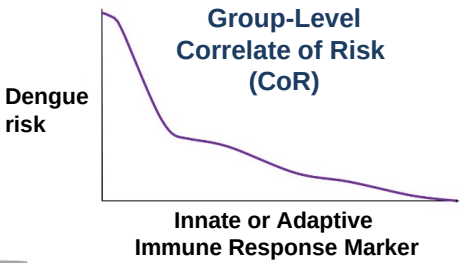
September 24–26, 2018

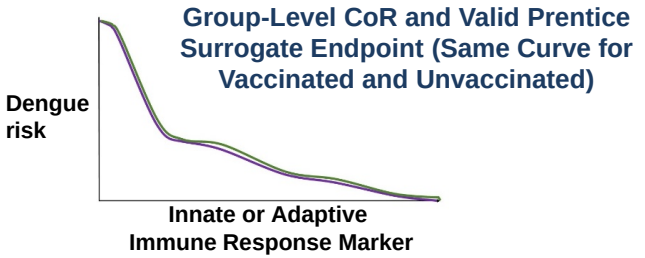
Outline

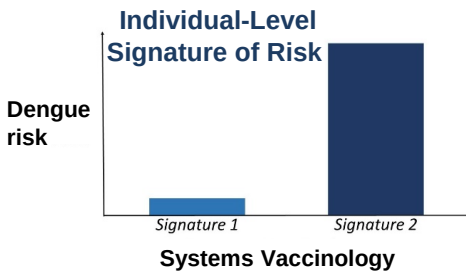
1. Introduction to CoR analysis
2. CoR analysis with a Cox model
   - Fixed-time CoR
   - Time-dependent CoR
3. CoR analysis (marker at a single fixed time point) with a logistic regression model
4. Selected issues
   - Marker sampling design
   - Marker measurement error
5. Improved CoR methods (Breslow et al., 2009; Rose and van der Laan, 2011)

Context: Eight Frameworks for Assessing Statistical Correlates of Vaccine Protection

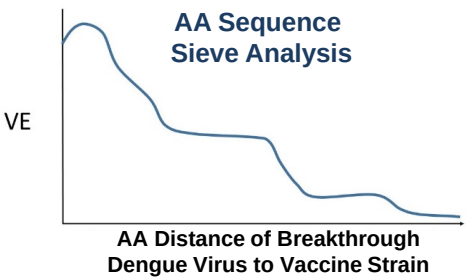


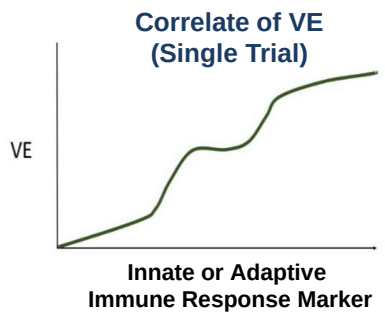




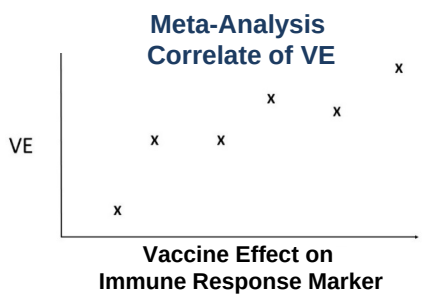


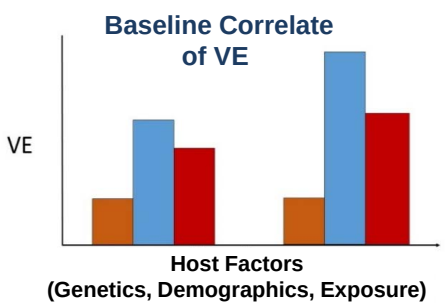
Multiple Types of "Correlates of Vaccine Protection"





AA





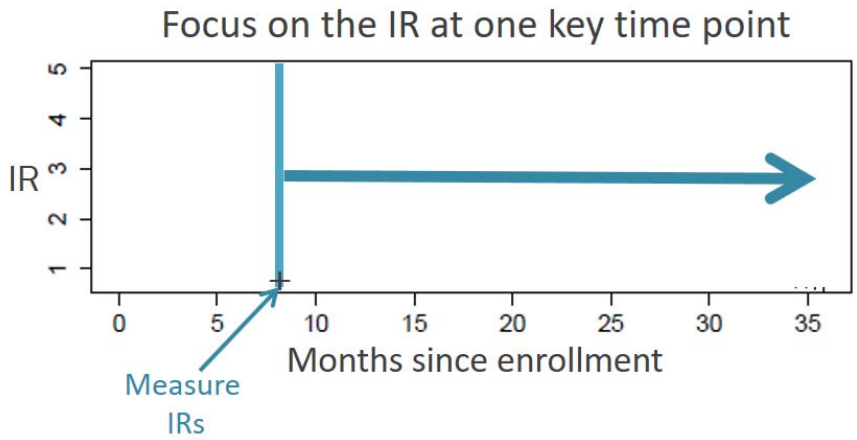
Applications of Statistical Correlates of Vaccine Protection

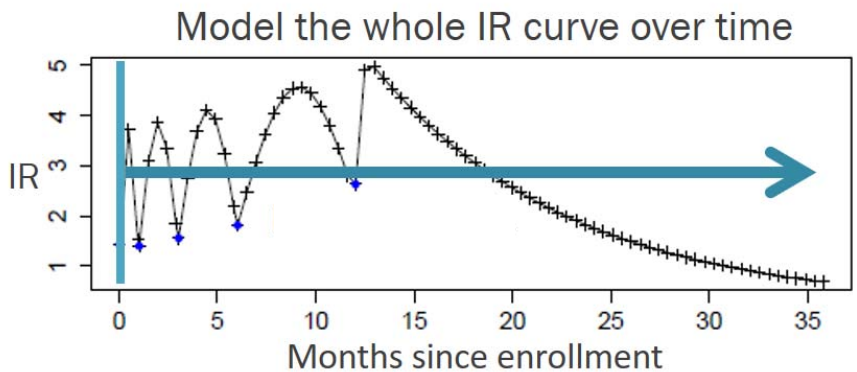
- Generate hypotheses about mechanistic correlates of protection that can be further evaluated
- Guide iterative development of vaccines between basic and clinical research
  - Refine vaccine regimens
- Guide regulatory decisions
- Guide immunization policy
- Model public-health impact and cost-effectiveness
- Shorten trials and reduce costs
- Bridge vaccine efficacy to new settings

Two Types of Correlates of Risk with respect to Time

1. **Fixed-time correlate:** IR marker measured at a **fixed time point** post-vaccination that associates with outcome or with VE against the outcome
   - ***Purpose:*** Practicable predictor of risk or VE / surrogate endpoint

2. **Time-dependent correlate:** IR marker **measured longitudinally** whose current level associates with instantaneous incidence of outcome or with VE against outcome
   - ***Purpose:*** Generate insights into mechanistic correlates of protection



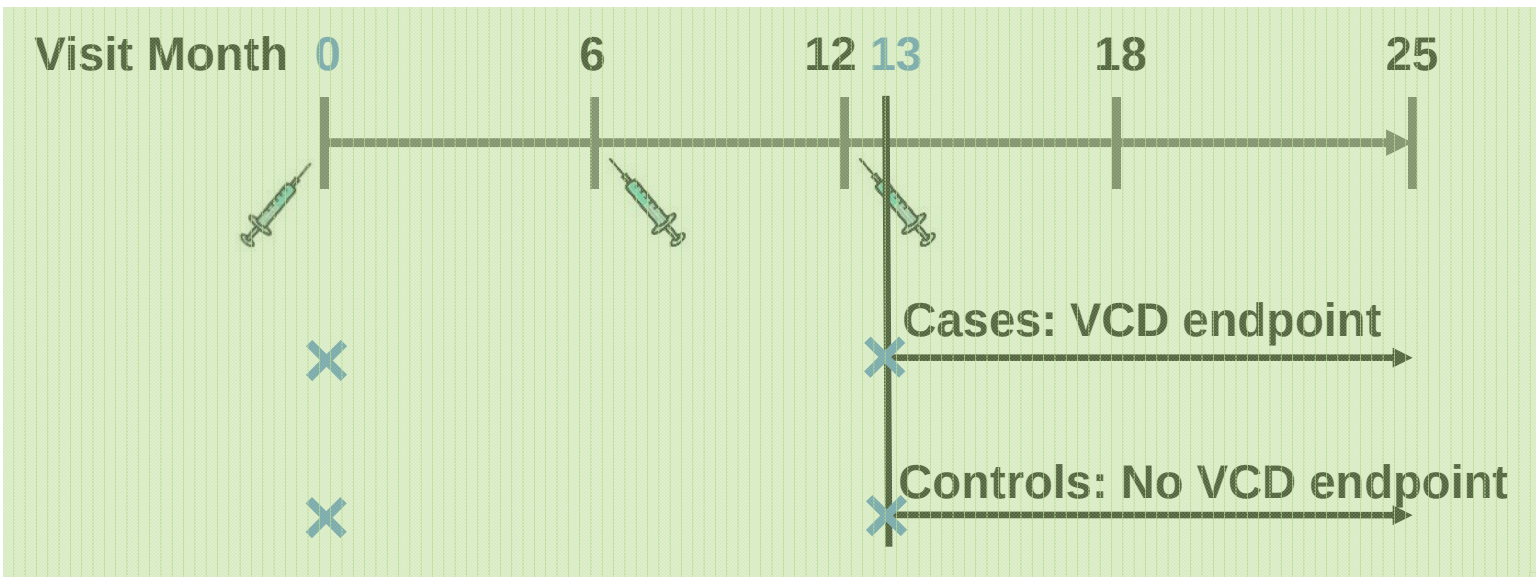


Prospective Cohort Study Sub-Sampling Design Nomenclature

- Terms used: case-cohort, case-control, 2-phase sampling
  - Case-cohort sampling originally meant taking a Bernoulli random sample of subjects at study entry for marker measurements (the "sub-cohort"), and also measuring the markers in all disease cases (Prentice, 1986, *Biometrika*)
  - Case-control sampling is Bernoulli or without replacement sampling done separately for observed diseased cases and observed non-diseased controls (retrospective sampling)
  - 2-phase sampling is the generalization of case-control sampling that samples within discrete levels of a covariate as well as within case and control strata (Breslow et al., 2009, *AJE*, *Stat Biosciences*)
  - Source of confusion: Some papers allow the term case-cohort to include retrospective sampling
- These slides use the original meaning of the term case-cohort

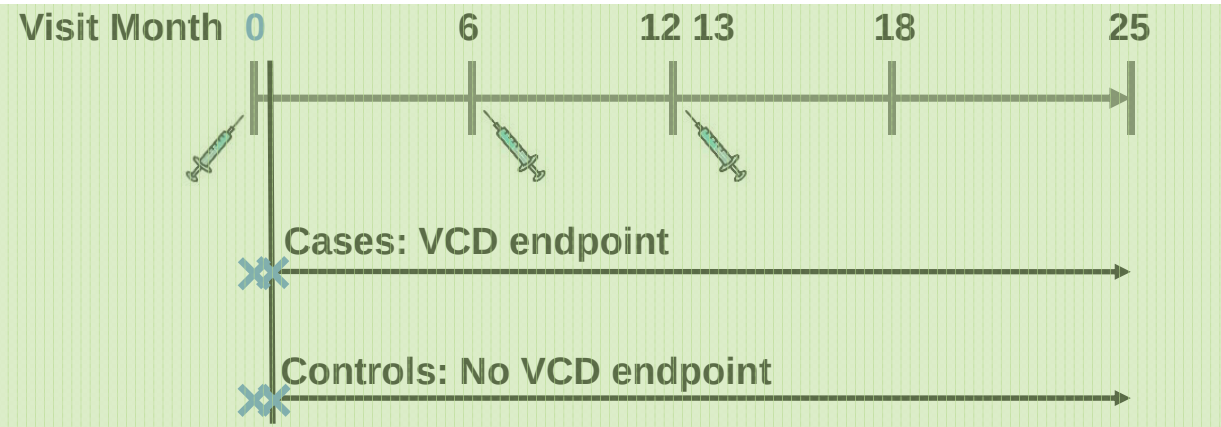
Assessing CoRs in Vaccine Recipients: Case-Control Design (Fixed-Time Adaptive CoR) (Design of CYD14/15)

- **Objective:** Develop Month 13 correlates of risk (CoRs) and protection (CoPs) against symptomatic VCD through Month 25
- Measure immune responses **(at Month 0, 13)** from all vaccinees with VCD after Month 13 through Month 25 and from a random sample of vaccinees free of VCD through Month 25



Assessing CoRs in Vaccine Recipients: Case-Control Design (Fixed-Time Innate CoR) (Hypothetical)

- **Objective:** Develop Day 1 correlates of risk (CoRs) and protection (CoPs) against symptomatic VCD through Month 25
- Measure immune responses **(at Day 0, Day 1)** from all vaccinees with VCD after Day 1 through Month 25



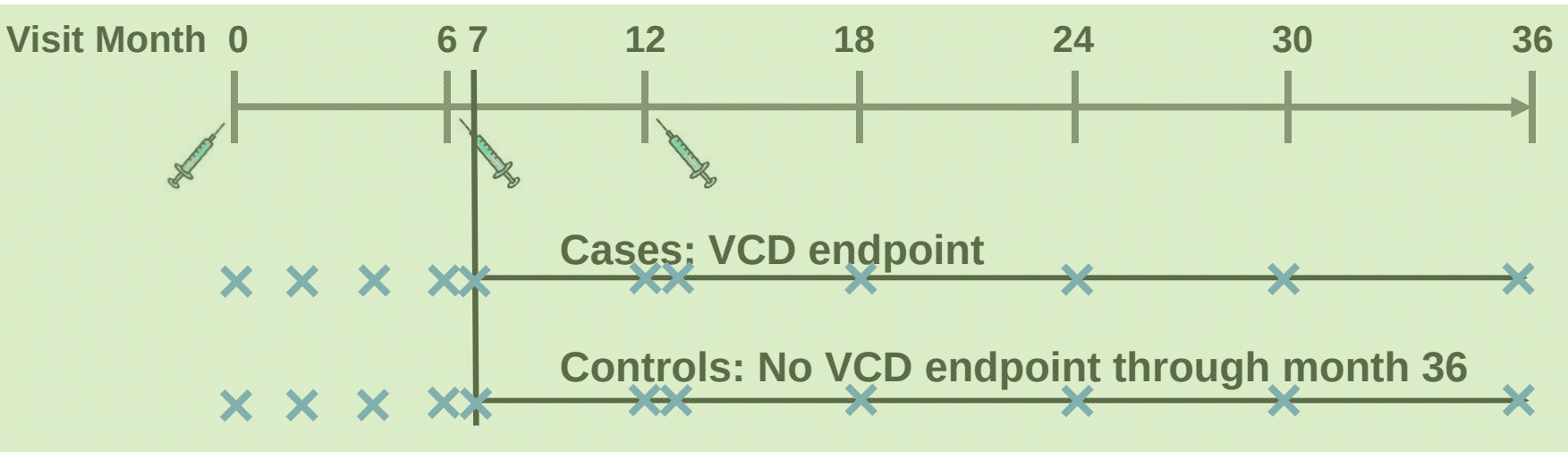
- Systems vaccinology: gene expression, cell sub-populations, etc.*

**E.g., Bali Pulendran (YFV, influenza), John Tsang (influenza, Cell 2014), Rafick Sekaly (HIV vaccines), Amy Chung, Galit Alter (HIV vaccines, Cell 2015) Zak, Andersen-Nissen, de Rosa et al. McElrath (HIV vaccines, PNAS, 2012) Gottardo, McElrath et al. within the HIP-C (Malaria vaccines, other vaccines) Andersen-Nissen, Fiore-Gartland et al. on HVTN 097 (HIV vaccines) and HVTN 602 (TB vaccines)*

FRED HUTCH

Assessing CoRs in Vaccine Recipients: Case-Control Design (Time-Dependent CoR) (Design of CYD14/15)

- **Objective:** Develop current-value correlates of instantaneous risk of outcome and VE against symptomatic VCD
- Measure immune responses longitudinally from all vaccinees with symptomatic VCD after Month 7 and from a random sample of vaccinees endpoint-free through Month 36



- 'Joint modeling' longitudinal + survival analysis methods (e.g., Fu and Gilbert, Lifetime Data Analysis, 2017)

Cox Model for Fixed-Time CoR Analysis (Fixed-Time CoR)

- Cox proportional hazards model

$$\lambda(t|Z) = \lambda_0(t) exp\left\{\beta_0^T Z\right\}$$

  - $\lambda(t|Z)$ = conditional failure hazard given covariate history until time t
  - β_0 = unknown vector-valued parameter
  - $\lambda_0(t) = \lambda(t|0)$ = unspecified baseline hazard function
    - $Z = (Z_1^T, Z_2^T)^T$, Z_1 are "Phase 1" baseline covariates measured in everyone and Z_2 are "Phase 2" (expensive) covariates only measured on failures and subjects in a random sub-sample
    - e.g.,
      Z_1 = treatment assignment, vaccination receipt, and baseline prognostic factors at enrollment;
      Z_2 = Immune response biomarkers measured at a fixed time point τ post-randomization

Notation and Set-Up (Similar to Kulich and Lin, 2004, *JASA*)

- T = failure time (e.g., time from Month 13 visit to dengue disease endpoint)
- C = censoring time
- $X = min(T, C), \Delta = I(T \leq C)$
- $N(t) = I(X \leq t, \Delta = 1)$
- $Y(t) = I(X \geq t)$
- Cases are subjects with $\Delta = 1$
- Controls are subjects with $\Delta = 0$

Notation and Set-Up (Matches Kulich and Lin, 2004, *JASA*) (Fixed-Time CoR)

- Consider a prospective cohort of N subjects, who are stratified by a variable V with K categories
  - V may contain *any* information available at the time of sampling (i.e., failure time, censoring time may be used as well as covariates)
- $\epsilon =$ indicator of whether a subject has Z_2 measured (i.e., the full vector Z measured)
  - $\alpha_k = Pr(\epsilon = 1 | V = k)$, where $\alpha_k > 0$
- $(X_{ki}, \Delta_{ki}, Z_{1ki}, V_{ki}, \epsilon_{ki})$ observed for all subjects
- $(X_{ki}, \Delta_{ki}, Z_{1ki}, Z_{2ki}, V_{ki}, \epsilon_{ki} = 1)$ observed for all subjects with $\epsilon_{ki} = 1$ (marker subcohort subjects and all cases)

Estimation of β_0 (Fixed-Time CoR)

- With full data, β_0 may be estimated by the MPLE, defined as the root of the score function

$$U_F(\beta) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ Z_i - \bar{Z}_F(t, \beta) \right\} dN_i(t), \tag{1}$$

where

$$\bar{Z}_F(t, \beta) = S_F^{(1)}(t, \beta) / S_F^{(0)}(t, \beta);$$

$$\begin{aligned} S_F^{(1)}(t, \beta) &= n^{-1} \sum_{i=1}^{n} Z_i exp\left\{ \beta^T Z_i \right\} Y_i(t) \\ S_F^{(0)}(t, \beta) &= n^{-1} \sum_{i=1}^{n} exp\left\{ \beta^T Z_i \right\} Y_i(t) \end{aligned}$$

Estimation of β_0 (Fixed-Time CoR)

- Due to missing data the previous equation (1) cannot be calculated under the sub-sampling design
- Most estimators are based on pseudoscores parallel to (1), with $\bar{Z}_F(t, \beta)$ replaced with an approximation $\bar{Z}_C(t, \beta)$

$$U_C(\beta) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \int_0^{\infty} \left\{ Z_{ki} - \bar{Z}_C(t, \beta) \right\} dN_{ki}(t)$$

  - The double indices k, i reflect the stratification

Estimation of β_0 (Fixed-Time CoR)

- The marker sampled cohort at-risk average is defined as

$$\bar{Z}_C(t,\beta) \equiv S_C^{(1)}(t,\beta)/S_C^{(0)}(t,\beta),$$

where

$$\begin{aligned} S_C^{(1)}(t,\beta) &= n^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\rho_{ki}(t)Z_{ki}exp\left\{\beta^T Z_{ki}\right\}Y_{ki}(t) \\ S_C^{(0)}(t,\beta) &= n^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\rho_{ki}(t)exp\left\{\beta^T Z_{ki}\right\}Y_{ki}(t) \end{aligned}$$

where $\rho_{ki}(t)$ is a weight

Estimation of β_0 (Fixed-Time CoR)

- $\rho_{ki(t)}$ is set to zero for subjects with incomplete data, eliminating them from the estimation
- Cases and subjects in the marker subcohort have $\rho_{ki}(t) > 0$
  - Usually $\rho_{ki}(t)$ is set as the **inverse estimated sampling probability** (Using the same idea as the weighted estimating equation methods of Robins, Rotnitzky, and Zhao, 1994, 1995)
- Different estimators are formed by different choices of weights $\rho_{ki}(t)$
- Two classes of estimators (case-cohort and 2-phase/case-control)

Case-cohort Estimators (Called N-estimators in Kulich and Lin, 2004) (Fixed-Time CoR)

- The subcohort is considered a sample from all study subjects regardless of failure status
- Original approaches:
  - Prentice (1986, *Biometrika*): $\rho_i(t) = \epsilon_i/\alpha$ for all controls and $\rho_i(t) = 1/\alpha$ for all cases
  - Self and Prentice (1988, *Ann Stat*): $\rho_i(t) = \epsilon_i/\alpha$ for all i
  - Equivalent if all cases have the phase-2 variables measured

Case-cohort N-estimators (Fixed-Time CoR)

- General stratified N-estimator
  - $\rho_{ki}(t) = \epsilon_i / \widehat{\alpha}_k(t)$ for all k, i for all controls and $\rho_{ki}(t) = 1$ for cases
    - $\alpha_k(t)$ is known by design, but nonetheless estimating $\alpha_k(t)$ provides greater efficiency for estimating β_0 (Robins, Rotnitzky, Zhao, 1994)

Two-phase Sampling Estimators (Called D-estimators in Kulich and Lin, 2004) (Fixed-Time CoR)

- D-estimators treat cases and controls **completely separately**
- Weight cases by 1
  - The $\alpha_k(t)$ apply to controls only, so that $\alpha_k(t)$ should be estimated using data only from controls
- Case-control estimators are the special case with one covariate sampling stratum $K = 1$

Two-phase Sampling D-estimators (Fixed-Time CoR)

- General D-estimator

$$\rho_{ki}(t) = \Delta_{ki} + (1 - \Delta_{ki})\epsilon_{ki}/\widehat{\alpha}_k(t)$$

  - Borgan et al. (2000, Estimator II) obtained by setting

$$\widehat{\alpha}_k(t) = \sum_i^n \epsilon_{ki}(1 - \Delta_{ki})Y_{ki}(t) / \sum_i^n (1 - \Delta_{ki})Y_{ki}(t),$$

    i.e., the proportion of the sampled controls among those who remain at risk at time t

Main Distinctions Between N- and D- Estimators

- For N-estimators, the sampling design is **specified in advance**, whereas for D-estimators, it can be **specified after the trial** (retrospectively)
    - D-estimators more flexible

Statistical Inference

- All of the methods provide Wald-based inference
  - Wald confidence intervals about elements of the β_0 vector
  - Wald p-values for testing hypotheses such as $H_0 : \beta_{0j} = 0$
  - Generalized Wald p-values for testing composite hypotheses such as all elements of β_0 are zero, $H_0 : \beta_0 = 0$

cch R Package for a Fixed-Time CoR

- Implements Cox regression for selected N and D estimators for a fixed-time marker(s), for unstratified or stratified sampling

Table: Method Options for the *cch* Package

Sampling design	*cch* Method
Unstratified sampling	
Case-cohort N	"Prentice" or "SelfPrentice"
Case-control D	"LinYing" (1993, *JASA*)
Stratified sampling	
Case-cohort N	"Borgan.I" (Generalized SelfPrentice) or
Two-phase D	"Borgan.II" (Generalized LinYing)

- In some applications the Case-cohort N vs. Case-control D/Two-Phase D sampling designs are equivalent, such that the methods are valid across the designs (e.g., CYD14, CYD15)

Example Fixed-Time CoR Analysis: RV144 HIV-1 VE Trial

Haynes et al. (2012, *NEJM*) assessed in vaccine recipients the association of 6 immune response biomarkers measured at Week 26 with HIV-1 infection through 3.5 years

- **2-phase sampling design:** Measured Week 26 responses from all HIV-1 infected cases ($n = 41$) and from a stratified random sample of controls ($n = 205$ by gender $\times$# vaccinations $\times$ per-protocol)

Immune Response Variable	Est. HR (95% CI)	2-Sided P-value
IgA Magnitude-Breadth to Env	1.58 (1.07–2.32)	0.02
Avidity to A244 Strain	0.90 (0.55–1.46)	0.66
ADCC to 92TH023 Strain	0.92 (0.62–1.37)	0.67
Neutralization M-B to Env	1.46 (0.87–2.47)	0.15
IgG to gp70-V1V2 Env	0.57 (0.37–0.90)	0.014
CD4 T cell Magn to 92TH023	1.17 (0.83–1.65)	0.37

Borgan et al. (2000) **estimator II** for the Cox model fit by *cch*

Fixed-Time CoR Analysis in the CYD14 and CYD15 Dengue VE Trials

The Journal of Infectious Diseases

MAJOR ARTICLE

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

Zoe Moodie,[1] Michal Juraska,[1] Ying Huang,[1,2] Yingying Zhuang,[2] Youyi Fong,[1,2] Lindsay N. Carpp,[1] Steven G. Self,[1,2] Laurent Chambonneau,[3] Robert Small,[4] Nicholas Jackson,[5] Fernando Noriega,[4] and Peter B. Gilbert[1,2]

[1]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; [2]Department of Biostatistics, University of Washington, Seattle; [3]Sanofi Pasteur, Marcy-L'Etoile, France; [4]Sanofi Pasteur, Swiftwater, Pennsylvania; [5]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.

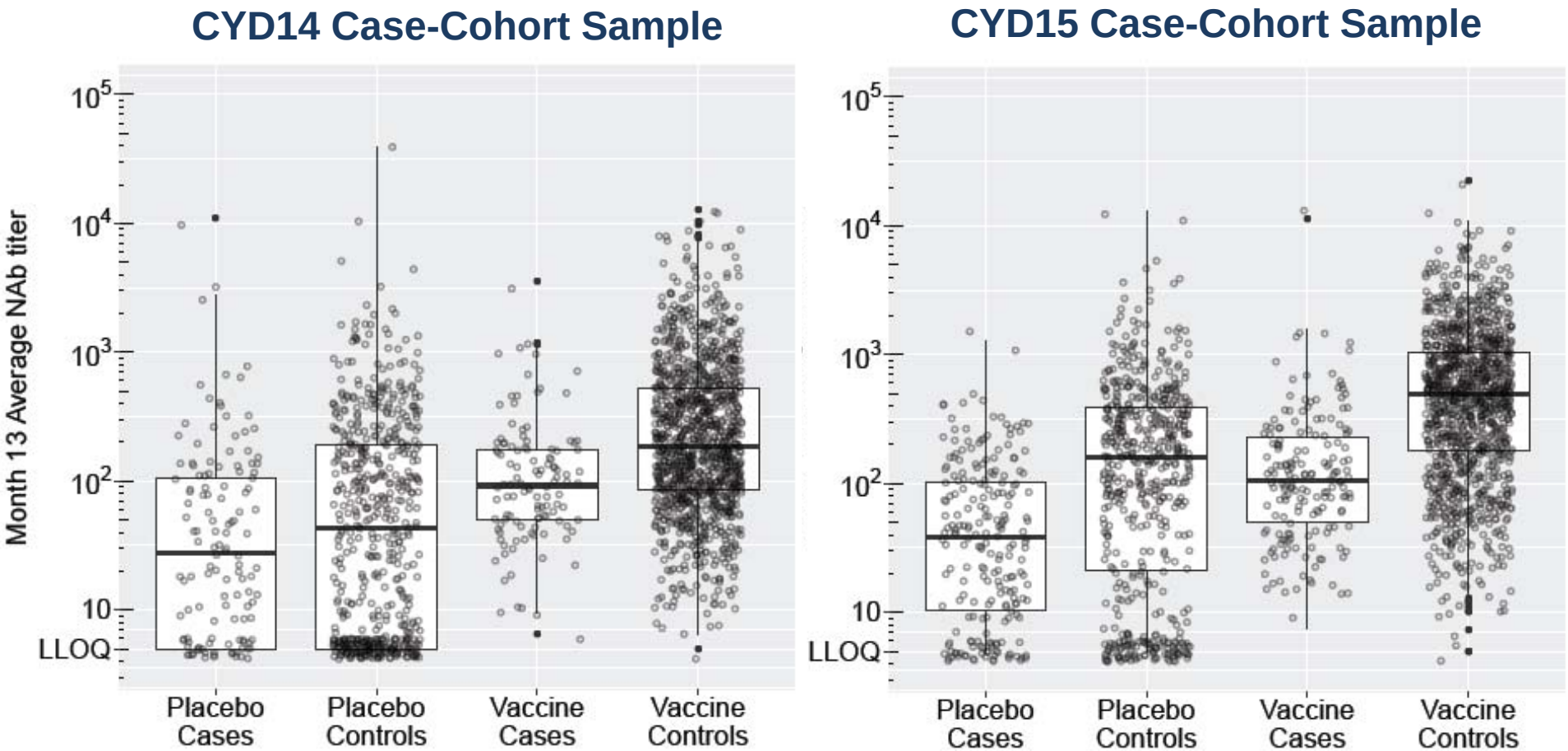
Methods. Neutralizing antibody titers to the 4 dengue serotypes in the CYD-TDV vaccine insert were measured at month 13 in a randomly sampled immunogenicity subcohort and in all VCD cases through month 25 (2848 vaccine, 1574 placebo) and studied for their association with VCD and with the level of VE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

Average nAb Titer Distributions (Moodie et al., 2018, JID)



*Average nAb titer = Geometric mean PRNT50 to the 4 dengue viruses in the vaccine construct

Cases = Dengue disease endpoint (VCD) between Month 13 and 25
Controls = Never experienced VCD through Month 25

FRED HUTCH

5

Fixed-Time CoR Analysis: CYD14 Dengue VE Trial

	CYD14 (n = 1390 Vaccine Recipients)		
Comparison	Hazard Ratio (95% CI)	*P* Value	Global *P* Value[b] (Holm *P* Value)[c]
Average Titer[d], DENV-Any			
Med vs low	0.37 (0.23–0.59)	<.001	<.001 (–)
High vs low	0.10 (0.05–0.19)	<.001	<.001 (–)
Per 10-fold increase	0.25 (0.17–0.38)	<.001	– (–)
DENV-1 Titer, DENV-1			
Med vs low	0.60 (0.33–1.10)	.10	.001 (.003)
High vs low	0.06 (0.01–0.28)	<.001	.001 (.003)
Per 10-fold increase	0.39 (0.25–0.62)	<.001	– (<.001)
DENV-2 Titer, DENV-2			
Med vs low	0.96 (0.38–2.45)	.93	.001 (.004)
High vs low	0.21 (0.07–0.64)	.006	.001 (.004)
Per 10-fold increase	0.43 (0.25–0.75)	.003	– (.009)
DENV-3 Titer, DENV-3			
Med vs low	0.61 (0.16–2.39)	.48	.29 (.42)
High vs low	0.16 (0.02–1.58)	.12	.29 (.42)
Per 10-fold increase	0.39 (0.13–1.15)	.09	– (.09)
DENV-4 Titer, DENV-4			
Med vs low	0.76 (0.27–2.20)	.62	.21 (.42)
High vs low	0.14 (0.02–1.25)	.08	.21 (.42)
Per 10-fold increase	0.30 (0.13–0.73)	.008	– (0.02)

*LinYing (1993, *JASA*) Cox model fit by *cch*

Another Application of the Case-Cohort Cox Model to the CYD14 and CYD15 Trials

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Effect of Dengue Serostatus on Dengue Vaccine Safety and Efficacy

S. Sridhar, A. Luedtke, E. Langevin, M. Zhu, M. Bonaparte, T. Machabert, S. Savarino, B. Zambrano, A. Moureau, A. Khromava, Z. Moodie, T. Westling, C. Mascareñas, C. Frago, M. Cortés, D. Chansinghakul, F. Noriega, A. Bouckenooghe, J. Chen, S.-P. Ng, P.B. Gilbert, S. Gurunathan, and C.A. DiazGranados

Case-Cohort Cox Model for a Baseline Covariate

- Phase-two covariate of interest:
  - Baseline nAb serostatus ($0 = \text{PRNT}_{50} < 10$ for all 4 serotypes; $1 =$ otherwise)
- Ming Zhu and Edith Langevin applied the case-cohort Cox model
- Constructed 10 phase-two baseline serostatus data sets by multiple-imputation, based on Month 13 NS1 assay readouts and other covariates
- Case-cohort Cox model fit 10 times and Rubin's variance rule applied

Sridhar et al. Results Support Vaccine Effect Modification by Baseline nAb Serostatus

Serostatus, End Point, and Method	Vaccine Group	Control Group	Relative Risk or Hazard Ratio (95% CI)
	no./total no.		
Seropositive			
Hospitalization for VCD			
MI, month 0 onward	58.8/1502.9	137.7/729.8	0.21 (0.14–0.31)
TMLE, month 0 onward	43.6/1442.6	121.3/699.3	0.19 (0.08–0.42)
NS1, T9, month 13 onward	49/1450	110/687	0.21 (0.15–0.30)
Severe VCD			
MI, month 0 onward	11.2/1502.9	33.4/729.8	0.16 (0.07–0.37)
TMLE, month 0 onward	8.6/1442.6	29.9/699.3	0.15 (0.07–0.35)
NS1, T9, month 13 onward	10/1450	27/687	0.18 (0.09–0.37)
Seronegative			
Hospitalization for VCD			
MI, month 0 onward	64.2/375.1	25.3/207.2	1.41 (0.74–2.68)
TMLE, month 0 onward	78.1/359.7	31.7/201	1.51 (0.73–3.11)
NS1, T9, month 13 onward	56/330	20/171	1.46 (0.85–2.49)
Severe VCD			
MI, month 0 onward	14.8/375.1	3.6/207.2	2.44 (0.47–12.56)
TMLE, month 0 onward	15.2/359.7	6.8/201	1.41 (0.44–4.46)
NS1, T9, month 13 onward	12/330	1/171	6.25 (0.81–48.32)

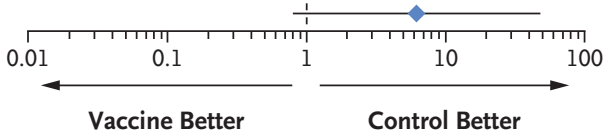


Figure 1. Risk of Hospitalization for Virologically Confirmed Dengue (VCD) and of Severe VCD in Participants 9 to 16 Years of Age, According to Baseline Serostatus.

The Cox Model with a Sub-Sampling Design (Time-Dependent Covariates)

- Cox proportional hazards model

$$\lambda(t|Z) = \lambda_0(t) exp\left\{\beta_0^T Z(t)\right\}$$

  - $\lambda(t|Z)$ = conditional failure hazard given covariate history until time t
  - β_0 = unknown vector-valued parameter
  - $\lambda_0(t) = \lambda(t|0)$ = unspecified baseline hazard function
    - $Z = (Z_1^T, Z_2(t)^T)^T$, Z_1 are "Phase 1" baseline covariates measured in everyone and $Z_2(t)$ are "Phase 2" (expensive) covariates only measured on failures and subjects in a random sub-sample
    - e.g.,
      Z_1 = treatment assignment, vaccination receipt, and baseline prognostic factors at enrollment;
      $Z_2(t)$ = Immune response biomarkers measured at longitudinal visits

Notation and Set-Up (Time-Dependent Covariates)

Same as for fixed-time covariates Z, except:

- $(X_{ki}, \Delta_{ki}, Z_{ki}(t), 0 \leq t \leq \tau, V_{ki}, \epsilon_{ki} \equiv 1)$ observed for all marker subcohort subjects
- At least $(X_{ki}, \Delta_{ki} \equiv 1, Z_{ki}(X_{ki}))$ observed for all cases

Estimation of β_0 (Time-Dependent Covariates)

Same as for fixed-time covariates except now terms depend on $Z_{ki}(t)$:

- The marker sampled cohort at-risk average is defined as

$$\bar{Z}_C(t,\beta) \equiv S_C^{(1)}(t,\beta)/S_C^{(0)}(t,\beta),$$

where

$$S_C^{(1)}(t,\beta) = n^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\rho_{ki}(t)Z_{ki}(t)exp\left\{\beta^T Z_{ki}(t)\right\}Y_{ki}(t)$$

$$S_C^{(0)}(t,\beta) = n^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}\rho_{ki}(t)exp\left\{\beta^T Z_{ki}(t)\right\}Y_{ki}(t)$$

Case-cohort N Estimators (Time-Dependent Covariates)

- The subcohort is considered a sample from all study subjects regardless of failure status
  - The whole covariate history $Z(t)$ is used for all subcohort subjects
  - For cases not in the subcohort, only $Z(T_i)$ (the covariate at the failure time) is used
- Prentice (1986, Biometrika): $\rho_i(t) = \epsilon_i/\alpha$ for $t < T_i$ and $\rho_i(T_i) = 1/\alpha$
- Self and Prentice (1988, Ann Stat): $\rho_i(t) = \epsilon_i/\alpha$ for all t

Case-cohort N-estimators (Time-Dependent Covariates)

- General stratified N-estimator
  - $\rho_{ki}(t) = \epsilon_i / \widehat{\alpha}_k(t)$ for $t < T_{ki}$ and $\rho_{ki}(T_{ki}) = 1$
    - $\widehat{\alpha}_k(t)$ is a possibly time-varying estimator of $\alpha_k(t)$
    - A time-varying weight can be obtained by calculating the fraction of the sampled subjects among those at risk at a given time point (Barlow, 1994; Borgan et al., 2000, Estimator I)

Two-phase Sampling D Estimators (Time-Dependent Covariates)

- Weight cases by 1 throughout their entire at-risk period
- D-estimators treat cases and controls **completely separately**
  - α_k apply to controls only, so that α_k should be estimated using data only from controls
- Case-control estimators are the special case with one covariate sampling stratum $K = 1$

Distinctions Between N- and D- Estimators (Time-Dependent Covariates)

- D-estimators require data on the complete covariate histories of cases
- N-estimators only require data at the failure time for cases
  - E.g., for the Vax004 HIV VE trial (Gilbert et al., 2005, *J Infect Dis*), the immune responses in cases were only measured at the visit prior to infection, so N-estimators are valid while D-estimators are not valid

Gaps of Both N- and D- Estimators (Time-Dependent Covariates)

Estimator	Does Not Need Full Covariate Histories in Cases	Allows Outcome-Dependent Sampling
N (Prosp. case-cohort)	Yes	No
D (Retrosp. 2-phase)	No	Yes

- For time-dependent correlates, none of the partial-likelihood based methods are flexible on both points
- All of the methods require full covariate histories in controls
- Critical implications for sample storage design

R Code for the Cox Model with Time-Dependent Covariates

- Therneau and Li (2000, *Lifetime Data Analysis*) describes how to implement several N and D estimators in R and SAS
- Code for implementing the Self-Prentice case-cohort estimation approach for a time-dependent phase-two covariate is at: http://faculty.washington.edu/peterg/SanofiPasteurCorrelatesRTraining.2018.html
  - E.g., very similar code was used in Gilbert et al. (2005, *J Infect Dis*)

Outline

1. Introduction to CoR analysis
2. CoR analysis with a Cox model
   - Fixed-time CoR
   - Time-dependent CoR
3. CoR analysis (marker at a single fixed time point) with a logistic regression model
4. Selected issues
   - Marker sampling design
   - Marker measurement error
5. Improved CoR methods (Breslow et al., 2009; Rose and van der Laan, 2011)

Notation and Set-Up

- Similar as for Cox model, except the outcome Y is binary
- $Y =$ failure outcome (e.g., dengue disease endpoint by 25 months)
- Cases have $Y = 1$; Controls $Y{=}0$
- Assume Y is known for all subjects (not exactly true due to participant dropout)
- $Z = (Z_1^T, Z_2^T)^T$, Z_1 are "Phase 1" baseline covariates measured in everyone and Z_2 are "Phase 2" (expensive) covariates only measured on failures and subjects in a random sub-sample
  - e.g., Phase 2 covariates = immune response biomarkers at fixed time τ (for a binary outcome only consider fixed-time correlates)

Notation and Set-Up

- Consider a prospective cohort of N subjects, who are stratified by a baseline covariate V with K categories
- $\epsilon =$ indicator of whether a subject has Z_2 measured (i.e., the full vector Z measured)
  - $\alpha_{yk} \equiv Pr(\epsilon = 1 | Y_i = y, V = k)$ for $y \in \{0, 1\}$ and $k \in \{1, \cdots, K\}$, where all $\alpha_{yk} > 0$
- $(Z_{1i}, V_i, Y_i, \epsilon_i)$ observed for all subjects
- $(Z_{1i}, Z_{2i}, V_i, Y_i, \epsilon_i)$ observed for all subjects with $\epsilon_i = 1$ [marker subcohort subjects and all cases (at least all cases with samples available)]

Two-Phase/Case-Control Sampling Design

Breslow and Cain (1988, *Biometrika*) two-phase outcome-dependent sampling design, also studied in Breslow and Holubkov (1997, *JRSS-B*)

Table: Two-Phase Sampling

	Phase 1 Baseline Covariate			
	1	2	·	K
$Y = 1$	n_{11}	n_{12}	·	n_{1K}
$Y = 0$	n_{01}	n_{02}	·	n_{0K}

- Within each of the $2 \times K$ strata, draw a random sample of n_{yk} subjects for phase 2 measurements
- Bernoulli sampling or without replacement sampling within each cell

Logistic Regression Model

$$P(Y = 1|Z = z) = \frac{\exp\left(Z^T\beta_0\right)}{1 + \exp\left(Z^T\beta_0\right)}$$

- If full sampling, estimate β_0 by maximum likelihood, e.g. *glm* in R
- Under case-control sampling, obtain valid inference on β_0 ignoring the sampling design (Prentice and Pyke, 1979, *Biometrika*)
  - But cannot estimate absolute conditional risks – for that, need weighting
  - Simplest approach fits *glm* to phase-two subjects only (complete cases) using the *weights* option
  - Sets *weights* to estimated inverse probability weights $1/\widehat{\alpha}_{yk}$ for all subjects in stratum (y, k)

Logistic Regression Model for Two-Phase Outcome-Dependent Sampling

Breslow and Holubkov (1997):

- Developed a pseudo-likelihood (PL) and a maximum likelihood (ML) estimator of β_0, with Wald 95% confidence intervals and p-values (same type of output as for the Cox model)
  - The ML estimator is fully efficient
  - The methods assume Bernoulli sampling of subjects within each of the $2 \times K$ strata
- All three estimation and inference procedures (WL, PL, ML) are implemented in the *osDesign* R package

Logistic Regression Model for Two-Phase Outcome-Dependent Sampling: Variance Estimation

- Variance estimators use the decomposition

$$Cov(\widehat{\beta}) = Cov_{Phase1}(\widehat{\beta}) + Cov_{Phase2}(\widehat{\beta}),$$

  $=$ Phase 1 variance if full data were available $+$ Phase 2 variance added because of incomplete data

- All variance estimators estimate $Cov_{Phase2}(\widehat{\beta})$ by the Horvitz-Thompson sandwich formula
  - They differ by whether a model-based or empirical sandwich estimator is used for $Cov_{Phase1}(\widehat{\beta})$
  - Results tend to be very similar unless the data set is very small or the logistic regression model is grossly mis-specified (Haneuse, Saegusa, Lumley, 2011, *J Stat Software*)
  - *osDesign* uses empirical-based for WL and model-based for (PL, ML) – seems safe to use these defaults in practice

Example Fixed-Time CoR Analysis: CYD14 Dengue VE Trial (Analysis Pools Over Vaccine and Placebo)

Table 10: Overall Dengue Endpoint: LR Multivariate Quantitative Variable (Pooled)

Immune Response Variable	Est. OR	95% LL	95% UL	P-value	Q-value	Interaction P-Value
Serotype 1	1.01	0.73	1.39	0.973	0.973	0.328
Serotype 2	1.08	0.77	1.53	0.647	0.863	0.033
Serotype 3	0.88	0.60	1.28	0.503	0.863	0.016
Serotype 4	0.73	0.45	1.18	0.201	0.804	0.698

p=0.217 for any association of the variables with endpoint.
$p < 0.001$ for a different association: vaccine vs. placebo.

*Breslow and Holubkov (1997) pseudo-likelihood (PL) estimator fit by *tps* in the *osDesign* R package

Logistic and Cox Regression Gave Very Similar Answers in RV144, HVTN 505, CYD14, CYD15

- Consistent results by method were reported in Haynes et al. (2012, *NEJM*), Hammer et al. (2013, *NEJM*), and Moodie et al. (2018, *J Infect Dis*)
- Conjecture that the results are consistent because all of the studies had a rare event, and low levels of right-censoring/loss to follow-up
- **Implication:** In such settings, it is reasonable to interpret power calculations based on logistic regression as applying to power calculations for Cox regression

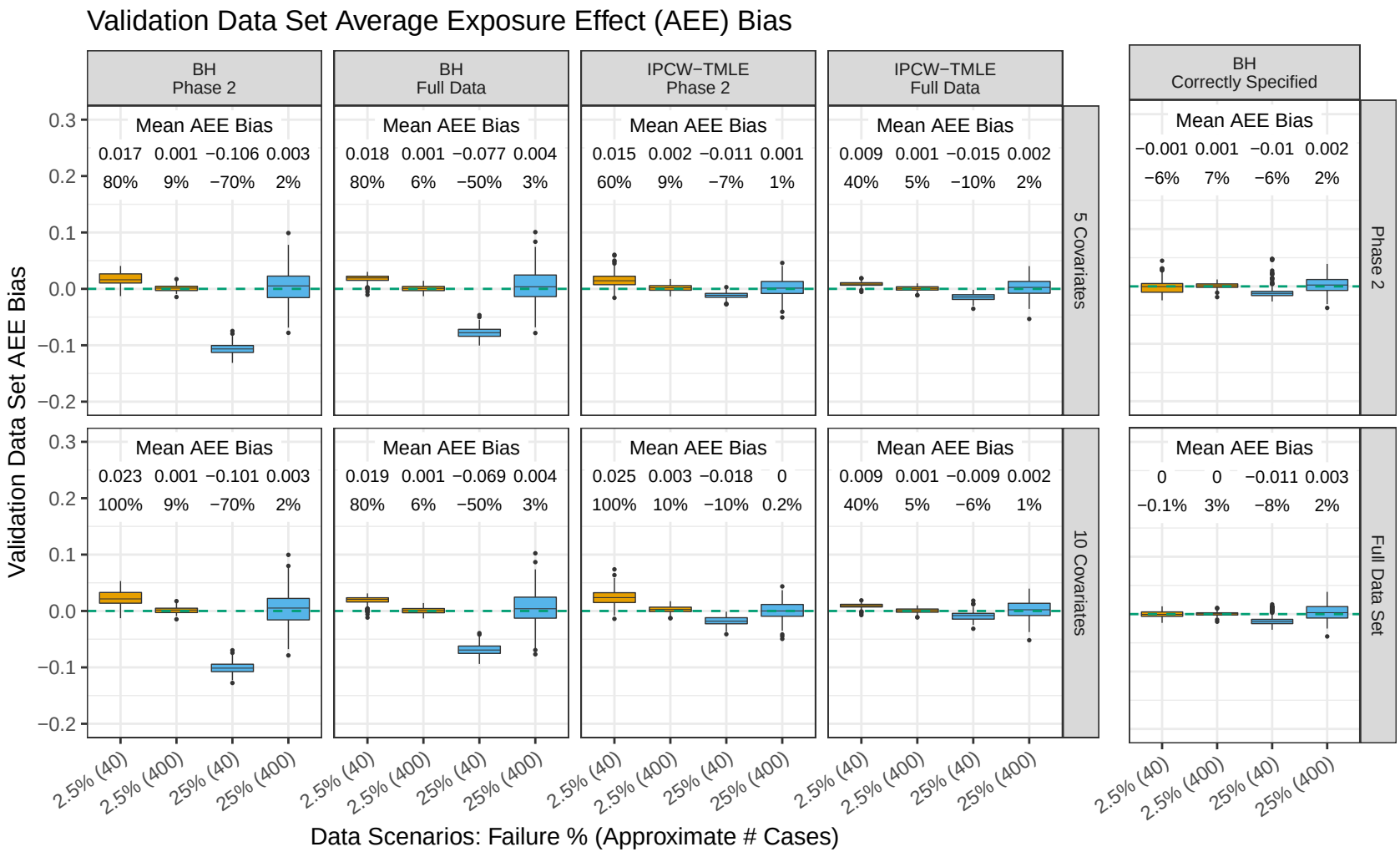
Remarks on Use of the Breslow-Holubkov Logistic Regression to Assess CoRs

- PL and ML tend to be more efficient than WL
- WL is more robust in a sense:
  - If the logistic regression model is mis-specified then the WL estimator correctly approximates the results that would have been obtained by fitting the (wrong) model to complete data for all N subjects at Phase 1 (not true for the PL and ML methods)
  - PL and ML assume $P(Y = 1|Z = z, V = v) = P(Y = y|Z = z)$; WL does not
    - (the assumption always holds by including V in the regression model)
- All of the Breslow-Holubkov methods are not robust
  - They require a correctly specified logistic regression model for consistent estimation

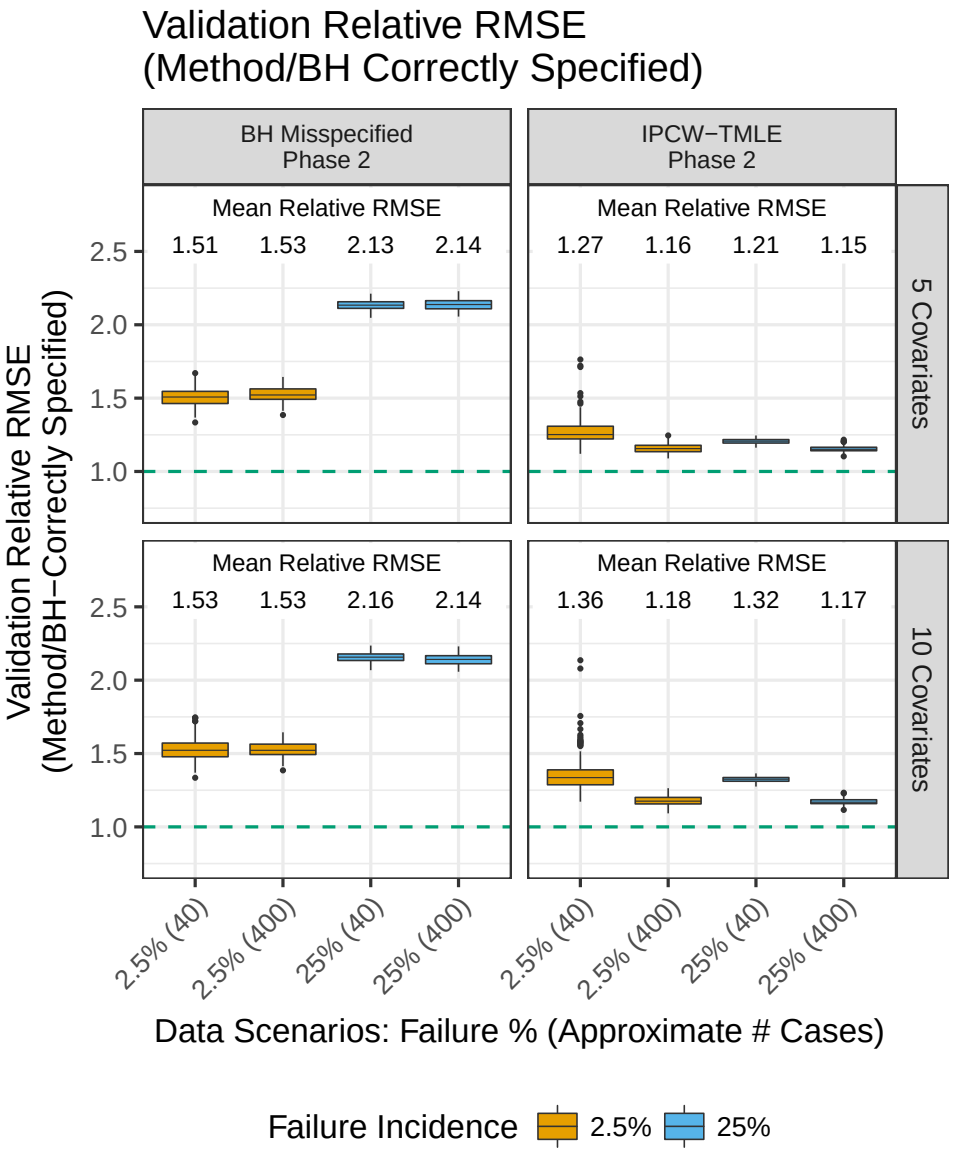
Remarks on Use of the Breslow-Holubkov Logistic Regression to Assess CoRs

- While the methods are designed for **Bernoulli** two-phase sampling, they are approximately correct if **without replacement** two-phase sampling is used
  - Consistent estimation and inference that is slightly conservative (confidence intervals a little too wide, p-values a little too big)
- More efficient and robust methods may be considered
  - Inverse probability of censoring weighted targeted minimum loss based estimation (IPCW-TMLE) for logistic regression (Rose and Van der Laan, 2011, *Int J Biost*)
  - My PhD student Brenda Price is elaborating this IPCW-TMLE approach

Teaser: Illustration of Greater Validity, Robustness, and Efficiency of IPCW-TMLE vs. Breslow-Holubkov



Teaser: Illustration of Greater Validity, Robustness, and Efficiency of IPCW-TMLE vs. Breslow-Holubkov



Outline

1. Introduction to CoR analysis
2. CoR analysis with a Cox model
   - Fixed-time CoR
   - Time-dependent CoR
3. CoR analysis (marker at a single fixed time point) with a logistic regression model
4. Selected issues
   - Marker sampling design
   - Marker measurement error
5. Improved CoR methods (Breslow et al., 2009; Rose and van der Laan, 2011)

Some Marker Sampling Questions to Consider Further

- Prospective or retrospective sampling?
- How much of the cohort to sample?
- Sampling design: Which subjects to sample?

Prospective or Retrospective Sampling?

Prospective case-cohort sampling: Select a random sample for immunogenicity measurement **at baseline**

- Advantages of prospective sampling
  - Can estimate case incidence for groups with certain immune responses
  - Can study correlations of immune response with multiple study endpoints
  - Straightforward to descriptively study the distribution of the immune responses in the whole study population at-risk when the immune responses are measured
  - **Practicality:** The lab will know what subjects to sample as early as possible, and there is one simple subcohort list

Prospective or Retrospective Sampling?

Retrospective 2-phase sampling: At or after the final analysis, select a random sample of control subjects for immunogenicity measurement

- Advantages of retrospective sampling
  - Can frequency match controls to cases to obtain balance on important covariates
    - E.g., enough representation of girls and boys
    - Can do "balanced sampling" on a prognostic factor to gain efficiency (balanced sampling = equal number of subjects in each of the $2 \times K$ sampling cells)
    - Can flexibly adapt the sampling design in response to the results of the trial
    - E.g., Suppose the results indicate effect modification, with VE $>>$ 0% in a subgroup and $VE \approx 0\%$ in other subgroups. Can over-sample controls in the 'interesting' subgroup.

Prospective or Retrospective Sampling?

- For applications where there is one primary endpoint and it is not of major interest to estimate absolute case incidence, retrospective sampling may be typically preferred

How Many Controls to Sample?

- In prevention trials, for which the clinical event rate is low, it is very expensive and unnecessary to sample all of the controls
  - Vax004 trial vaccine recipients: 225 HIV infected cases; $\approx$ 3000 controls
  - RV144 trial vaccine recipients: 41 HIV infected cases; $\approx$ 7000 controls
  - **Rule of thumb:** Under the null hypothesis, a $K:1$ Control:Case ratio achieves relative efficiency of $1 - \frac{1}{1+K}$ compared to complete sampling

K	Relative Efficiency
1	0.50
2	0.67
3	0.75
4	0.80
5	0.83
10	0.91

  - Simulations useful for studying the trade-offs of different K under alternative CoR hypotheses

Which Controls to Sample?

Principle: Well-powered CoR evaluation requires broad variability in the biomarker and in the risk of the clinical endpoint

- Can improve efficiency by over-sampling the "most informative" subjects
  - Disease cases (usually sampled at 100%)
  - Rare or unusual immune responses; or rare covariate patterns believed to affect immune response (e.g., HLA subgroups)
- Auxiliary Phase I variables measured in everyone are most valuable when they predict the missing data (i.e., the biomarker of interest)
- In general, optimal sampling obtained with sampling probabilities proportional to the cost-adjusted square-root variance of the efficient influence function (Gilbert, Yu, Rotnitzky, 2014, *Stat Med*)

Practical Thoughts for Consideration on Sampling

Practical questions for studying a given phase-two immune response biomarker as a CoR:

- Is there a Phase 1 covariate that should be reasonably predictive that a vaccine recipient will have a negative response?
- Is there a Phase 1 covariate that should be reasonably predictive of whether a vaccine recipient has a high response?
- If yes on both, then may gain efficency by oversampling controls predicted to have a negative response, and also oversampling controls predicted to have a high response

The trick is how good does the prediction need to be to be worth it? The 'cost' of Phase 1 covariate-dependent sampling is a little more complexity in analysis (e.g., all down-stream analyses should account for the oversampling).

Outline

1. Introduction to CoR analysis
2. CoR analysis with a Cox model
   - Fixed-time CoR
   - Time-dependent CoR
3. CoR analysis (marker at a single fixed time point) with a logistic regression model
4. Selected issues
   - Marker sampling design
   - Marker measurement error
5. Improved CoR methods (Breslow et al., 2009; Rose and van der Laan, 2011)

Measurement Error Reduces Power to Detect a CoR

Illustrative Example

- 'True' CoR $S \sim N(0,1)$
- 'Measured CoR' $S^* = S + \epsilon,\ \epsilon \sim N(0, \sigma^2)$
- Disease outcome status Y generated from $\Phi(\alpha + \beta S)$

  with α set to give $P(Y = 1|S = 0) = 0.20$ and β set to give $P(Y = 1|S = 1) = 0.15$

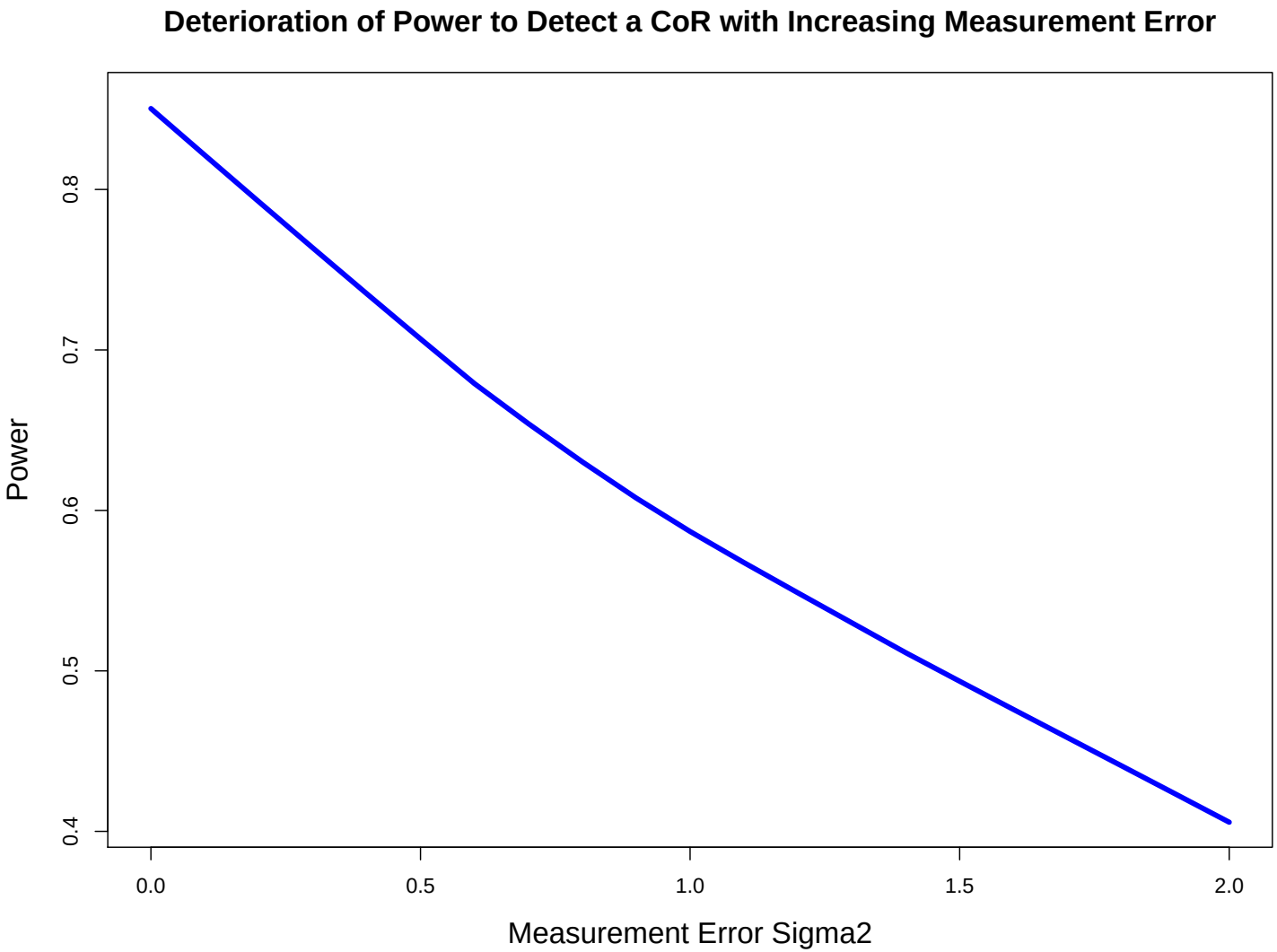
  σ^2 ranges from 0 to 2 (no-to-large measurement error)

Measurement Error Reduces Power to Detect a CoR

Simple Simulation Study

- Consider a study with $n = 500$ participants
- Consider power of a logistic regression model to detect an association between S^* (the observed variable) and Y

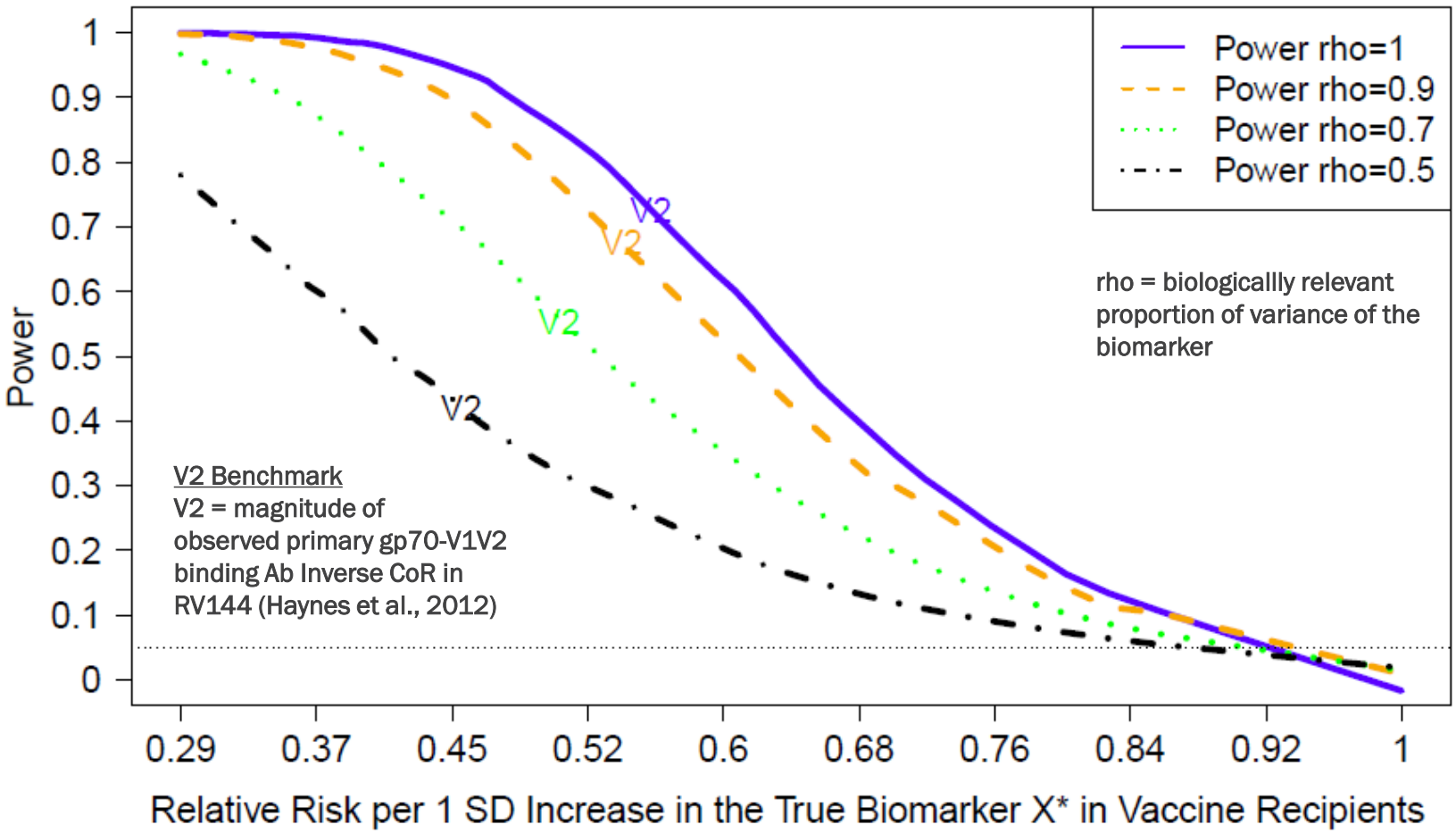
Measurement Error Reduces Power to Detect a CoR



Power Calculations for Assessing CoRs

- Ideally, the power/sample size calculations should explicitly account for measurement error in the assay
  - E.g., Gilbert, Janes, Huang (2016, *Stat Med*), implemented in the R package *CoRpower* (Michal to describe)
  - E.g., specify $\rho \equiv \sigma^2/\sigma^2_{obs}$, the proportion of inter-vaccinee variability of the biomarker that is biologically relevant
  - **Rule of thumb:** $\rho =$relative efficiency for estimating a CoR odds ratio for the underlying perfect biomarker compared to the observed biomarker (McKeown-Eyssen, Tibshirani, 1994, *AJE*)
  - 'Noise' components of σ^2_{obs} may be estimated, especially from laboratory assay validation studies
    - Within-vaccinee variability of replicates
    - Between-vaccinee variability due to variability in the time from the last immunization to marker sampling

Power to Detect a CoR of HIV Infection in Vaccinees in HVTN 505 ($\alpha = 0.05$)



Method: 2-phase logistic regression (Holubkov and Breslow, 1997)

Outline

1. Introduction to CoR analysis
2. CoR analysis with a Cox model
   - Fixed-time CoR
   - Time-dependent CoR
3. CoR analysis (marker at a single fixed time point) with a logistic regression model
4. Selected issues
   - Marker sampling design
   - Marker measurement error
5. Improved CoR methods (Breslow et al., 2009; Rose and van der Laan, 2011)

Typical Correlates Assessments are Inefficient

- Broadly in epidemiology studies, biomarker-disease associations are commonly assessed ignoring much data collected in the study
- That is, only subjects with the biomarker measured are included in the analysis
- Standard analyses use inverse probability weighting of the biomarker sampled subcohort
- These ubiquitously-used methods are implemented in the R packages *cch* (Breslow and Lumley) and *osDesign* (Haneuse, Saegusa, Chatterjee, Breslow, Smoot)

Typical Correlates Assessments are Inefficient

- Breslow et al.* urge statisticians/epidemiologists to consider using the whole cohort in the analysis of case-cohort/2-phase sampling data
- Baseline data on demographics and potential confounders are typically collected in all subjects (the Phase I data measured in everyone)
- These Phase I data are most valuable when they predict "missing" data

*Breslow, Lumley et al. (2009, *AJE*, *Stat Biosciences*)

How to Leverage All of the Data?

- **Question:** How can we use the Phase I data to improve the assessment of CoRs?
- **One Answer:** One approach adjusts the sampling weights used in the standard analyses described above to obtain approximately efficient estimators (e.g., Breslow et al., 2009, *AJE*, *Stat Biosciences*)

Some Lessons Learned from Breslow et al. (2009)

1. Obtain 'worthwhile' efficiency gain for the CoR assessment if baseline covariates can explain at least 40% of the variation in the immunological biomarker ($R^2 \geq 0.40$)
2. If interested in interactions (evaluation of whether a baseline covariate measured in everyone modifies the association of the biomarker and the clinical endpoint), can obtain worthwhile efficiency gain with a lower R^2
3. Even if no gain for the CoR assessment, will usually dramatically improve efficiency for assessing the associations of the Phase I covariates with outcome
4. Therefore it may often be the preferred method, and practicioners should have methods accounting for all of the data in their analytic toolkit
5. Additional research needed to make these more-efficient methods work well for multivariate markers and for time-dependent markers

How to Leverage All of the Data?

- **Question:** How can we use the Phase I data to improve the assessment of CoRs?
- **Another Answer:** Use an efficient and double-robust method: Inverse probability of censoring weighted targeted minimum loss based estimation (IPCW-TMLE) (Rose and Van der Laan, 2011, *Int J Biost*)
- Accessible R code does not seem to be available. Brenda Price is developing R code as part of her dissertation research. The code will handle a case-cohort or two-phase sampling study, for a binary outcome or a failure time outcome subject to right-censoring.

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"

References for the *cch* R Package Methods

1. Prentice, RL (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 73: 1-11.
2. Self, S and Prentice, RL (1988). Asymptotic distribution theory and efficiency results for case-cohort studies. Annals of Statistics 16: 64-81.
3. Lin, DY and Ying, Z (1993). Cox regression with incomplete covariate measurements. Journal of the American Statistical Association 88: 1341-1349.
4. Barlow, WE (1994). Robust variance estimation for the case-cohort design. Biometrics 50: 1064-1072
5. Therneau, TM and Li, H (1999). Computing the Cox model for case-cohort designs. Lifetime Data Analysis 5: 99-112.
6. Borgan, O, Langholz, B, Samuelsen, SO, Goldstein, L and Pogoda, J (2000). Exposure stratified case-cohort designs. Lifetime Data Analysis 6, 39-58.