Evaluating immune correlates of protection in vaccine efficacy trials with stochastic-interventional causal effects

Nima Hejazi

Wednesday, 26$^{\text{the}}$ June 2024

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Department of Biostatistics,
T.H. Chan School of Public Health,
Harvard University

nshejazi
nhejazi
nimahejazi.org

Session on *Causal Inference for Studying Vaccine Effects*
International Symposium on Nonparametric Statistics
Joint with P.B. Gilbert and M.J. van der Laan

Prelude:
Immune correlates of protection

The fights against HIV and COVID-19

- The HIV epidemic:
  - 1.5 million new infections occurring annually worldwide;
  - new infections outpace patients starting antiretroviral therapy;
  - HIV Vaccine Trials Network's (HVTN) *505* trial evaluated a novel antibody boost vaccine (Hammer et al. 2013).
- The COVID-19 ~~epi~~ ~~pan~~ endemic (Antia and Halloran 2021):
  - ~~270~~ ~~331~~ ~~619~~ ~~643~~ ~~686~~ 772 million total cases globally (WHO);
  - new variants emerging, with vaccine uptake globally stalled;
  - COVID-19 Prevention Network's (CoVPN) *COVE* trial focused on Moderna's (mRNA-1273) vaccine (Baden et al. 2021).

Evaluating vaccine protection for HIV and COVID-19

- In *505*: How would HIV infection risk have differed had the boost vaccine modulated antibody responses differently?
- In *COVE*: How would COVID-19 disease risk have differed for alternative vaccine-induced immunogenic response profiles?
- **Question**: Can [HIV-1, COVID-19] vaccines be improved by the modulation of immunogenic response profiles? How?

Why measure and analyze immune correlates?

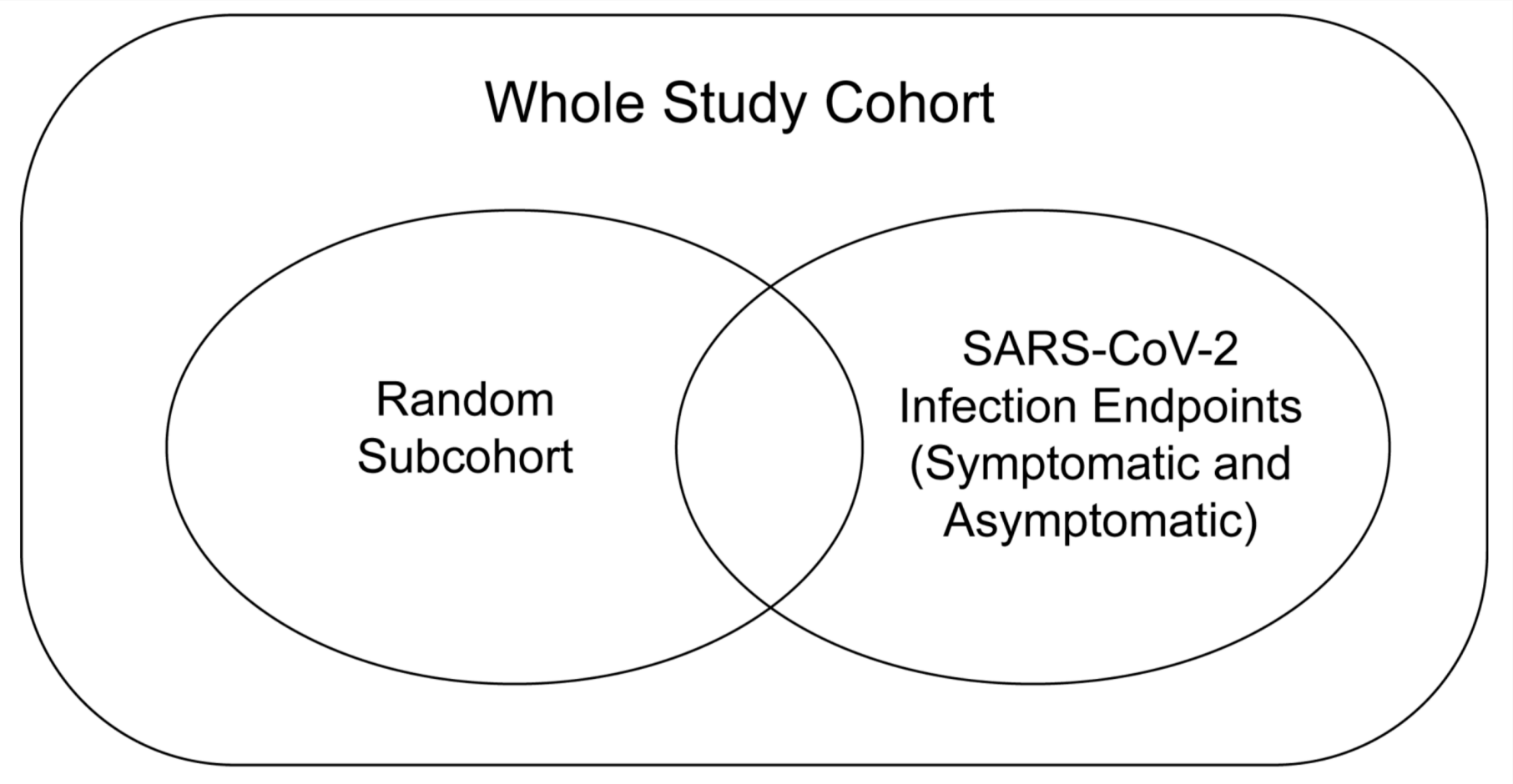
- Two interrelated goals of immune correlates analyses are to
  - identify/validate possible *surrogate endpoints* (Prentice 1989);
  - understand/delineate *protective mechanisms* of vaccination.
- If an immune correlate is established to reliably predict VE, subsequent efficacy trials may use it as a primary endpoint.
- Such surrogate endpoints may accelerate approval of
  - existing vaccines in different populations (e.g., in children);
  - new vaccines within the same class (e.g., based on mRNA);
  - inform the development of future vaccines (e.g., Ab targets);
  - shed light on immunologic mechanisms of disease occurrence.

Measuring correlates: Two-phase designs

- Often, use case-cohort design (Prentice 1986), a special case of two-phase sampling (Breslow et al. 2003).
- Phase 1: measure baseline, vaccination, endpoint on everyone.
- Phase 2: given baseline, vaccine, endpoint, select members of immune response subcohort with (possibly known) probability.
  - In *505*: phase-two sample with 100% of HIV-1 cases and matching of non-cases ($n = 189$ per Janes et al. 2017)).
  - In *COVE*: stratified random subcohort ($n \approx 1600$) and all SARS-CoV-2 infection and COVID-19 disease endpoints.

A simple two-phase design: Case-cohort

Assaying over 30000 samples is expensive, statistically unnecessary.



Case-cohort design, per Prentice (1986), as applied to COVE.

Two-phase sampling masks the complete data structure

- Complete unobserved data $X = (L, A, S, Y) \sim P_0^X \in \mathcal{M}^X$:
  - L (baseline covariates): sex, age, BMI, behavioral HIV risk,
  - A (treatment): randomized assignment to vaccine/placebo,
  - S (exposure): immune response profile for relevant markers,
  - Y (outcome of interest): infection status at trial's end.
- Observed data $O = (B, BX) = (L, A, B, BS, Y) \sim P_0 \in \mathcal{M}$.
  - $B \in \{0, 1\}$ indicates inclusion in the phase-two sample.
  - $\pi_0 := \mathbb{P}(B = 1 \mid Y, L)$ must be *known by design* or estimated.

Act I:
Causal effects for quantitative exposures

Static interventions: What are they good for?

- Describe the manner in which X is hypothetically generated via an NPSEM-IE (Pearl 2009):

  $$L = f_L(U_L); A \sim \text{Bern}(0.5); S = f_S(A, L, U_S); Y = f_Y(S, A, L, U_Y)$$

- Implies a model for distribution of counterfactual random variables induced by interventions on the system under study.
- A *static* intervention replaces (by "graph surgery") f_S with a specific value $S = s$ in the conditional support $S \mid A = a, L$.
- This requires specifying *a priori* a particular value of exposure under which to evaluate the outcome — but what value?

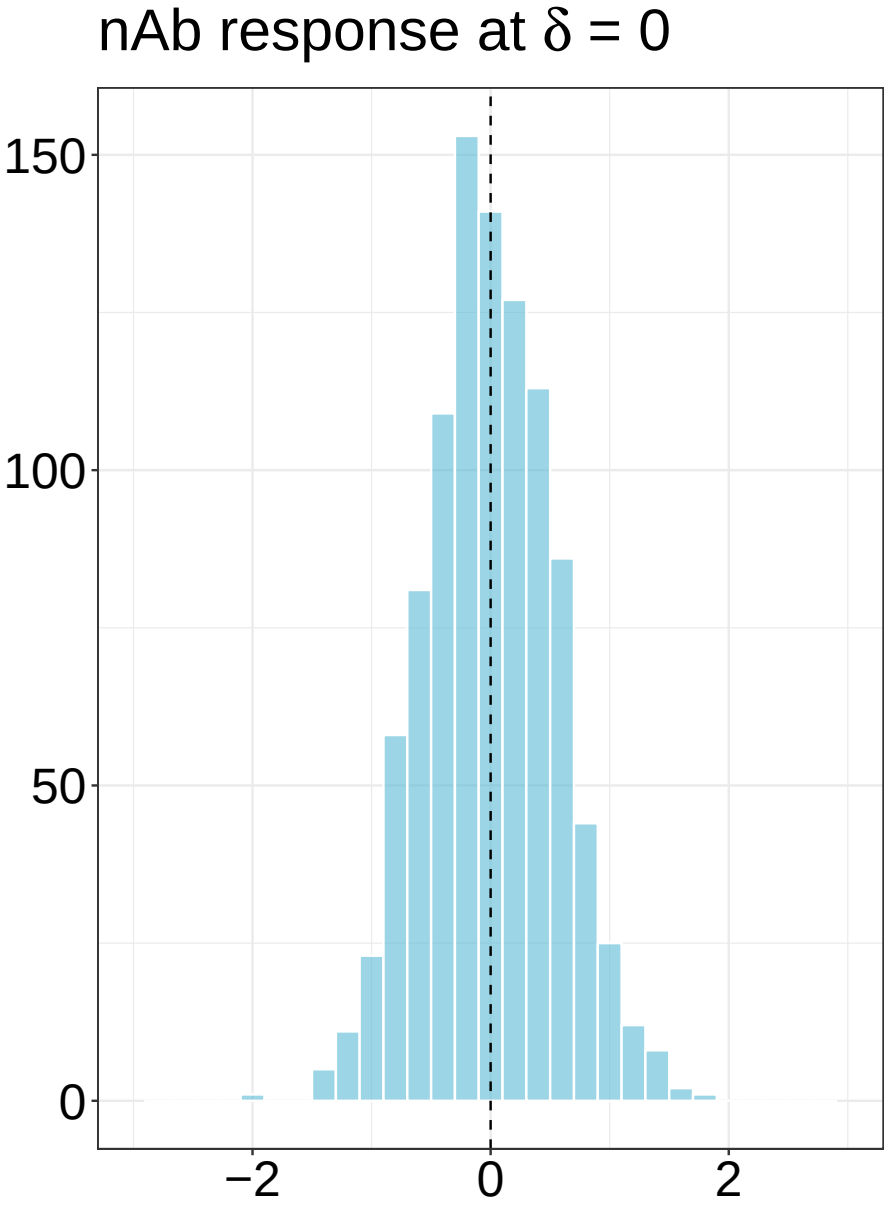
Controlled vaccine efficacy

- For a hypothetical value $s \in \mathcal{S}$, the *controlled direct effect* (CDE) quantifies the effect of A on Y while fixing $S = s$.
- The hypothetical value $S = s$ must be chosen carefully — to be scientifically informative *and* to avoid positivity violations.
- For two hypothetical values $s_0, s_1 \in \mathcal{S}$, *Controlled Vaccine Efficacy* (CVE) (Gilbert et al. 2022) is

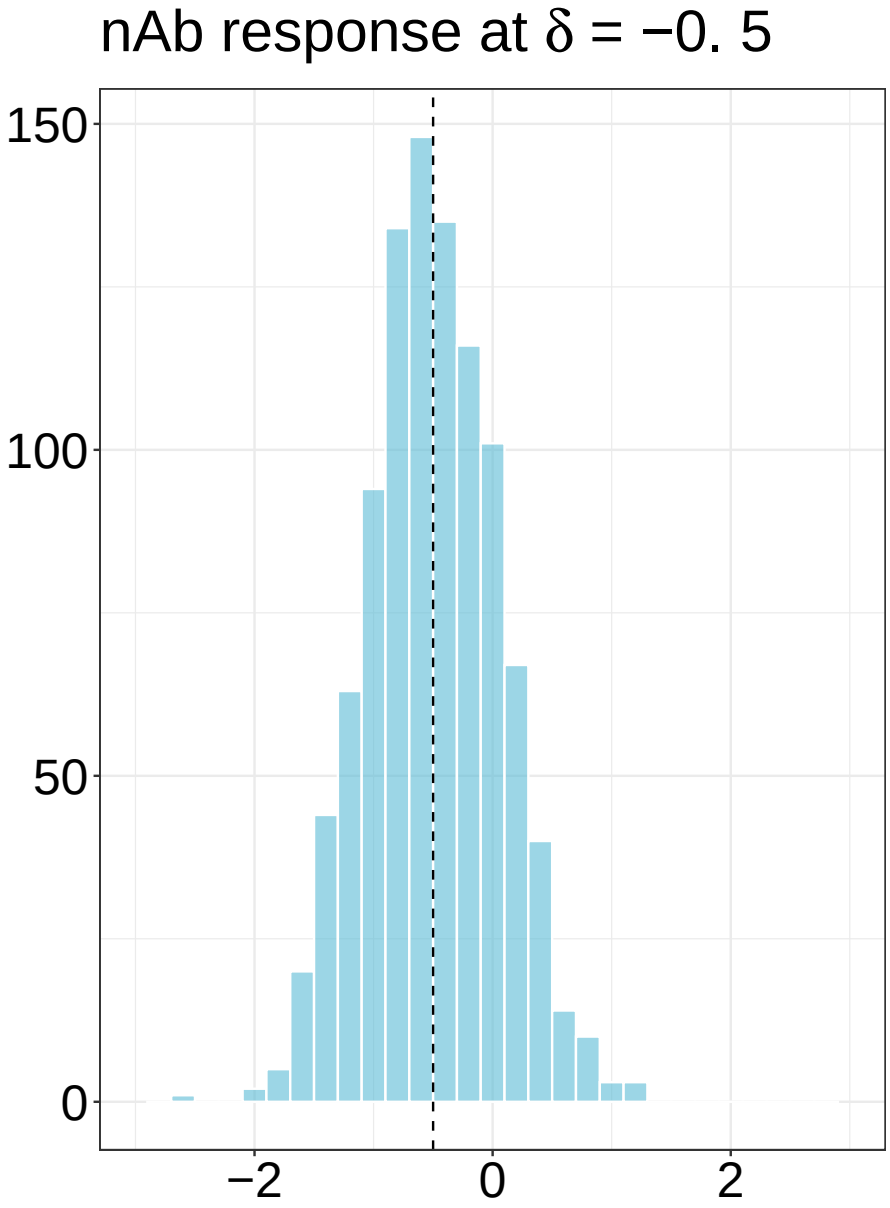
$$\text{CVE}(\cancel{s_0}, s_1) = 1 - \frac{\mathbb{E}[\mathbb{P}(Y = 1 \mid S = s_1, A = 1, L = l)]}{\mathbb{E}[\mathbb{P}(Y = 1 \mid \cancel{S = s_0}, A = 0, L = l)]},$$

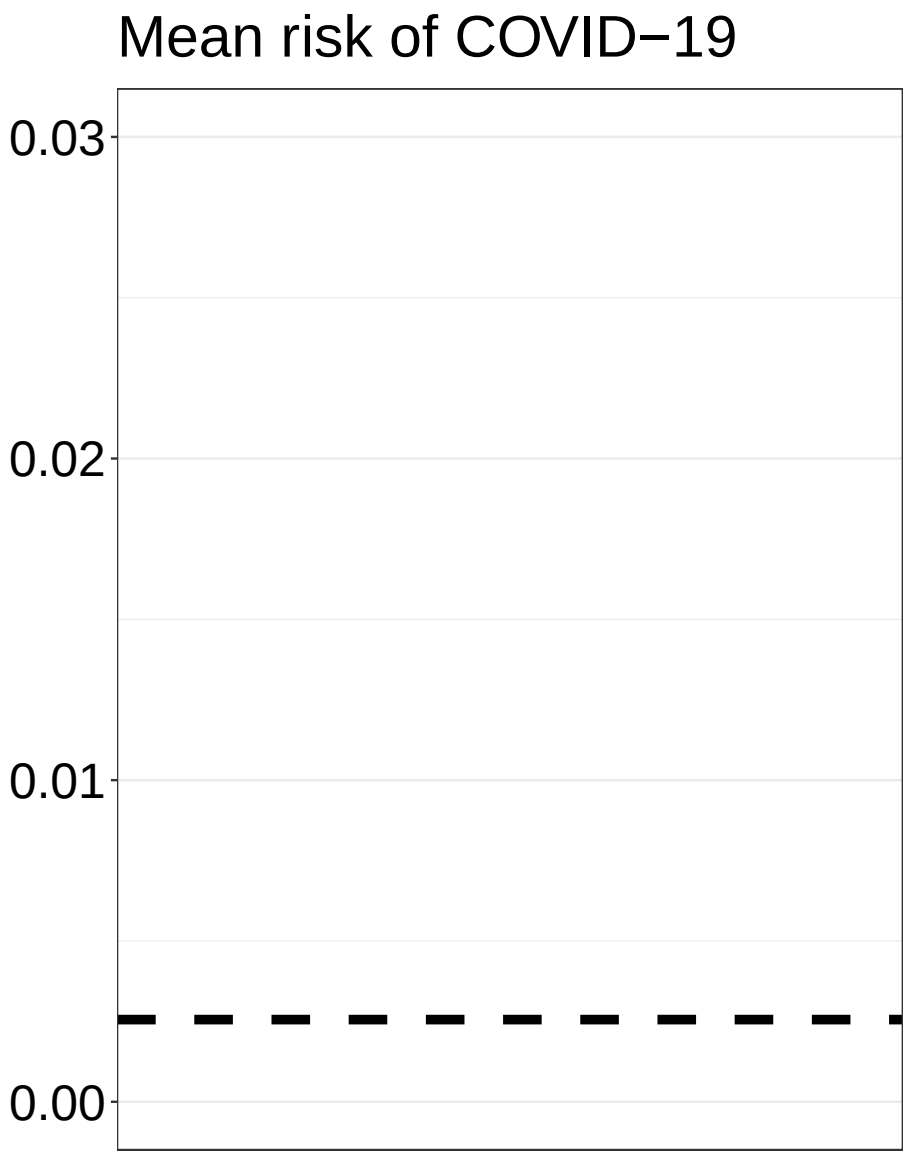
contrasting risk for vaccine receipt and $S = s_1$ vs. placebo and $S = s_0$, where $s_0 = 0$ is a plausible simplifying assumption in pathogen-naïve populations (Gilbert et al. 2024).

Disease risk under modified immunogenic responses

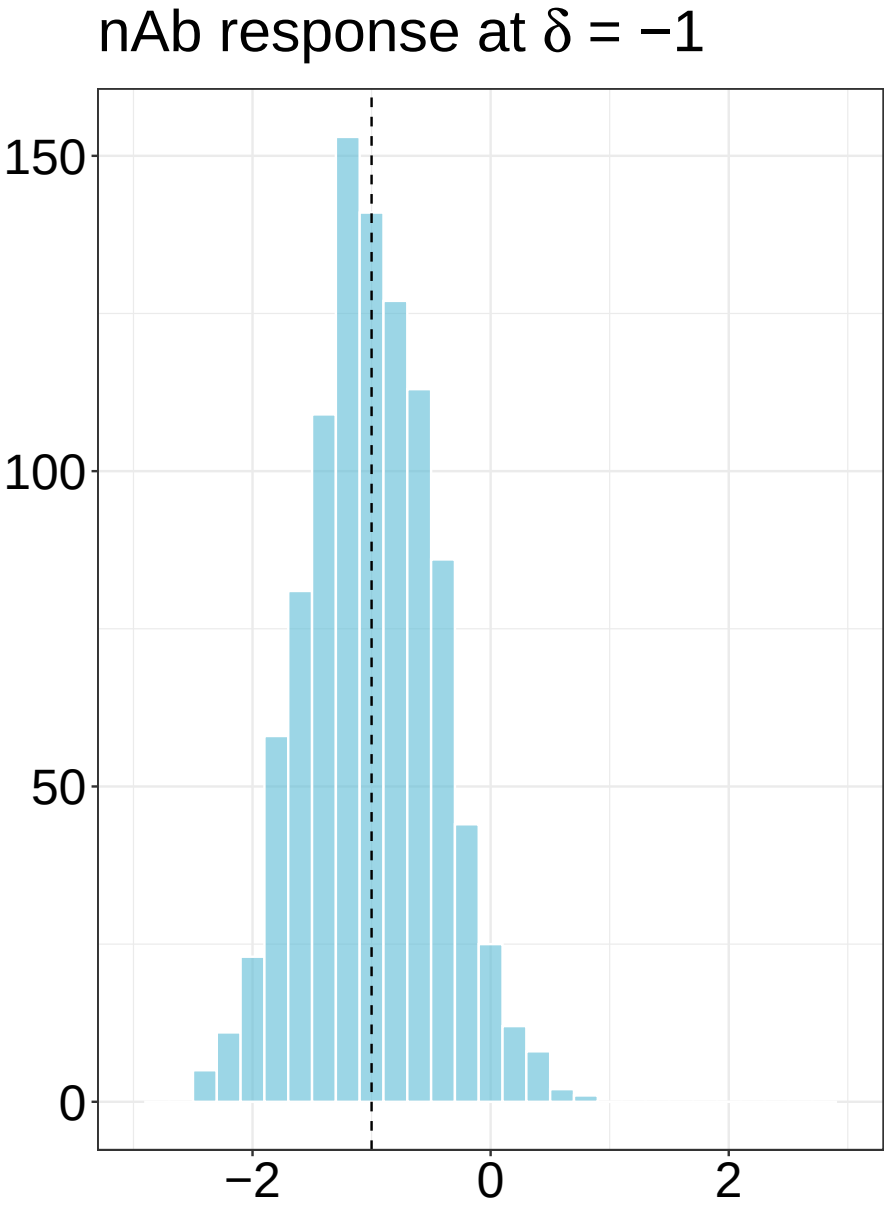


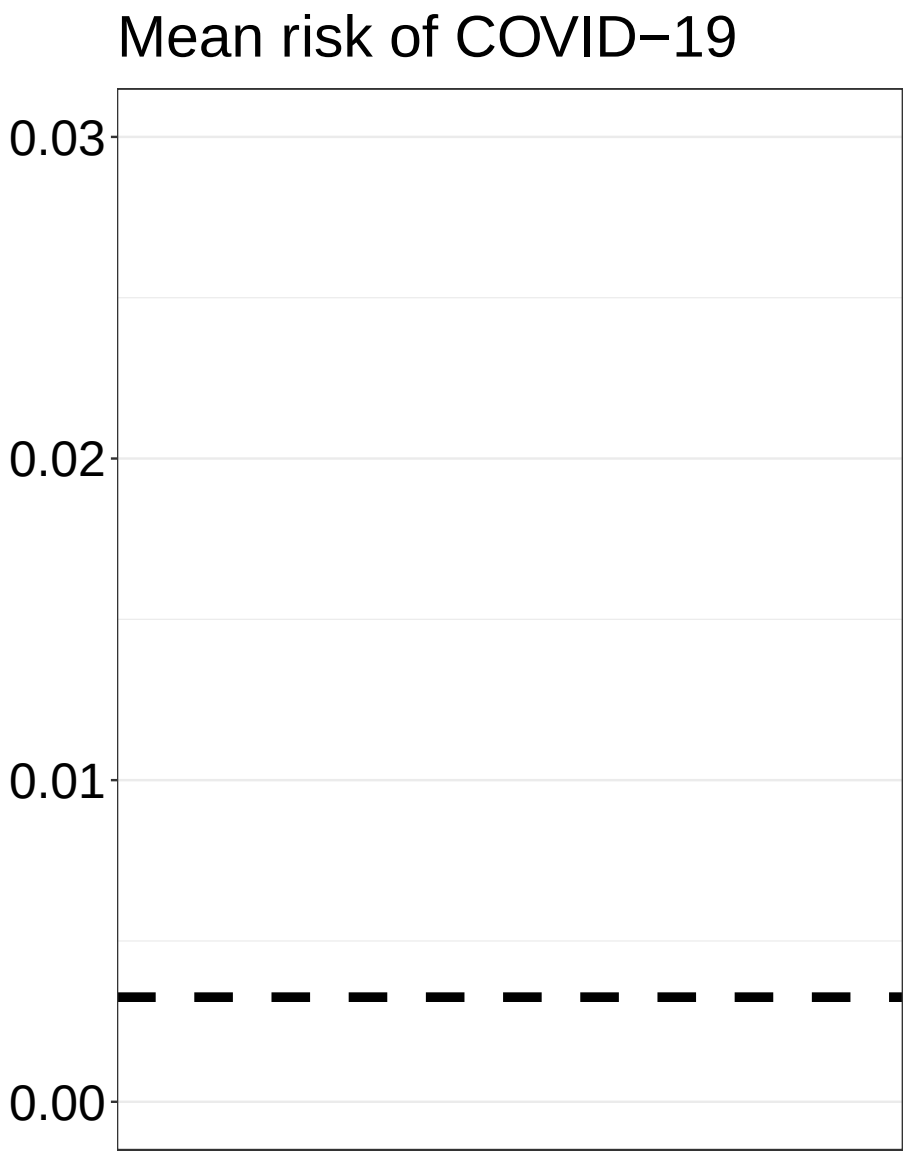
Disease risk under modified immunogenic responses



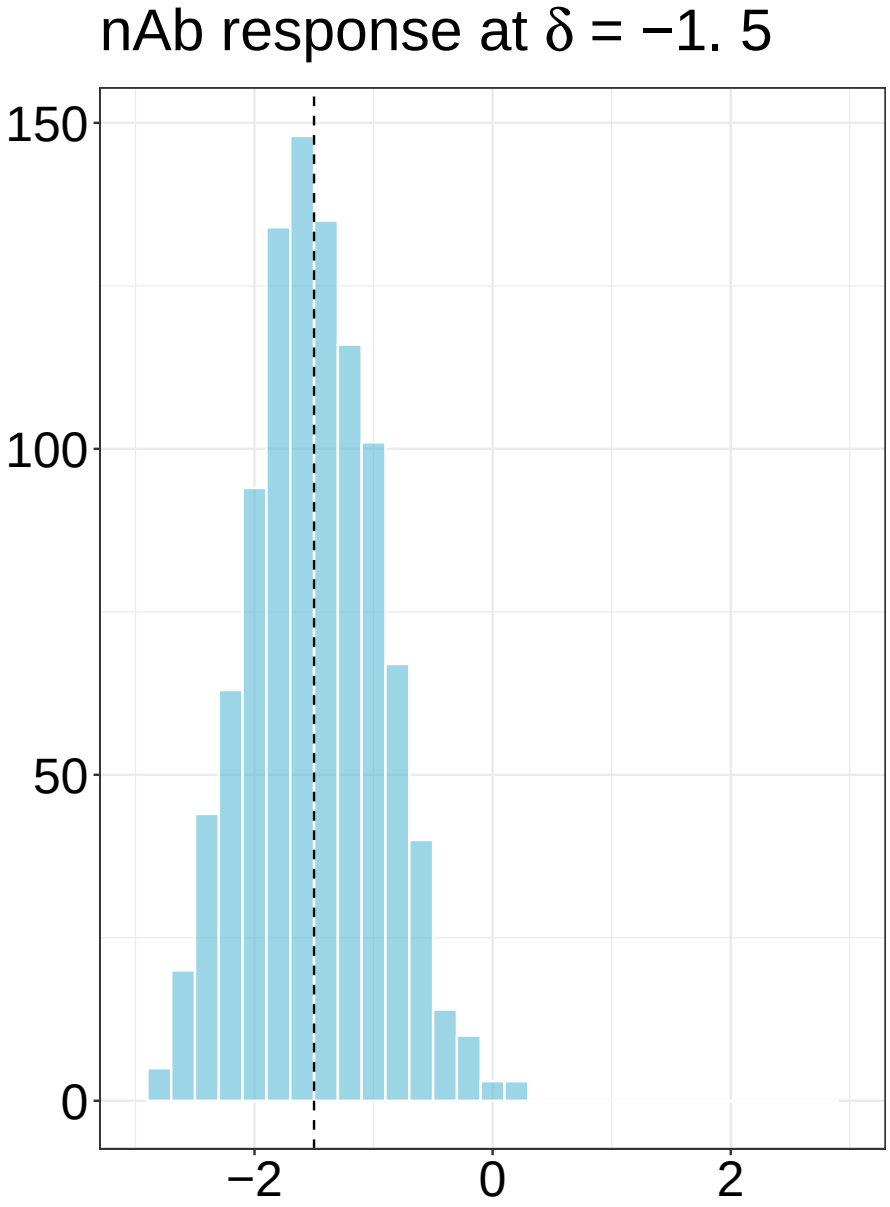


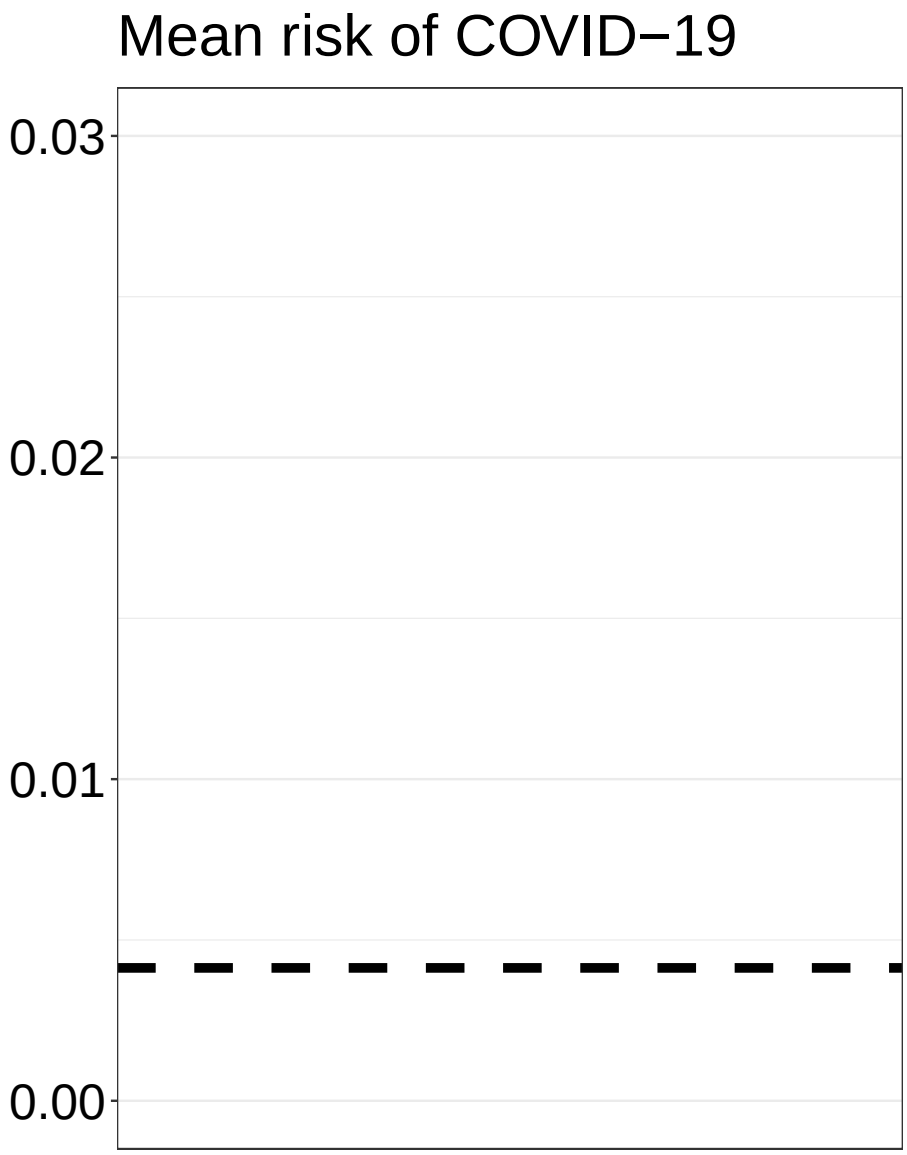
Disease risk under modified immunogenic responses





Disease risk under modified immunogenic responses





Stochastic interventions and modified treatment policies

- Stochastic interventions modify the value S would naturally assume by *modifying* the "natural" exposure distribution.
- Díaz and van der Laan (2012; 2018)'s shift interventions[1]

$$d(s, l) = \begin{cases} s + \delta, & s + \delta < u(l) \quad \text{(if plausible)} \\ s, & s + \delta \geq u(l) \quad \text{(otherwise)} \end{cases}$$

- Causal estimand $\psi_{0,\delta} := \mathbb{E}\{Y^{d(S,L)}\}$ is identified by

$$\psi_{0,\delta} = \int_{\mathcal{L}} \int_{\mathcal{S}} \mathbb{E}_{P_0}\{Y \mid S = d(s, l), L = l\} \\ g_{0,S}(s \mid L = l) q_{0,L}(l) d\mu(s) d\nu(l)$$

[1] Haneuse and Rotnitzky (2013) introduced *modified treatment policies* (MTPs), extended by Díaz and van der Laan (2018) and Díaz et al. (2021).

Causally interpreting the statistical target parameter

Assumption 1: *Stable Unit Treatment Value (SUTVA)*

- $Y_i^{d(s_i,l_i)}$ *does not depend on* $d(s_j, l_j)$ *for* $i = 1, \ldots, n$ *and* $j \neq i$*, or lack of interference (Cox 1958)*
- $Y^{d(s,l)} = Y$ *in the event* $S = d(s, l)$

Assumption 2: *No Unmeasured Confounding*

$Y^{d(s,l)} \perp\!\!\!\perp S \mid L = l$

Assumption 3: *Structural positivity*

$s \in \mathcal{S} \implies d(s, l) \in \mathcal{S}$ *for all* $l \in \mathcal{L}$*, where* $\mathcal{S}$ *denotes the support of* S *conditional on* $L = l$

Stochastic–interventional vaccine efficacy (SVE)

- Stochastic–interventional vaccine efficacy (SVE) estimand:

$$\begin{aligned}\mathrm{SVE}(\delta) &= 1 - \frac{\mathbb{E}[\mathbb{P}(Y = 1 \mid S = d(s, l), A = 1, L = l)]}{\mathbb{P}(Y(0) = 1)} \\ &= 1 - \frac{\psi_{0,\delta}}{\mathbb{P}(Y(0) = 1)}\end{aligned}$$

- $\mathbb{P}(Y(0) = 1)$: counterfactual infection risk in the placebo arm — under randomization, $\mathbb{P}(Y(0) = 1) \equiv \mathbb{P}(Y = 1 \mid A = 0)$.
- Summarizes VE as shifts of S by δ, yielding a VE measure based on hypothetically modifying S in the vaccine arm.

[2] CoVPN SAP: `https://doi.org/10.6084/m9.figshare.13198595`.

Estimation of the counterfactual mean $\psi_{0,\delta}$

A RAL *estimator* $\psi_{n,\delta}$ of $\psi_{0,\delta} := \Psi(P_0)$ is asymptotically *efficient* if

$$\psi_{n,\delta} - \psi_{0,\delta} = \frac{1}{n} \sum_{i=1}^{n} D^{\star}(P_0)(O_i) + o_P(n^{-1/2}) \ ,$$

where $D^{\star}(P)$ is the *efficient influence function* (EIF) of $\psi_{0,\delta}$ with respect to the nonparametric model $\mathcal{M}$ at some distribution P.

The EIF of $\psi_{0,\delta}$ is indexed by two key *nuisance parameters*

$$\overline{Q}_Y(S, L) := \mathbb{E}_P(Y \mid S, L) \qquad \textit{conditional outcome mean}$$
$$g_S(S \mid L) := f_P(S \mid L) \qquad \textit{generalized propensity score}$$

Ensemble machine learning to estimate $\overline{Q}_Y(S, L)$ and $g_S(S \mid L)$.

Flexible, efficient, and doubly robust estimation

- Efficient influence function (EIF) of $\psi_{0,\delta}$ with respect to $\mathcal{M}$ is

$$D^\star(P_0)(o) = \frac{g_{0,S}(d^{-1}(s,l) \mid l)}{g_{0,S}(s \mid l)}(y - \overline{Q}_{0,Y}(s,l)) + \overline{Q}_{0,Y}(d(s,l),l) - \psi_{0,\delta} \ .$$

- One-step estimator performs additive bias-correction:

$$\psi_n^+ = \frac{1}{n}\sum_{i=1}^{n} \overline{Q}_{n,Y}(d(S_i,L_i),L_i) + D_n^\star(O_i) \ .$$

- TML estimator tilts initial estimator $\overline{Q}_{n,Y}$ for bias-correction:

$$\psi_n^\star = \frac{1}{n}\sum_{i=1}^{n} \overline{Q}_{n,Y}^\star(d(S_i,L_i),L_i) \ .$$

- *Doubly robust*: Consistent even if $\overline{Q}_{n,Y}$ or $g_{n,S}$ incorrect.

Act II:

Revenge of the two-phase sampling design

(“There ain’t nothing in this world for free”)

Augmented Estimators for Two-Phase Sampling Designs

- Rose and van der Laan (2011) suggested inverse probability of censoring weighted (IPCW) loss functions:

$$\mathcal{L}(P_0^X)(O) = \frac{B}{\pi_0(Y, L)} \mathcal{L}(P_0^X)(X)$$

- When the sampling mechanism $\pi_0(Y, L)$ is known by design, this procedure yields a reasonably reliable estimator.
- When data-adaptive regression must be used — that is, when $\pi_0(Y, L)$ is not known by design[2]— this is insufficient.

[3]Sampling of non-cases in HVTN 505 used matching (Janes et al. 2017).

Efficiency and Multiple Robustness (Hejazi et al. 2021)

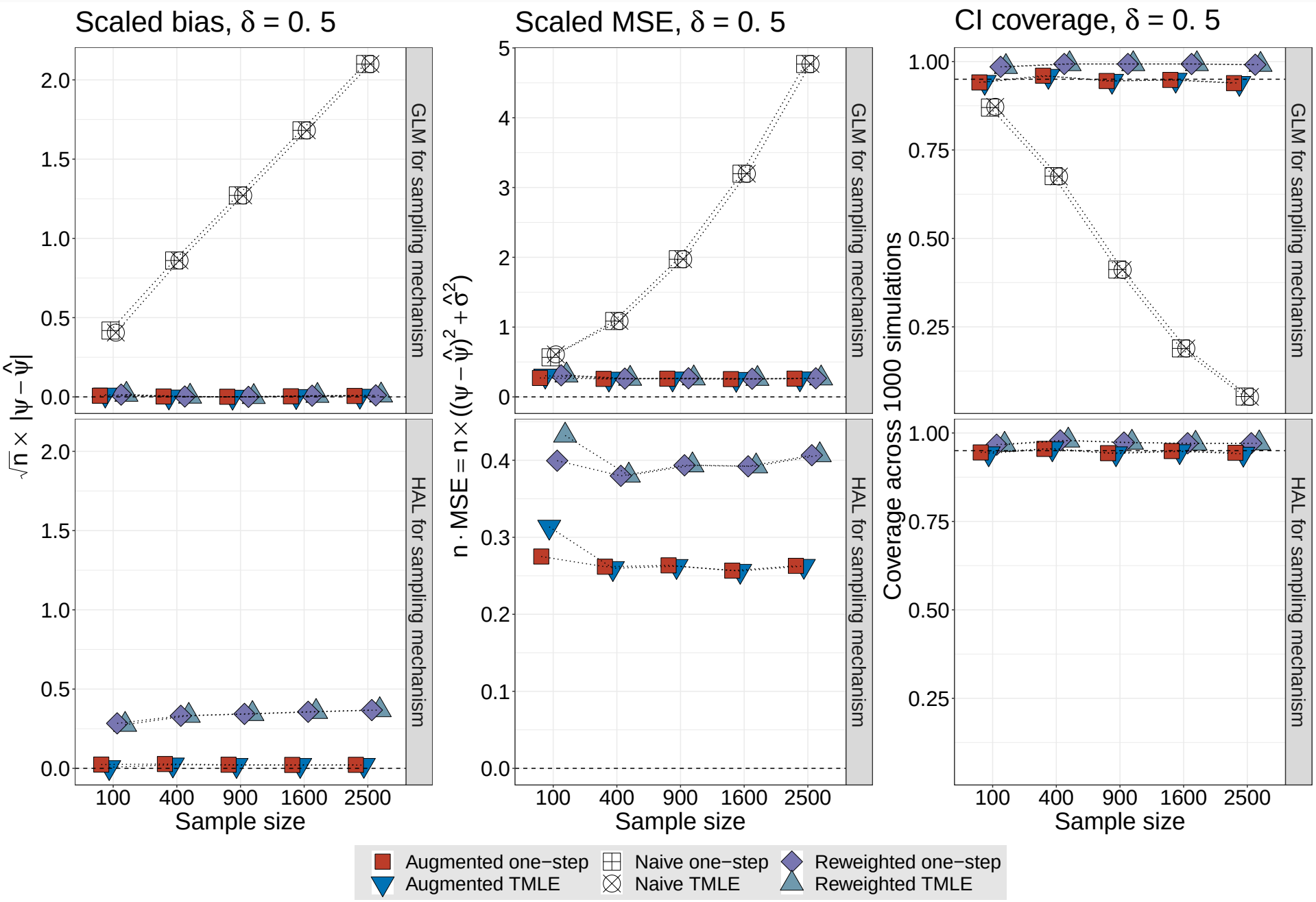
- Then, the IPCW augmentation must be applied to the EIF[3]:

$$D^\star(P_0^X)(o) = \frac{b}{\pi_0(y, l)} D_F^\star(P_0^X)(x) - \left(1 - \frac{b}{\pi_0(y, l)}\right) \mathbb{E}(D_F^\star(P_0^X)(x) \mid B = 1, Y = y, L = l).$$

- Expresses observed data EIF $D^\star(P_0^X)(o)$ via complete data EIF $D_F^\star(P_0^X)(x)$; inclusion of second term improves efficiency.
- An emergent multiple robustness property — combinations of $\{g_0(S \mid L), \overline{Q}_0(S, L)\} \times \{\pi_0(Y, L), \mathbb{E}(D_F^\star(P_0^X)(x) \mid B = 1, Y, L)\}$.
- Our `txshift` `R` package implements our estimators of $\psi_{0,\delta}$.

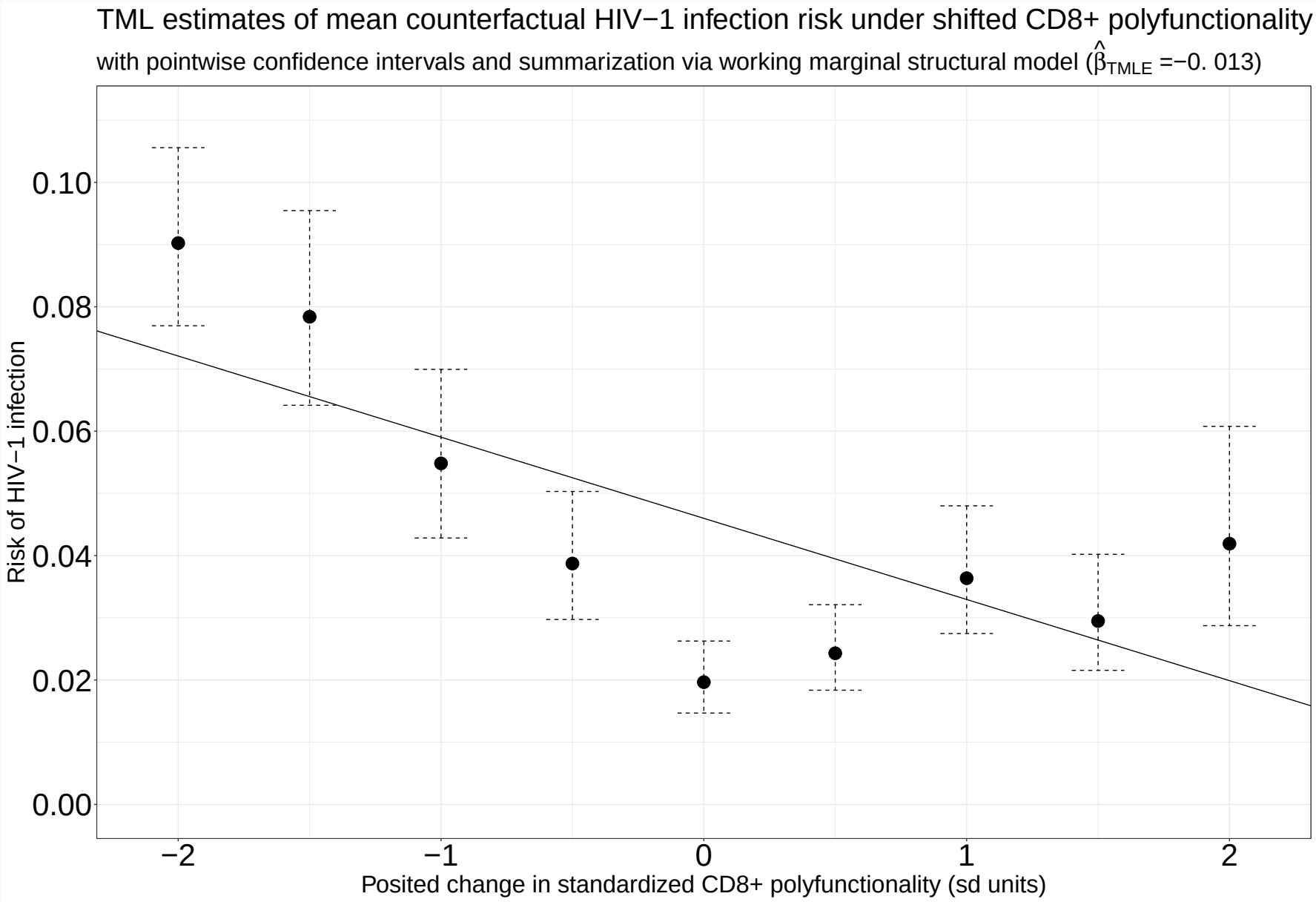
[4] Robins et al. (1994) explored a similar correction for such designs.

Comparing Reweighted and Augmented Estimators



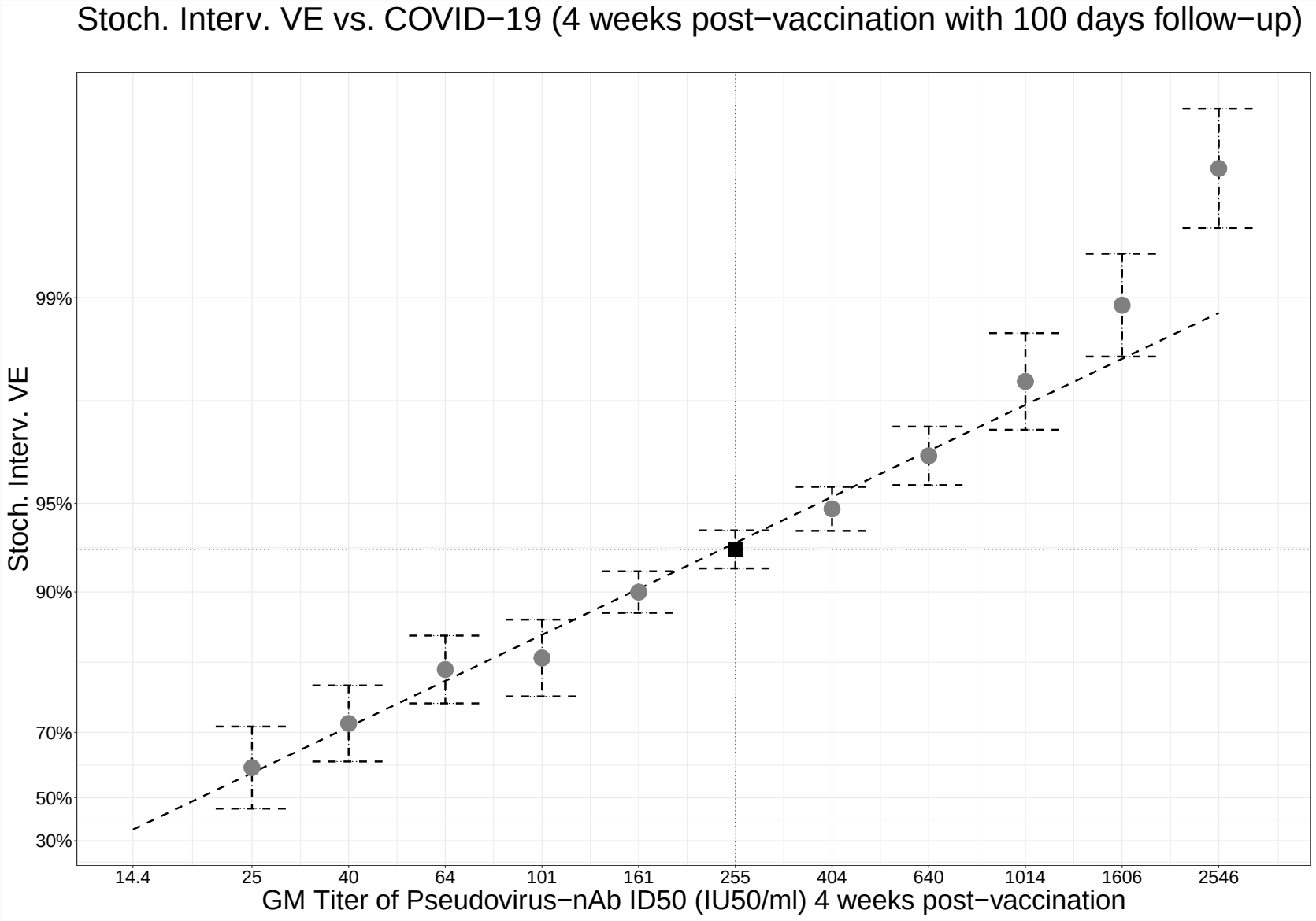
Act III:
Prediction and immunobridging via SVE

Prediction of HIV-1 risk (Hejazi et al. 2021)

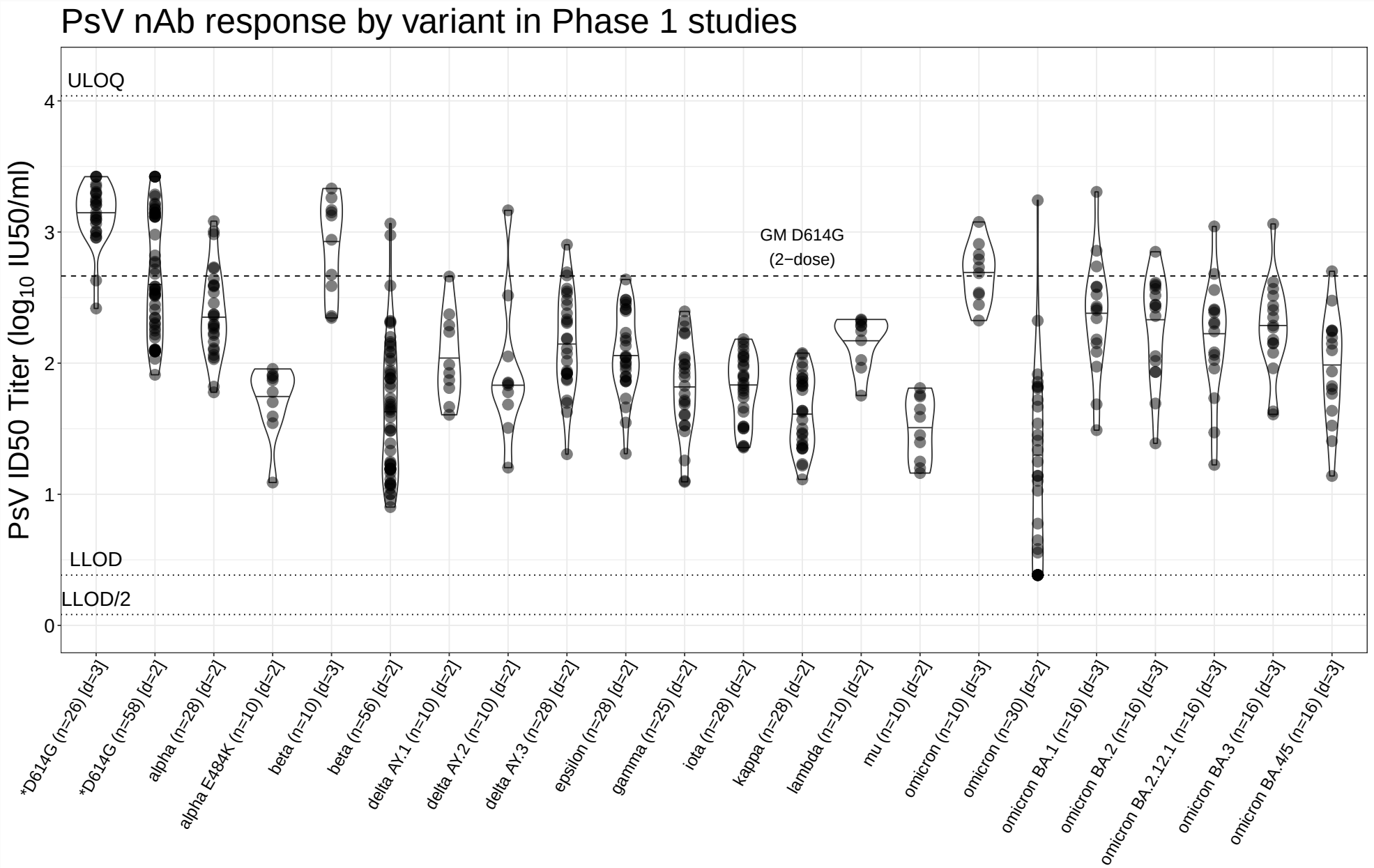


HIV-1 risk change across CD8+ score (`txshift` `R` package).

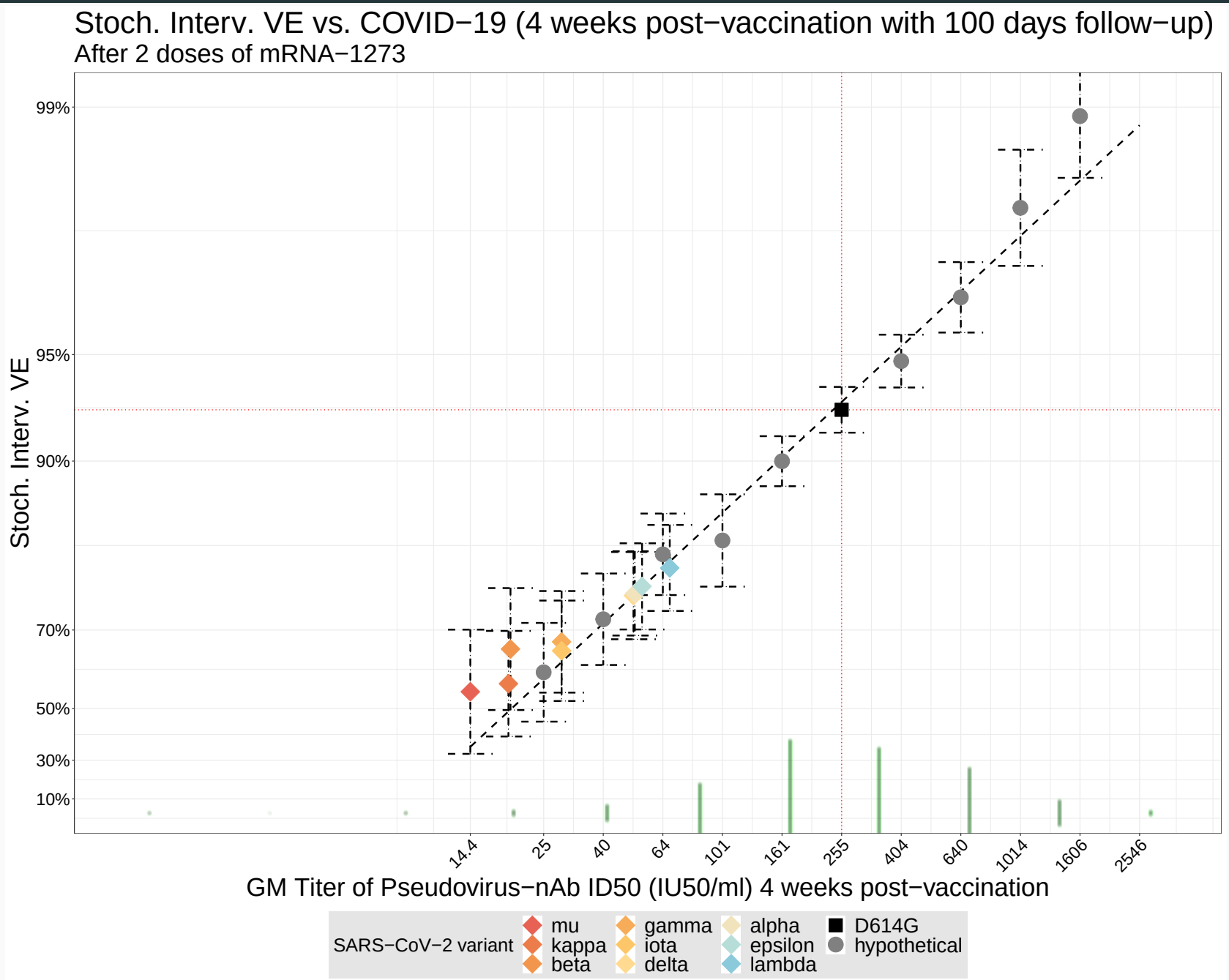
SVE *prediction* of mRNA-1273 VE (Huang et al. 2023)



Pooled phase 1 studies: PsV nAb responses across variants



SVE *bridging* of mRNA-1273 VE (Hejazi et al. 2023)



Summary of SVE prediction for immunobridging

- SVE prediction shows sharp changes in VE with shifts to the GM titer of the PsV nAb correlate in vaccinees.
- Bridging VE across variants indicates VE drops but stabilizes at 50%, if the model based on ancestral D614G strain holds.
- Post-2^{nd} dose: For most variants (excepting omicron), the VE estimate ranges from 50% (mu) to about 80% (lambda).
- SVE predictions and real-world VE estimates well-correlated, but SVE predictions may be underestimates, as PsV nAb correlate is an *imperfect* causal *mediator* of total VE.

The big picture

- Flexible, realistic interventions can be used to formulate modified treatment policies (based on "natural" exposure).
- *Modified treatment policies* address causal questions about *hypothetical* manipulations of quantitative variables.
- Efficient estimators with double/multiple robustness can safely answer such questions *while* incorporating machine learning.
- Applying machine learning with causal inference yields robust predictions of VE and evidence for immunobridging.
- Open source software for such statistical analyses is critical for the methods to have any impact on real-world studies.

Thank you

Thanks for your attention! Any questions?

https://nimahejazi.org

https://github.com/nhejazi

https://twitter.com/nshejazi

Biometrics: https://doi.org/10.1111/biom.13375

IJID: https://doi.org/10.1016/j.ijid.2023.09.012

Viruses: https://doi.org/10.3390/v15102029

Vaccine: https://doi.org/10.1016/j.vaccine.2024.02.071

Appendix

Immune Correlates of Protection (Plotkin and Gilbert 2012)

- Correlate of Protection (CoP): immune marker statistically predictive of vaccine efficacy, not necessarily mechanistic.
- Mechanistic CoP (mCoP): immune marker that is causally and mechanistically responsible for protection.
- Nonmechanistic CoP (nCoP): immune marker that is predictive but not a causal agent of protection.
- A CoP is a *candidate surrogate* endpoint (Prentice 1989) — primary endpoint in future trials if reliably predictive.

Literature: Díaz and van der Laan (2012; 2018)

- *Proposal:* Evaluate outcome under an altered *intervention distribution* — e.g., $P_\delta(g_{0,S})(S = s \mid L) = g_{0,S}(s - \delta(L) \mid L)$.
- Identification conditions for a statistical parameter of the counterfactual outcome $\psi_{0,\delta}$ under such an intervention.
- Show that the causal quantity of interest $\mathbb{E}_{P_0^\delta}\{Y_{d(S,L)}\}$ is identified by a functional of the distribution of O, i.e.,

$$\psi_{0,\delta} = \int_{\mathcal{L}} \int_{\mathcal{S}} \mathbb{E}_{P_0}\{Y \mid S = d(s, l), L = l\} \\ g_{0,S}(s \mid L = l) \cdot q_{0,L}(l) d\mu(s) d\nu(l)$$

Literature: Haneuse and Rotnitzky (2013)

- *Proposal:* Characterization of stochastic interventions as *modified treatment policies* (MTPs).
- Assumption of *piecewise smooth invertibility* allows for the post-intervention distribution of any MTP to be recovered:

$$g_{0,S}(s \mid l; \delta) = \sum_{j=1}^{J(l)} \mathbb{I}_{\delta,j}\{h_j(s,l), l\} g_0\{h_j(s,l) \mid l\} h_j'(s,l)$$

- MTPs account for the natural value of exposure S yet may be interpreted as imposing an altered intervention mechanism.

Literature: Young et al. (2014)

- Establishes equivalence between g-formula when proposed intervention depends on natural value and when it does not.
- This equivalence leads to a sufficient positivity condition for estimating the counterfactual mean under MTPs via the same statistical functional studied in Díaz and van der Laan (2012).
- Extends earlier identification results, providing a way to use the same statistical functional to assess $\mathbb{E}Y_{d(S,L)}$ or $\mathbb{E}Y_{d(L)}$.
- The authors also consider limits on implementing shifts $d(S, L)$, and address working in a longitudinal setting.

Flexible conditional density estimation of $g_{0,S}$

- Díaz and van der Laan (2011)'s conditional density estimator:

$$g_{n,\alpha}(s \mid L) = \frac{\mathbb{P}(s \in [\alpha_{t-1}, \alpha_t) \mid L)}{\alpha_t - \alpha_{t-1}}.$$

  - Re-expressed as hazard regressions in repeated measures data.
  - Tuning parameter $t \approx$ bandwidth in kernel density estimation.

- When càdlàg (RCLL) with finite sectional variation, we have

$$\text{logit}\{\mathbb{P}(s \in [\alpha_{t-1}, \alpha_t) \mid L)\} = \beta_0 + \sum_{w \subset \{1,\ldots,d\}} \sum_{i=1}^{n} \beta_{w,i} \phi_{w,i},$$

for appropriate basis functions $\{\phi_{w,i}\}_{i=1}^{n}$ (Gill et al. 1995).

Flexible conditional density estimation of $g_{0,S}$

- Utilizing a particular basis construction for ϕ_w, van der Laan (2017)'s HAL estimator achieves $n^{-1/4}$ convergence rate[4].
- Loss-based cross-validation allows selection of a suitable HAL estimator, which has only the ℓ_1 regularization term λ:

$$\beta_{n,\lambda} = \underset{\beta:|\beta_0|+\sum_{w\subset\{1,\ldots,d\}}\sum_{i=1}^{n}|\beta_{w,i}|<\lambda}{\arg\min} P_n\mathcal{L}(g_{\beta,\lambda,S}),$$

  where $\mathcal{L}(\cdot)$ is an appropriate loss function, giving $\{\lambda_n, \beta_n\}$.

- We denote by $g_{n,\lambda,S} := g_{\beta_{n,\lambda},S}$, the HAL estimate of $g_{0,S}$.
- Our `haldensify` `R` package implements our estimator of $g_{0,S}$.

[6]Similar rates can be achieved via *local* (vs. global) smoothness assumptions on $g_{n,S}$ (see, e.g., Robins et al. 2008, Mukherjee et al. 2017, Liu et al. 2021).

Algorithm for TML estimation

1. Construct initial estimators g_n of $g_0(S, L)$ and Q_n of $\overline{Q}_0(S, L)$, perhaps using data-adaptive regression techniques.
2. For each observation i, compute an estimate $H_n(s_i, l_i)$ of the auxiliary covariate $H(s_i, l_i)$.
3. Estimate the parameter ϵ in the logistic regression model

$$\text{logit}\overline{Q}_{\epsilon,n}(s, l) = \text{logit}\overline{Q}_n(s, l) + \epsilon H_n(s, l),$$

   or an alternative regression model incorporating weights.
4. Compute TML estimator Ψ_n of the target parameter, defining update $\overline{Q}_n^\star$ of the initial estimate $\overline{Q}_{n,\epsilon_n}$:

$$\Psi_n = \Psi(P_n^\star) = \frac{1}{n}\sum_{i=1}^{n}\overline{Q}_n^\star(d(S_i, L_i), L_i).$$

Algorithm for IPCW-TML estimation

1. Using all observed units (X), estimate sampling mechanism $\pi(Y, L)$, perhaps using data-adaptive regression methods.

2. Using only observed units in the phase-two sample $B = 1$, construct initial estimators $g_n(S, L)$ and $\overline{Q}_n(S, L)$, weighting by the sampling mechanism estimate $\pi_n(Y, L)$.

3. With the approach described for the full data case, compute $H_n(s_i, l_i)$, and fluctuate submodel via logistic regression.

4. Compute IPCW-TML estimator Ψ_n of the target parameter, by solving the IPCW-augmented EIF estimating equation.

5. Iteratively update estimated sampling weights $\pi_n(Y, L)$ and IPCW-augmented EIF, updating TMLE in each iteration.

References

Antia, R. and Halloran, M. E. (2021). Transition to endemicity: Understanding covid-19. *Immunity*, 54(10):2172–2176.

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., Gilbert, P., Janes, H., Follmann, D., Marovich, M., Mascola, J., Polakowski, L., Ledgerwood, J., Graham, B. S., Bennett, H., Pajon, R., Knightly, C., Leav, B., Deng, W., Zhou, H., Han, S., Ivarsson, M., Miller, J., Zaks, T., and the COVE Study Group (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine*, 384(5):403–416.

Breslow, N., McNeney, B., Wellner, J. A., et al. (2003). Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *The Annals of Statistics*, 31(4):1110–1139.

Cox, D. R. (1958). *Planning of Experiments*. Wiley.

Díaz, I. and van der Laan, M. J. (2011). Super learner based conditional density estimation with application to marginal structural models. *International Journal of Biostatistics*, 7(1):1–20.

Díaz, I. and van der Laan, M. J. (2012). Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549.

Díaz, I. and van der Laan, M. J. (2018). Stochastic treatment regimes. In *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*, pages 167–180. Springer Science & Business Media.

Díaz, I., Williams, N., Hoffman, K. L., and Schenck, E. J. (2021). Nonparametric causal effects based on longitudinal modified treatment policies. *Journal of the American Statistical Association*.

Gilbert, P. B., Fong, Y., Hejazi, N. S., Kenny, A., Huang, Y., Carone, M., Benkeser, D., and Follmann, D. (2024). Four statistical frameworks for assessing an immune correlate of protection (surrogate endpoint) from a randomized, controlled, vaccine efficacy trial. *Vaccine*, 42(9):2181–2190.

Gilbert, P. B., Fong, Y., Kenny, A., and Carone, M. (2022). A controlled effects approach to assessing immune correlates of protection. *Biostatistics*, (in press).

Gill, R. D., van der Laan, M. J., and Wellner, J. A. (1995). Inefficient estimators of the bivariate survival function for three models. In *Annales de l'IHP Probabilités et statistiques*, volume 31, pages 545–597.

Hammer, S. M., Sobieszczyk, M. E., Janes, H., Karuna, S. T., Mulligan, M. J., Grove, D., Koblin, B. A., Buchbinder, S. P., Keefer, M. C., Tomaras, G. D., et al. (2013). Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *New England Journal of Medicine*, 369(22):2083–2092.

Haneuse, S. and Rotnitzky, A. (2013). Estimation of the effect of interventions that modify the received treatment. *Statistics in Medicine*, 32(30):5260–5277.

Hejazi, N. S., Shen, X., Carpp, L. N., Benkeser, D., Follmann, D., Janes, H. E., Baden, L. R., El Sahly, H. M., Deng, W., Zhou, H., Leav, B., Montefiori, D. C., and Gilbert, P. B. (2023). Stochastic interventional correlates of protection analysis of the COVE trial, with application to predict mRNA-1273 vaccine efficacy against SARS-CoV-2 variants. under review at *Lancet Microbe*.

Hejazi, N. S., van der Laan, M. J., Janes, H. E., Gilbert, P. B., and Benkeser, D. C. (2021). Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*.

Huang, Y., Hejazi, N. S., Blette, B., Carpp, L. N., Benkeser, D., Montefiori, D. C., McDermott, A. B., Fong, Y., Janes, H. E., Deng, W., Zhou, H., Houchens, C. R., Martins, K. A., Jayashankar, L., Flach, B., Lin, B. C., O'Connell, S., McDanal, C., Eaton, A., Sarzotti-Kelsoe, M., Lu, Y., Yu, C., Kenny, A., Carone, M., Huynh, C., Miller, J., El Sahly, H. M., Baden, L. R., Jackson, L. A., Campbell, T. B., Clark, J. L., Andrasik, M. P., Kublin, J. G., Corey, L., Neuzil, K. M., Pajon, R., Follmann, D. A., Donis, R. O., Koup, R. A., Gilbert, P. B., and on behalf of the Immune Assays Team; Moderna, Inc., Team; Coronavirus Prevention Network (CoVPN)/Coronavirus Efficacy (COVE) Team; and the United States Government (USG)/CoVPN Biostatistics Teams (2023). Stochastic interventional vaccine efficacy and principal surrogate analyses of antibody markers as correlates of protection against symptomatic COVID-19 in the COVE mRNA-1273 trial. *Viruses*, 15(10):2029.

Janes, H. E., Cohen, K. W., Frahm, N., De Rosa, S. C., Sanchez, B., Hural, J., Magaret, C. A., Karuna, S., Bentley, C., Gottardo, R., et al. (2017). Higher t-cell responses induced by dna/rad5 hiv-1 preventive vaccine are associated with lower hiv-1 infection risk in an efficacy trial. *The Journal of Infectious Diseases*, 215(9):1376–1385.

Liu, L., Mukherjee, R., Robins, J. M., and Tchetgen Tchetgen, E. (2021). Adaptive estimation of nonparametric functionals. *Journal of Machine Learning Research*, 22(99):1–66.

Mukherjee, R., Newey, W. K., and Robins, J. M. (2017). Semiparametric efficient empirical higher order influence function estimators. *arXiv preprint arXiv:1705.07577*.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Plotkin, S. A. and Gilbert, P. B. (2012). Nomenclature for immune correlates of protection after vaccination. *Clinical Infectious Diseases*, 54(11):1615–1617.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73(1):1–11.

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine*, 8(4):431–440.

Robins, J. M., Li, L., Tchetgen Tchetgen, E., and van der Vaart, A. W. (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and statistics: essays in honor of David A. Freedman*, pages 335–421. Institute of Mathematical Statistics.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association*, 89(427):846–866.

Rose, S. and van der Laan, M. J. (2011). A targeted maximum likelihood estimator for two-stage designs. *The International Journal of Biostatistics*, 7(1):1–21.

van der Laan, M. J. (2017). A generally efficient targeted minimum loss based estimator based on the highly adaptive lasso. *International Journal of Biostatistics*, 13(2).

Young, J. G., Hernán, M. A., and Robins, J. M. (2014). Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods*, 3(1):1–19.