Addressing Confounding and Continuous Exposure Measurement Error Using Corrected Score Functions

Brian Richardson

January 10, 2024

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH
Biostatistics

Acknowledgements

Bryan Blette, PhD

Vanderbilt University

Peter Gilbert, PhD

University of Washington Medical Center

Michael Hudgens, PhD

UNC Chapel Hill

This research was supported by the U.S. Public Health Service Grant AI068635, the National Institute of Allergy And Infectious Diseases of the National Institutes of Health (NIH) under Award Number R37 AI054165, and the National Institute of Environmental Health Sciences of the NIH under Award Number T32 ES007018

Motivation: HVTN 505 trial

- **HVTN 505 trial**: trial of a preventive HIV vaccine
- stopped administering immunizations early after reaching predetermined cutoffs for efficacy futility [Hammer et al., 2013]

The NEW ENGLAND JOURNAL of MEDICINE

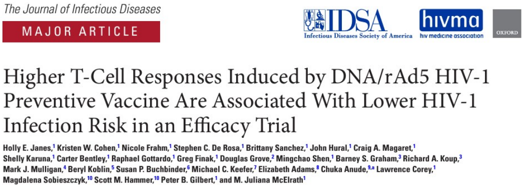
ESTABLISHED IN 1812 NOVEMBER 28, 2013 VOL. 369 NO. 22

Efficacy Trial of a DNA/rAd5 HIV-1 Preventive Vaccine

Scott M. Hammer, M.D., Magdalena E. Sobieszczyk, M.D., M.P.H., Holly Janes, Ph.D., Shelly T. Karuna, M.D., Mark J. Mulligan, M.D., Doug Grove, M.S., Beryl A. Koblin, Ph.D., Susan P. Buchbinder, M.D., Michael C. Keefer, M.D., Georgia D. Tomaras, Ph.D., Nicole Frahm, Ph.D., John Hural, Ph.D., Chuka Anude, M.D., Ph.D., Barney S. Graham, M.D., Ph.D., Mary E. Enama, M.A., P.A.-C., Elizabeth Adams, M.D., Edwin DeJesus, M.D., Richard M. Novak, M.D., Ian Frank, M.D., Carter Bentley, Ph.D., Shelly Ramirez, M.A., Rong Fu, M.S., Richard A. Koup, M.D., John R. Mascola, M.D., Gary J. Nabel, M.D., Ph.D., David C. Montefiori, Ph.D., James Kublin, M.D., M.P.H., M. Juliana McElrath, M.D., Ph.D., Lawrence Corey, M.D., and Peter B. Gilbert, Ph.D., for the HVTN 505 Study Team*

Motivation: HVTN 505 trial

- later analyses identified several immunologic biomarker correlates of HIV acquisition among HIV vaccine recipients [Janes et al., 2017, Fong et al., 2018, Neidich et al., 2019]
- of interest to assess the effect of these biomarkers on risk of HIV acquisition
- measurement of the biomarkers is subject to error and the association between the biomarkers and HIV risk is likely confounded


The Journal of Infectious Diseases
MAJOR ARTICLE

Higher T-Cell Responses Induced by DNA/rAd5 HIV-1 Preventive Vaccine Are Associated With Lower HIV-1 Infection Risk in an Efficacy Trial

Holly E. Janes,[1] Kristen W. Cohen,[1] Nicole Frahm,[1] Stephen C. De Rosa,[1] Brittany Sanchez,[1] John Hural,[1] Craig A. Magaret,[1] Shelly Karuna,[1] Carter Bentley,[1] Raphael Gottardo,[1] Greg Finak,[1] Douglas Grove,[2] Mingchao Shen,[1] Barney S. Graham,[3] Richard A. Koup,[3] Mark J. Mulligan,[4] Beryl Koblin,[5] Susan P. Buchbinder,[6] Michael C. Keefer,[7] Elizabeth Adams,[8] Chuka Anude,[9,a] Lawrence Corey,[1] Magdalena Sobieszczyk,[10] Scott M. Hammer,[10] Peter B. Gilbert,[1] and M. Juliana McElrath[1]

Goal

To estimate the effect of a continuous exposure on an outcome when

(i) the exposure-outcome association is potentially confounded

(ii) the exposure is measured with error

Approach

To estimate the effect of a continuous exposure on an outcome when

(i) the exposure-outcome association is potentially confounded
   (a) g-formula
   (b) inverse probability weighting (IPW)
   (c) doubly-robust (DR)

(ii) the exposure is measured with error
   (a) corrected score (CS) method

Notation

- true exposure: $\boldsymbol{A} = (A_1, \ldots, A_m)$
- measured exposure: $\boldsymbol{A}^* = (A_1^*, \ldots, A_m^*) = \boldsymbol{A} + \boldsymbol{\epsilon}$
- measurement error: $\boldsymbol{\epsilon} \sim \mathcal{N}_m(0, \boldsymbol{\Sigma}_{me})$
- potential outcome: $Y(\boldsymbol{a})$
- observed outcome: Y
- confounders: $\boldsymbol{L} = (L_1, L_2, ..., L_p)$

Observe: iid copies of $(Y_i, \boldsymbol{A}_i^*, \boldsymbol{L}_i)$.

Estimand: dose-response curve $\eta(\boldsymbol{a}) \equiv \mathrm{E}[Y(\boldsymbol{a})]$ for $\boldsymbol{a} \in \boldsymbol{\mathcal{A}}$.

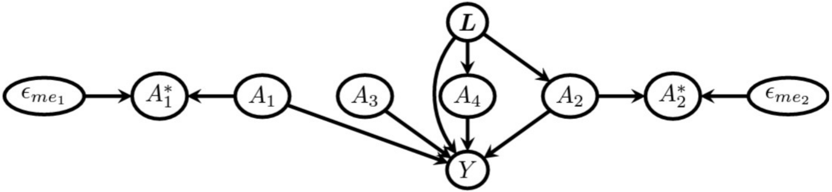
Assumptions

(i) **causal consistency**: $Y = Y(\boldsymbol{a})$ when $\boldsymbol{A} = \boldsymbol{a}$

(ii) **conditional exchangeability**: $Y(\boldsymbol{a}) \perp\!\!\!\perp \boldsymbol{A}|\boldsymbol{L}$ for all $\boldsymbol{a} \in \boldsymbol{\mathcal{A}}$

(iii) **positivity**: $f_{\boldsymbol{A}|\boldsymbol{L}}(\boldsymbol{a}|\boldsymbol{l}) > 0$ for all $\boldsymbol{l}$ such that $f_{\boldsymbol{L}}(\boldsymbol{l}) > 0$ and for all $\boldsymbol{a} \in \boldsymbol{\mathcal{A}}$

(iv) **independent measurement error**: $\boldsymbol{\epsilon} \perp\!\!\!\perp (Y, \boldsymbol{A}, \boldsymbol{L})$



Addressing Confounding: G-Formula

- fit the **outcome model** $\mu(\boldsymbol{L}, \boldsymbol{A}; \beta) \equiv \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{A})$
- estimate the dose-response curve by marginalizing over the distribution of confounders: $\widehat{\eta}(\boldsymbol{a}) = n^{-1} \sum_{i=1}^{n} \mu(\boldsymbol{L}_i, \boldsymbol{a}; \widehat{\beta})$
- This can be expressed as an M-estimator with estimating function

$$\boldsymbol{\Psi}_{0-GF}(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}_{GF}) = \begin{bmatrix} \{Y - \mu(\boldsymbol{L}, \boldsymbol{A}; \beta)\} \, \partial_{\boldsymbol{\beta}} \mu(\boldsymbol{L}, \boldsymbol{A}; \beta) \\ \eta(\boldsymbol{a}) - \mu(\boldsymbol{L}, \boldsymbol{a}; \beta) \end{bmatrix}$$

Addressing Confounding: IPW

- obtain/estimate **standardized propensity score weights**

$$SW(\boldsymbol{L}, \boldsymbol{A}) = \frac{f_{\boldsymbol{A}}(\boldsymbol{A})}{f_{\boldsymbol{A}|\boldsymbol{L}}(\boldsymbol{A}|\boldsymbol{L})}$$

- use weighted observations to estimate the dose-response curve $\eta(\boldsymbol{a}; \boldsymbol{\gamma})$
- This can be expressed as an M-estimator with estimating function

$$\boldsymbol{\Psi}_{0-IPW}(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}_{IPW}) = \begin{bmatrix} \boldsymbol{\Psi}_{PS}(\boldsymbol{L}, \boldsymbol{A}) \\ SW(\boldsymbol{L}, \boldsymbol{A}) \left\{ Y - \eta(\boldsymbol{A}; \boldsymbol{\gamma}) \right\} \partial_{\boldsymbol{\gamma}} \eta(\boldsymbol{A}; \boldsymbol{\gamma}) \end{bmatrix}$$

Addressing Confounding: DR

- obtain/estimate **standardized propensity score weights** $SW(\boldsymbol{L}, \boldsymbol{A})$
- use weighted observations to estimate the **outcome model** $\mu(\boldsymbol{L}, \boldsymbol{A}; \beta) \equiv \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{A})$
- estimate the dose-response curve by marginalizing over the distribution of confounders
- This can be expressed as an M-estimator with estimating function

$$\psi_{0-DR}(Y, \boldsymbol{L}, \boldsymbol{A}, \boldsymbol{\theta}_{DR}) = \begin{bmatrix} \boldsymbol{\Psi}_{PS}(\boldsymbol{L}, \boldsymbol{A}) \\ SW(\boldsymbol{L}, \boldsymbol{A})\{Y - \mu(\boldsymbol{L}, \boldsymbol{A}; \beta)\}\partial_{\boldsymbol{\beta}}\mu(\boldsymbol{L}, \boldsymbol{A}; \beta) \\ \eta(\boldsymbol{a}) - \mu(\boldsymbol{L}, \boldsymbol{a}; \beta) \end{bmatrix}$$

- doubly robust* to models for $\mu(\boldsymbol{L}, \boldsymbol{A}; \beta)$ and $f_{\boldsymbol{A}|\boldsymbol{L}}(\boldsymbol{A}|\boldsymbol{L})$.

How to Address Measurement Error?

- The three estimators proposed so far are solutions to **estimating equations** $\sum_{i=1}^{n} \boldsymbol{\Psi}_0(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i; \boldsymbol{\theta}) = \mathbf{0}$ that are **unbiased** in the sense that $\mathrm{E}\{\boldsymbol{\Psi}_0(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta})\} = \mathbf{0}$.
- **Problem**: the true exposure values $\boldsymbol{A}$ are not observed. Instead we observe the mismeasured version $\boldsymbol{A}^* = \boldsymbol{A} + \boldsymbol{\epsilon}$.
- **Solution**: Create a **corrected score** function $\boldsymbol{\Psi}_{CS}$ that takes $\boldsymbol{A}^*$ as an argument and satisfies

$$
\begin{aligned}
&\mathrm{E}\left\{\boldsymbol{\Psi}_{CS}\left(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}\right) \mid Y, \boldsymbol{L}, \boldsymbol{A}\right\} = \boldsymbol{\Psi}_0\left(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}\right) \\
\implies\ &\mathrm{E}\left[\mathrm{E}\left\{\boldsymbol{\Psi}_{CS}\left(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}\right) \mid Y, \boldsymbol{L}, \boldsymbol{A}\right\}\right] = \mathrm{E}\left\{\boldsymbol{\Psi}_0\left(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}\right)\right\} \\
\implies\ &\mathrm{E}\left\{\boldsymbol{\Psi}_{CS}\left(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}\right)\right\} = \mathbf{0}
\end{aligned}
$$

Addressing Exposure Measurement Error: Corrected Score Functions

- Suppose the oracle estimating function is **conditionally unbiased**, meaning

$$\mathrm{E}\{\boldsymbol{\Psi}_0(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}) | \boldsymbol{A}\} = \boldsymbol{0}.$$

- Then we can create a corrected score function (following Novick and Stefanski [2002]) as

$$\boldsymbol{\Psi}_{CS}(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}) = \mathrm{E}\left[\mathrm{Re}\left\{\boldsymbol{\Psi}_0(Y, \boldsymbol{L}, \widetilde{\boldsymbol{A}}; \boldsymbol{\theta})\right\} | Y, \boldsymbol{L}, \boldsymbol{A}^*\right],$$

where $\widetilde{\boldsymbol{A}} = \boldsymbol{A}^* + i\widetilde{\boldsymbol{\epsilon}}$, $i = \sqrt{-1}$, $\mathrm{Re}(\cdot)$ denotes the real component of a complex number, and $\widetilde{\boldsymbol{\epsilon}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{me})$.

Computing Conditional Score Functions

The corrected score method involves evaluating an expectation of the form

$$\mathrm{E}[\mathrm{Re}\{\boldsymbol{\Psi}_0(Y, \boldsymbol{L}, \widetilde{\boldsymbol{A}}; \boldsymbol{\theta})\} | Y, \boldsymbol{L}, \boldsymbol{A}^*]$$

- sometimes this expectation has a closed form
- can also be approximated with the **Monte-Carlo corrected score** (MCCS) function

$$\boldsymbol{\Psi}^B_{MCCS}(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}) = B^{-1} \sum_{b=1}^{B} \mathrm{Re}\left\{\boldsymbol{\Psi}_0(Y, \boldsymbol{L}, \widetilde{\boldsymbol{A}}_b; \boldsymbol{\theta})\right\},$$

where $\widetilde{\boldsymbol{A}}_b = \boldsymbol{A}^* + i\widetilde{\boldsymbol{\epsilon}}_b$, and $\widetilde{\boldsymbol{\epsilon}}_b$ are iid simulated measurement errors.

Addressing Confounding and Exposure Measurement Error

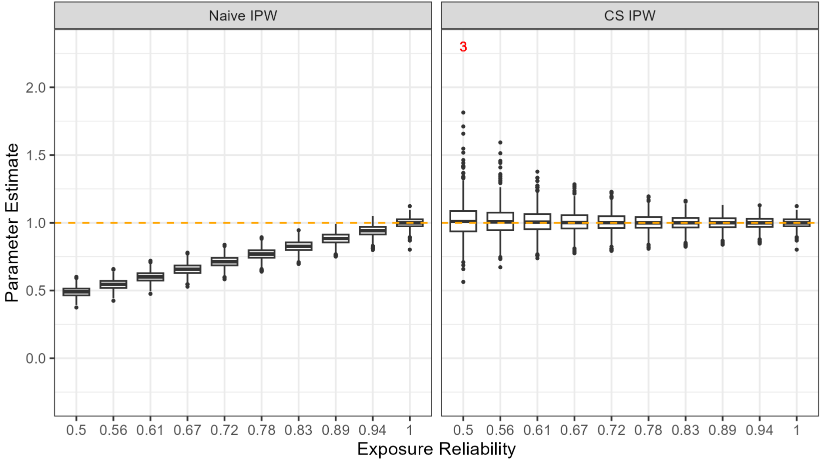
The corrected score method can be applied to the g-formula, IPW, and DR estimators

$$\boldsymbol{\Psi}_{0-GF}(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}_{GF}) \longrightarrow \boldsymbol{\Psi}_{CS-GF}(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}_{GF})$$
$$\boldsymbol{\Psi}_{0-IPW}(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}_{IPW}) \longrightarrow \boldsymbol{\Psi}_{CS-IPW}(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}_{IPW})$$
$$\boldsymbol{\Psi}_{0-DR}(Y, \boldsymbol{L}, \boldsymbol{A}; \boldsymbol{\theta}_{DR}) \longrightarrow \boldsymbol{\Psi}_{CS-DR}(Y, \boldsymbol{L}, \boldsymbol{A}^*; \boldsymbol{\theta}_{DR})$$

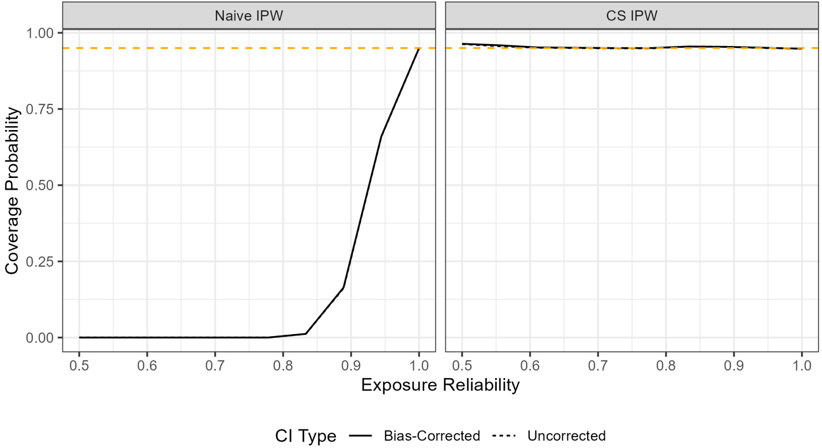
Simulation Setting

- confounder $L \sim \mathcal{N}(0, 0.36)$
- exposure $\boldsymbol{A} = (A_1, A_2)$ with $\boldsymbol{A}|L \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{I})$
- exposure measurement error $\boldsymbol{\epsilon} \sim \mathcal{N}_2(\mathbf{0}, \sigma^2_{me}\boldsymbol{I})$
- outcome Y with $Y|L, \boldsymbol{A} \sim \mathcal{N}(A_1 + A_2 + L, 1)$
- implied MSM of $\eta(\boldsymbol{a}; \boldsymbol{\gamma}) = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_2$ for $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2) = (0, 1, 1)$
- sample size $n = 800$

Simulation Results



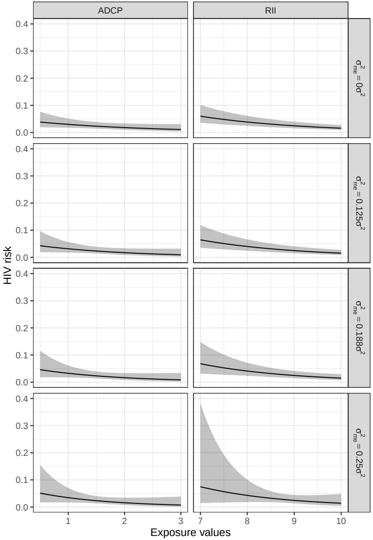
Simulation Results



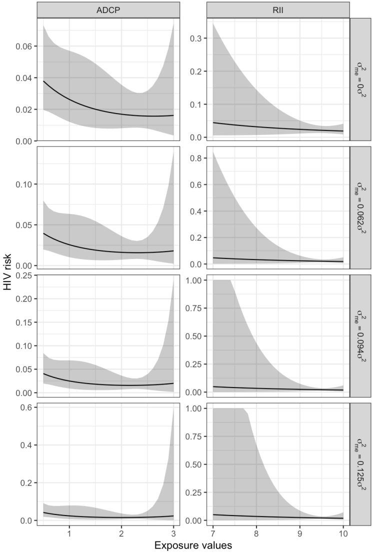
Application: HVTN 505 Trial

- **two exposures**:
  (i) antibody-dependent cellular phagocytosis (ADCP)
  (ii) recruitment of FcγRIIa of the H131-Con S gp140 protein (RII)
- **case-cohort sampling**: immunologic markers only measured in stratified random sample of controls
- **covariates**: age, race, BMI, behavior risk, CD4-P, and CD8-P
- **two analyses**:
  (i) DR estimator with a linear outcome model
  (ii) g-formula with a quadratic outcome model

Application: DR Method with Linear Outcome Model



Application: G-Formula with Quadratic Outcome Model



Mismex: Mismeasured Exposures

Paper on arXiv

GitHub R package

References

Scott M Hammer, Magdalena E Sobieszczyk, Holly Janes, Shelly T Karuna, Mark J Mulligan, Doug Grove, Beryl A Koblin, Susan P Buchbinder, Michael C Keefer, Georgia D Tomaras, et al. Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *New England Journal of Medicine*, 369(22): 2083–2092, 2013.

Holly E Janes, Kristen W Cohen, Nicole Frahm, Stephen C De Rosa, Brittany Sanchez, John Hural, Craig A Magaret, Shelly Karuna, Carter Bentley, Raphael Gottardo, et al. 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, 2017.

Youyi Fong, Xiaoying Shen, Vicki C Ashley, Aaron Deal, Kelly E Seaton, Chenchen Yu, Shannon P Grant, Guido Ferrari, Allan C deCamp, Robert T Bailer, et al. Modification of the association between T-cell immune responses and human immunodeficiency virus type 1 infection risk by vaccine-induced antibody responses in the HVTN 505 trial. *The Journal of Infectious Diseases*, 217(8):1280–1288, 2018.

Scott D Neidich, Youyi Fong, Shuying S Li, Daniel E Geraghty, Brian D Williamson, William Chad Young, Derrick Goodman, Kelly E Seaton, Xiaoying Shen, Sheetal Sawant, et al. Antibody Fc effector functions and IgG3 associate with decreased HIV-1 risk. *Journal of Clinical Investigation*, 129(11):4838–4849, 2019.

Steven J Novick and Leonard A Stefanski. Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association*, 97(458):472–481, June 2002. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214502760047005. URL `http://www.tandfonline.com/doi/abs/10.1198/016214502760047005`.