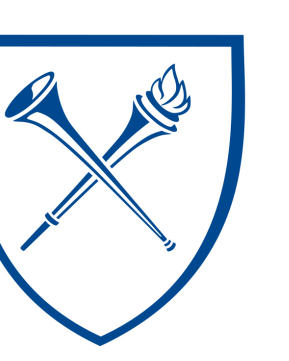




# Robust inference on the causal effects of stochastic interventions in two-phased vaccine efficacy trials

Nima Hejazi, David Benkeser, and Mark van der Laan

Graduate Group in Biostatistics & Department of Statistics, University of California, Berkeley  
Department of Biostatistics and Bioinformatics, Emory University



EMORY  
UNIVERSITY

## OVERVIEW & MOTIVATIONS

- We consider estimating the effect of a stochastic intervention in vaccine trials with two-phase sampling of the treatment.
- We present robust, efficient estimators of the mean counterfactual outcome under a stochastic intervention.
- The proposed estimators are asymptotically normal with a variance estimator based on the efficient influence function.
- Our `txshift` R package [3] implements these estimators and leverages state-of-the-art machine learning in the procedure.
- Our `haldensify` R package [2] estimates a generalized propensity score (a conditional density) via highly adaptive lasso (HAL).

## HIV VACCINE EFFICACY TRIALS

- Question:** How does HIV infection risk vary across shifts of an immunogenic response profile in a trial's vaccine arm.
- Analysis of data from the HIV Vaccine Trials Network 505 efficacy trial, as in [4]:
  - RCT with  $n = 2504$ ; only  $n = 189$  from vaccine arm in second-stage sample.
  - Background covariates ( $W$ ): sex, age, BMI, behavioral HIV risk score.
  - Treatment ( $A$ ): immunogenic response profiles, post-vaccination.
  - Outcome ( $Y$ ): HIV-1 infection status at week 28 of trial.
  - All individuals with HIV-1 infections by week 28 in second-stage sample.
- Goal:** Develop a ranking of immunogenic responses as study endpoints for future HIV-1 vaccine efficacy trials.

## EFFICIENT CORRECTIONS FOR TWO-PHASE SAMPLING

- In the HVTN 505 HIV-1 trial, immunogenic responses  $A$  are sequenced for infected individuals and a matched subset of controls, making  $O = (W, \Delta, \Delta A, Y)$  the observed data structure.
  - $\Delta \in \{0, 1\}$  is the censoring indicator introduced by the two-phase sampling mechanism, under which observed immunogenic response ( $\Delta A$ ) is arbitrarily set to 0 when unobserved.
  - Given  $V := (W, Y)$ , we assume  $\Delta \sim \text{Bern}(\pi_0(V))$  — that is,  $\pi_0(V) = \mathbb{P}(\Delta = 1 \mid V)$ .
- The IPCW-TMLE [5] estimates the target parameter by incorporating inverse probability weights  $\frac{\Delta}{\pi_n(V)}$  in the estimation of the outcome regression and generalized propensity score.
- Efficiency improvements are attainable by constructing estimators from the observed-data EIF:

$$D(P_0)(o) = \frac{\Delta}{\pi(v)} D^F(P_0^X)(X) - \left(1 - \frac{\Delta}{\pi(v)}\right) \mathbb{E}(D^F(P_0^X)(x) \mid \Delta = 1, V = v)$$

- The observed-data EIF, based on full-data EIF  $D^F(P_0^X)(X)$ , allows construction of estimators that
  - achieve the efficiency bound for the class of regular asymptotically linear estimators;
  - are *multiply robust*, with consistency of parameter estimates when one of  $\{g, Q\}$  and one of  $\{\pi_0(V), \mathbb{E}_0(D^F(P_0^X)(X) \mid \Delta = 1, V)\}$  are consistently estimated (e.g., via HAL);
  - allow valid statistical inference even when  $\pi_0$  is estimated nonparametrically.

## THE EFFECT OF A STOCHASTIC TREATMENT REGIME

- Consider  $X = (W, A, Y) \sim P_0^X \in \mathcal{M}$ , where  $W$  is a set of baseline covariates,  $A$  a treatment, and  $Y$  an outcome of interest, with no assumptions placed on the statistical model  $\mathcal{M}$ .
- Consider a shift of the treatment  $d(A, W) = A + \delta$  for a given shift  $\delta$ . To protect against violations of the positivity assumption, make the shifting mechanism a function of the observed data, where  $u(w)$  is the maximum shift supported in the observed data:

$$d(a, w) = \begin{cases} a + \delta, & a + \delta < u(w) \\ a, & \text{otherwise} \end{cases}$$

- The causal parameter has been shown to be identified by a functional of the observed data [1]:

$$\Psi(P_0^X) = \mathbb{E}_{P_0} \bar{Q}(d(A, W), W),$$

where  $\bar{Q}(d(A, W), W)$  is the conditional mean of the outcome given  $A = d(A, W)$  and  $W$ .

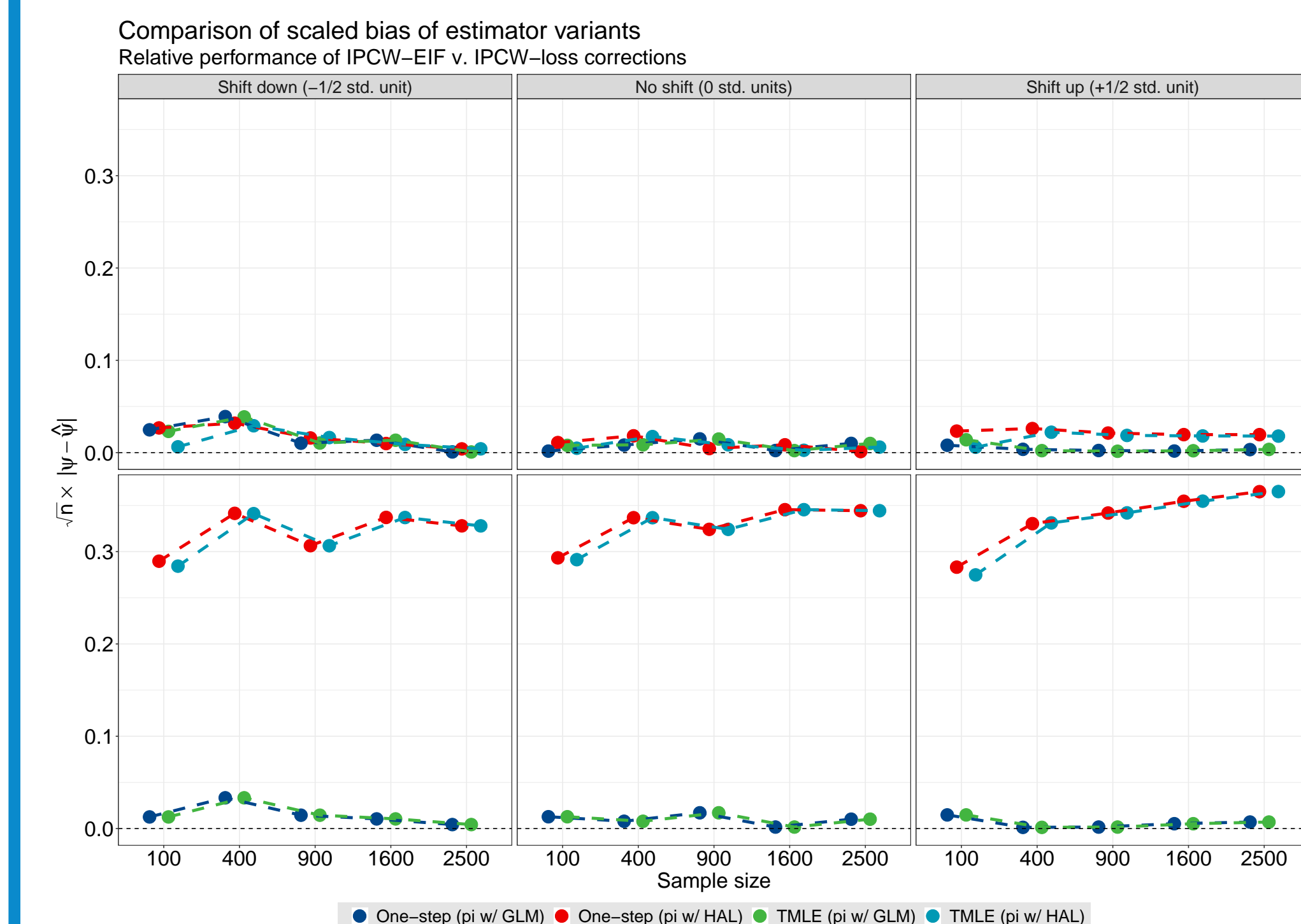
- The efficient influence function (EIF) of  $\Psi$  relative to the nonparametric model  $\mathcal{M}$  is

$$D^F(P_0^X)(x) = H(a, w)(y - \bar{Q}(a, w)) + \bar{Q}(d(a, w), w) - \Psi(P_0^X)(x),$$

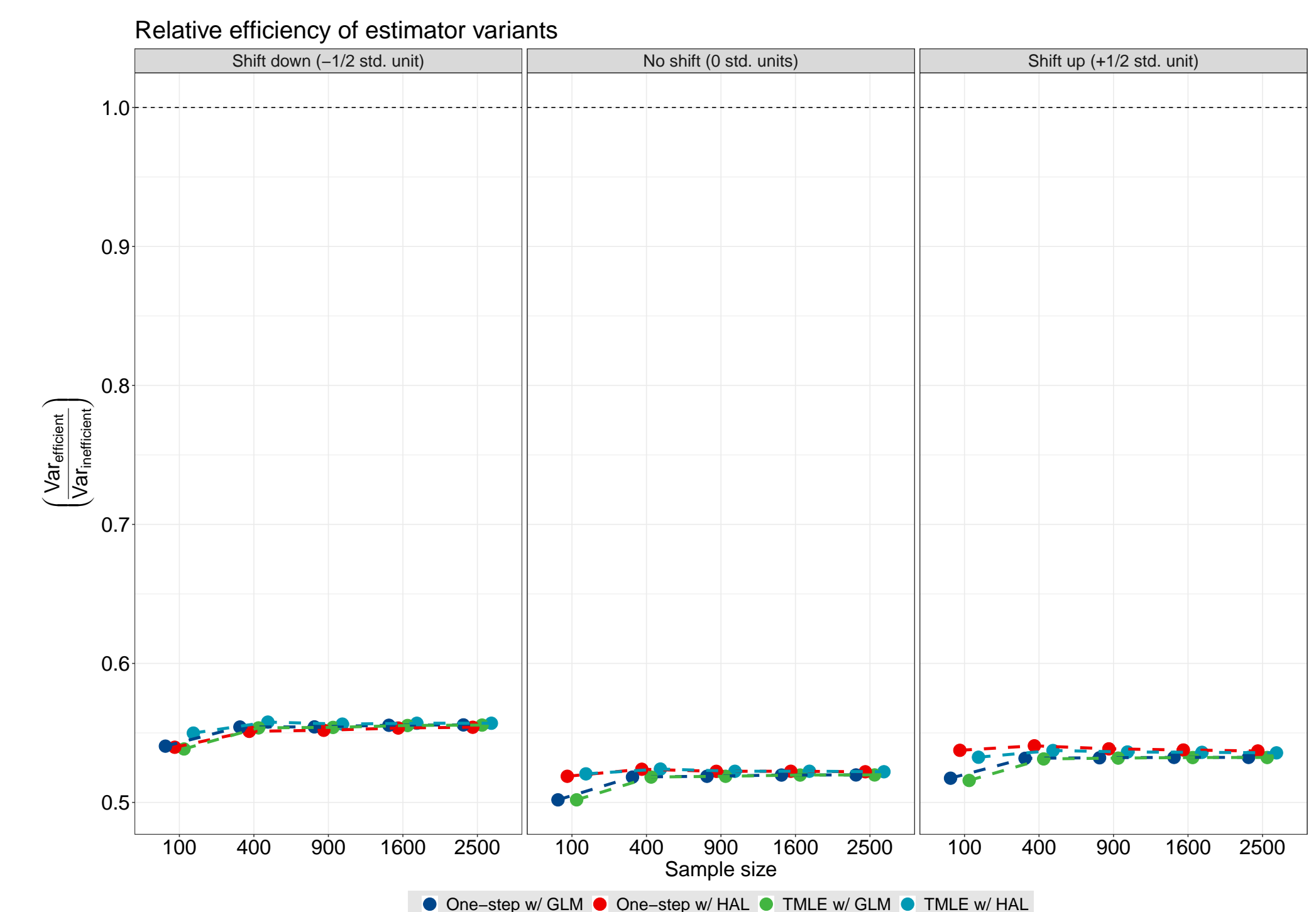
where the auxiliary covariate term  $H(a, w)$  may be expressed as

$$H(a, w) = \mathbb{I}(a < u(w)) \frac{g(a - \delta \mid w)}{g(a \mid w)} + \mathbb{I}(a \geq u(w) - \delta).$$

## NUMERICAL STUDY & RESULTS



(a) Most estimator variants unbiased in large samples; inefficient variants with HAL fail to converge.



(b) Fitting  $\pi$  with HAL or GLM, efficient estimators and simpler variants show comparable relative efficiency.

## REFERENCES

- Iván Díaz and Mark J van der Laan. Population intervention causal effects based on stochastic interventions. *Biometrics*, 68(2):541–549, 2012.
- Nima S Hejazi and David C Benkeser. *haldensify: Nonparametric conditional density estimation with the highly adaptive lasso in R*, 2019. URL <https://github.com/nhejazi/haldensify>. R package version 0.0.3.
- Nima S Hejazi and David C Benkeser. *txshift: Targeted Learning of the Causal Effects of Stochastic Interventions in R*, 2019. URL <https://github.com/nhejazi/txshift>. R package version 0.2.7.
- Holly E Jones, 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.
- Sherri Rose and Mark J van der Laan. A targeted maximum likelihood estimator for two-stage designs. *The International Journal of Biostatistics*, 7(1):1–21, 2011.

## CONTACT INFORMATION

- N. Hejazi**, Graduate Group in Biostatistics, UC Berkeley, [nhejazi@berkeley.edu](mailto:nhejazi@berkeley.edu)
- D. Benkeser**: Asst. Prof. of Biostatistics, Emory University, [benkeser@emory.edu](mailto:benkeser@emory.edu)
- M. van der Laan**, Prof. of Biostatistics, UC Berkeley, [laan@berkeley.edu](mailto:laan@berkeley.edu)