Leveraging the causal effects of stochastic interventions to evaluate vaccine efficacy in two-phase trials

Nima Hejazi

Wednesday, 16 December 2020

Graduate Group in Biostatistics, and
Center for Computational Biology,
University of California, Berkeley

nshejazi
nhejazi
nimahejazi.org

with , D. Benkeser, M. van der Laan, H. Janes, P. Gilbert
SER: "Methods for the thorny challenges of real studies"

The burden of HIV-1

- The HIV-1 epidemic — the facts:
  - now in its fourth decade,
  - 2.5 million new infections occurring annually worldwide,
  - new infections outpace patients starting antiretroviral therapy.
- *Most efficacious* preventive vaccine: 31% reduction rate.
- **Question**: To what extent can HIV-1 vaccines be improved by modulating immunogenic CD4+/CD8+ response profiles?

HVTN 505 trial examined new antibody boost vaccines

- HIV Vaccine Trials Network's (HVTN) 505 vaccine efficacy; randomized controlled trial, $n = 2504$ (Hammer et al. 2013).
- **Question:** How would HIV-1 infection risk in week 28 have changed had immunogenic response (due to vaccine) differed?
- Immunogenic response profiles only available for second-stage sample of $n = 189$ (Janes et al. 2017) due to cost limitations.
- Two-phased sampling mechanism: 100% inclusion rate if HIV-1 positive in week 28; based on matching otherwise.

- Baseline covariates(L): sex, age, BMI, behavioral HIV risk.
- Intervention(s) (A): post-vaccination T-cell activity markers.
- Outcome (Y): HIV-1 infection status at week 28 of tiral.
- 12-color intracellular cytokine staining (ICS) assay.
- Cryopreserved peripheral blood mononuclear cells were stimulated with synthetic HIV-1 peptide pools.
- All immune responses are assayed *after* the endpoints of interest (HIV-1 infection status) are collected.
- **Conclusion:** Understanding which immune responses impact vaccine efficacy helps develop more efficacious vaccines.
- A vaccine effective at preventing HIV-1 acquisition would be a cost-effective and durable approach to halting the worldwide epidemic.

Two-phase sampling censors the complete data structure

- Complete (<u>unobserved</u>) data $X = (L, A, Y) \sim P_0^X \in \mathcal{M}^X$, as per the full HVTN 505 trial cohort (Hammer et al. 2013):
  - L (baseline covariates): sex, age, BMI, behavioral HIV risk,
  - A (exposure): immune response profile for CD4+ and CD8+,
  - Y (outcome of interest): HIV-1 infection status at week 28.
- Observed data $O = (C, CX) = (L, C, CA, Y)$; $C \in \{0, 1\}$ is an indicator for inclusion in the second-stage sample.

- P_0^X — true (unknown) distribution of the full data X,
- $\mathcal{M}_{NP}^X$ — nonparametric statistical model.

NPSEM with static interventions

- Use a nonparametric structural equation model (NPSEM) to describe the generation of X (Pearl 2009), specifically

$$L = f_L(U_L); A = f_A(L, U_A); Y = f_Y(A, L, U_Y)$$

- Implies a model for the distribution of counterfactual random variables generated by interventions on the process.
- A *static intervention* replaces f_A with a specific value a in its conditional support $A \mid L$.
- This requires specifying a particular value of the exposure under which to evaluate the outcome *a priori*.

NPSEM with stochastic interventions

- *Stochastic interventions* modify the value A would naturally assume by drawing from a modified exposure distribution.
- Consider the post-intervention value $A^\star \sim G^\star(\cdot \mid L)$; static interventions are a special case (degenerate distribution).
- Such an intervention generates a counterfactual random variable $Y_{G^\star} := f_Y(A^\star, L, U_Y)$, with distribution P_0^δ, .
- We aim to estimate $\psi_{0,\delta} := \mathbb{E}_{P_0^\delta}\{Y_{G^\star}\}$, the counterfactual mean under the post-intervention exposure distribution $G^\star$.

Stochastic interventions for the causal effects of shifts

- Díaz and van der Laan (2012; 2018)'s *stochastic* interventions

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

- Our estimand is $\psi_{0,d} := \mathbb{E}_{P_0^d}\{Y_{d(A,L)}\}$, mean of $Y_{d(A,L)}$.
- Statistical target parameter is $\Psi(P_0^X) = \mathbb{E}_{P_0^X}\overline{Q}(d(A, L), L)$, counterfactual mean of the *shifted* outcome mechanism.
- For HVTN 505, $\psi_{0,d}$ is the counterfactual risk of HIV-1 infection, had the observed value of the immune response been altered under the rule $d(A, L)$ defining $G^\star(\cdot \mid L)$.

- Causal estimand: counterfactual mean of HIV-1 infection (risk) under a *shifted* immunogenic response distribution.

Flexible, efficient estimation

- The efficient influence function (EIF) is:

$$D(P_0^X)(x) = H(a,l)(y - \overline{Q}(a,l)) + \overline{Q}(d(a,l),l) - \Psi(P_0^X).$$

- The one-step estimator corrects bias by adding the empirical mean of the estimated EIF to the substitution estimator:

$$\Psi_n^+ = \frac{1}{n}\sum_{i=1}^{n} \overline{Q}_n(d(A_i,L_i),L_i) + D_n(O_i).$$

- The TML estimator is built by updating initial estimates of $\overline{Q}_n$ via a (logistic) tilting model, yielding

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

- Both estimators are CAN even when nuisance parameters are estimated via flexible, machine learning techniques.

- Semiparametric-efficient estimation thru solving efficient influence function estimating equation wrt the model $\mathcal{M}$.
- The auxiliary covariate simplifies when the treatment is in the limits (conditional on W) — i.e., for $A_i \in (u(l) - \delta, u(l))$, then we have $H(a, l) = \frac{g_0(a-\delta|l)}{g_0(a|l)} + 1$.
- Need to explicitly remind the audience what $u(l)$ is again. It's only appeared once at this point, and only been mentioned in passing.

Augmented estimators for two-phase sampling designs

- Rose and van der Laan (2011) introduce the IPCW-TMLE, to be used when observed data is subject to two-phase sampling.
- *Initial proposal:* correct for two-phase sampling by using a loss function with inverse probability of censoring weights:

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

- When the sampling mechanism $\pi_0(Y, L)$ can be estimated by a parametric form, this procedure yields an efficient estimator.
- However, when machine learning is used (e.g., when $\pi_0(Y, L)$ is not *known by design*), this is insufficient.

Efficient estimation and multiple robustness

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

$$D(P_0^X)(o) = \frac{c}{\pi_0(y, l)} D^F(P_0^X)(x) - \left(1 - \frac{c}{\pi_0(y, l)}\right) \cdot \mathbb{E}(D^F(P_0^X)(x) \mid C = 1, Y = y, L = l),$$

  - Expresses observed data EIF $D^F(P_0^X)(o)$ in terms of full data EIF $D^F(P_0^X)(x)$; inclusion of second term ensures efficiency.
  - The expectation of the full data EIF $D^F(P_0^X)(x)$, taken only over units selected by the sampling mechanism (i.e., $C = 1$).

- A unique multiple robustness property — combinations of $(g_0(L), \overline{Q}_0(A, L)) \times (\pi_0(Y, L), \mathbb{E}(D^F(P_0^X)(x) \mid C = 1, Y, L))$.

Helping to fight the HIV-1 epidemic

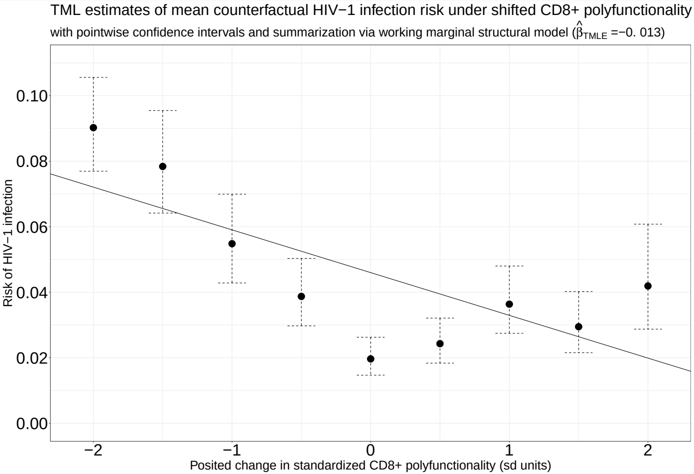

Figure 1: Analysis of HIV-1 risk as a function of CD8+ immunogenicity, using R package `txshift` (`https://github.com/nhejazi/txshift.`)

Big picture takeaways

- Vaccine efficacy evaluation helps to develop enhanced vaccines better informed by biological properties of the target disease.
- HIV-1 vaccines modulate immunogenic response profiles as part of their mechanism for lowering HIV-1 infection risk.
- *Stochastic* interventions constitute a flexible framework for considering **realistic** treatment/intervention policies.
- Large-scale (vaccine) trials often use two-phase designs — need to (carefully!) accommodate for sampling complications.
- We've developed robust, open source statistical software for assessing stochastic interventions in observational studies.

Thank you!

- https://nimahejazi.org
- https://twitter.com/nshejazi
- https://github.com/nhejazi
- https://doi.org/10.1111/biom.13375

Appendix

From the causal to the statistical target parameter

Assumption 1: *Consistency*

$Y_i^{d(a_i,l_i)} = Y_i$ *in the event* $A_i = d(a_i, l_i)$, *for* $i = 1, \ldots, n$

Assumption 2: *SUTVA*

$Y_i^{d(a_i,l_i)}$ *does not depend on* $d(a_j, l_j)$ *for* $i = 1, \ldots, n$ *and* $j \neq i$, *or lack of interference (Rubin 1978; 1980)*

Assumption 3: *Strong ignorability*

$A_i \perp\!\!\!\perp Y_i^{d(a_i,l_i)} \mid L_i$, *for* $i = 1, \ldots, n$

From the causal to the statistical target parameter

Assumption 4: *Positivity (or overlap)*

$a_i \in \mathcal{A} \implies d(a_i, l_i) \in \mathcal{A}$ *for all* $l \in \mathcal{L}$, *where* $\mathcal{A}$ *denotes the support of* A *conditional on* $L = l_i$ *for all* $i = 1, \dots n$

- This positivity assumption is not quite the same as that required for categorical interventions.
- In particular, we do not require that the intervention density place mass across all strata defined by L.
- Rather, we merely require the post-intervention quantity be seen in the observed data for given $a_i \in \mathcal{A}$ and $l_i \in \mathcal{L}$.

Literature: Díaz and van der Laan (2012)

- *Proposal:* Evaluate outcome under an altered *intervention distribution* — e.g., $P_\delta(g_0)(A = a \mid L) = g_0(a - \delta(L) \mid L)$.
- Identification conditions for a statistical parameter of the counterfactual outcome $\psi_{0,d}$ under such an intervention.
- Show that the causal quantity of interest $\mathbb{E}_0\{Y_{d(A,L)}\}$ is identified by a functional of the distribution of X:

$$\psi_{0,d} = \int_{\mathcal{L}} \int_{\mathcal{A}} \mathbb{E}_{P_0^X}\{Y \mid A = d(a,l), L = l\} \cdot \\ q_{0,A}^X(a \mid L = l) \cdot q_{0,L}^X(l) d\mu(a) d\nu(l)$$

- Provides a derivation based on the efficient influence function (EIF) with respect to the nonparametric model $\mathcal{M}$.

- The identification result allows us to write down the causal quantity of interest in terms of a functional of the observed data.
- Key innovation: loosening standard assumptions through a change in the observed intervention mechanism.
- Problem: globally altering an intervention mechanism does not necessarily respect individual characteristics.
- The authors build IPW, A-IPW, and TML estimators, comparing the three different approaches.
- IMPORTANT: gives the g-computation formula for identification of this estimator from the observed data structure.

Literature: Haneuse and Rotnitzky (2013)

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

$$g_{0,\delta}(a \mid l) = \sum_{j=1}^{J(l)} I_{\delta,j}\{h_j(a,l), l\} g_0\{h_j(a,l) \mid l\} h_j'(a,l)$$

- Such intervention policies account for the natural value of the intervention A directly yet are interpretable as the imposition of an altered intervention mechanism.
- Identification conditions for assessing the parameter of interest under such interventions appear technically complex (at first).

- Shifts of the form $d(A, L)$ are considerably more interesting since these are realistic intervention policies.
- Example: consider an individual with an extremely high immune response but whose baseline covariates L suggest we shift the response still higher. Such a shift may not be biologically plausible (impossible, even) but we cannot account for this if the shift is only a function of L.
- The authors build IPW, outcome regression, and non-iterative doubly robust estimators, as well as an approach based on MSMs.
- Piecewise smooth invertibility: This assumption ensures that we can use the change of variable formula when computing integrals over A and it is useful to study the estimators that we propose in this paper.

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 $\mathbb{E}Y_{d(A,L)}$ or $\mathbb{E}Y_{d(L)}$.
- The authors also consider limits on implementing shifts $d(A, L)$, and address working in a longitudinal setting.

Literature: Díaz and van der Laan (2018)

- Builds on the original proposal, accomodating MTP-type shifts $d(A, L)$ proposed after their earlier work.
- To protect against positivity violations, considers a specific shifting mechanism:

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

- Proposes an improved "1-TMLE" algorithm, with a single auxiliary covariate for constructing the TML estimator.
- Our (first) contribution: implementation of this algorithm.

Nonparametric conditional density estimation

- To compute the auxiliary covariate $H(a, l)$, we need to estimate conditional densities $g(A \mid L)$ and $g(A - \delta \mid L)$.
- There is a rich literature on density estimation, we follow the approach proposed in Díaz and van der Laan (2011).
- To build a conditional density estimator, consider

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

  for $\alpha_{t-1} \leq a < \alpha_t$.
  - This is a classification problem, where we estimate the probability that a value of A falls in a bin $[\alpha_{t-1}, \alpha_t)$.
  - The choice of the tuning parameter t corresponds roughly to the choice of bandwidth in classical kernel density estimation.

Nonparametric conditional density estimation

- Díaz and van der Laan (2011) propose a re-formulation of this classification approach as a set of hazard regressions.
- To effectively employ this proposed re-formulation, consider

$$\begin{aligned}\mathbb{P}(A \in [\alpha_{t-1}, \alpha_t) \mid L) =& \mathbb{P}(A \in [\alpha_{t-1}, \alpha_t) \mid A \geq \alpha_{t-1}, L) \times \\ & \prod_{j=1}^{t-1} \{1 - \mathbb{P}(A \in [\alpha_{j-1}, \alpha_j) \mid A \geq \alpha_{j-1}, L)\}\end{aligned}$$

  - The likelihood of this model may be expressed to correspond to the likelihood of a binary variable in a data set expressed via a long-form repeated measures structure.
  - Specifically, the observation of X_i is repeated as many times as intervals $[\alpha_{t-1}, \alpha_t)$ are before the interval to which A_i belongs, and the binary variables indicating $A_i \in [\alpha_{t-1}, \alpha_t)$ are recorded.

Density estimation with the Super Learner algorithm

- To estimate $g(A \mid L)$ and $g(A - \delta \mid L)$, use a pooled hazard regression, spanning the support of A.
- We rely on the Super Learner algorithm of van der Laan et al. (2007) to build an ensemble learner that optimally weights each of the proposed regressions, using cross-validation (CV).
- The Super Learner algorithm uses V-fold CV to train each proposed regression model, weighting each by the inverse of its average risk across all V holdout sets.
- By using a library of regression estimators, we invoke the result of van der Laan et al. (2004), who prove this likelihood-based cross-validated estimator to be asymptotically optimal.

- The auxiliary covariate simplifies when the treatment is in the limits (conditional on L) — i.e., for $A_i \in (u(l) - \delta, u(l))$, then we have $H(a, l) = \frac{g_0(a-\delta|l)}{g_0(a|l)} + 1$.
- Asymptotically optimal in the sense that it performs as well as the oracle selector as the sample size increases.

Key properties of TML estimators

- **Asymptotic linearity:**

$$\Psi(P_n^\star) - \Psi(P_0^X) = \frac{1}{n}\sum_{i=1}^{n} D(P_0^X)(X_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

- **Gaussian limiting distribution:**

$$\sqrt{n}(\Psi(P_n^\star) - \Psi(P_0^X)) \to N(0, Var(D(P_0^X)(X)))$$

- **Statistical inference:**

$$\text{Wald-type confidence interval} : \Psi(P_n^\star) \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_n}{\sqrt{n}},$$

where σ_n^2 is computed directly via $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} D^2(\cdot)(X_i)$.

Under the additional condition that the remainder term $R(\hat{P}^\star, P_0)$ decays as $o_P\left(\frac{1}{\sqrt{n}}\right)$, we have that $\Psi_n - \Psi_0 = (P_n - P_0) \cdot D(P_0) + o_P\left(\frac{1}{\sqrt{n}}\right)$, which, by a central limit theorem, establishes a Gaussian limiting distribution for the estimator, with variance $V(D(P_0))$, the variance of the efficient influence function when Ψ admits an asymptotically linear representation.

The above implies that Ψ_n is a $\sqrt{n}$-consistent estimator of Ψ, that it is asymptotically normal (as given above), and that it is locally efficient. This allows us to build Wald-type confidence intervals, where σ_n^2 is an estimator of $V(D(P_0))$. The estimator σ_n^2 may be obtained using the bootstrap or computed directly via $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n D^2(\bar{Q}_n^\star, g_n)(O_i)$

We obtain semiparametric-efficient estimation and robust inference in the nonparametric model $\mathcal{M}$ by solving the efficient influence function.

1. If $D(\bar{Q}_n^*, g_n)$ converges to $D(P_0)$ in $L_2(P_0)$ norm.
2. The size of the class of functions $\bar{Q}_n^\star$ and g_n is bounded (technically, $\exists \mathcal{F}$ st $D(\bar{Q}_n^*, g_n) \in \mathcal{F}$ whp, where $\mathcal{F}$ is a Donsker class)

Algorithm for TML estimation

1. Construct initial estimators g_n of $g_0(A, L)$ and Q_n of $\overline{Q}_0(A, L)$, perhaps using data-adaptive regression techniques.
2. For each observation i, compute an estimate $H_n(a_i, l_i)$ of the auxiliary covariate $H(a_i, l_i)$.
3. Estimate the parameter ϵ in the logistic regression model
$$\text{logit}\overline{Q}_{\epsilon,n}(a, l) = \text{logit}\overline{Q}_n(a, l) + \epsilon H_n(a, l),$$
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 $\overline{Q}_{n,\epsilon_n}$:
$$\Psi_n = \Psi(P_n^\star) = \frac{1}{n}\sum_{i=1}^{n} \overline{Q}_n^\star(d(A_i, L_i), L_i).$$

- We recommend using nonparametric methods for the initial estimators, as consistent estimation is necessary for efficiency of the estimator Ψ_n.
- Intuition for the submodel fluctuation?

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 second-stage sample $C = 1$, construct initial estimators $g_n(A, L)$ and $\overline{Q}_n(A, L)$, weighting by the sampling mechanism estimate $\pi_n(Y, L)$.

3. With the approach described for the full data case, compute $H_n(a_i, l_i)$, 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 TML estimate in each iteration, until $\frac{1}{n} \sum_{i=1}^{n} \text{EIF}_i < \frac{1}{n}$.

- We recommend using nonparametric methods for the initial estimators, as consistent estimation is necessary for efficiency of the estimator Ψ_n.
- Intuition for the submodel fluctuation?
- This process includes the use of HAL to fit the regression of the EIF contributions on the sampling node $\{Y, L\}$.

Identifying the best efficient estimator

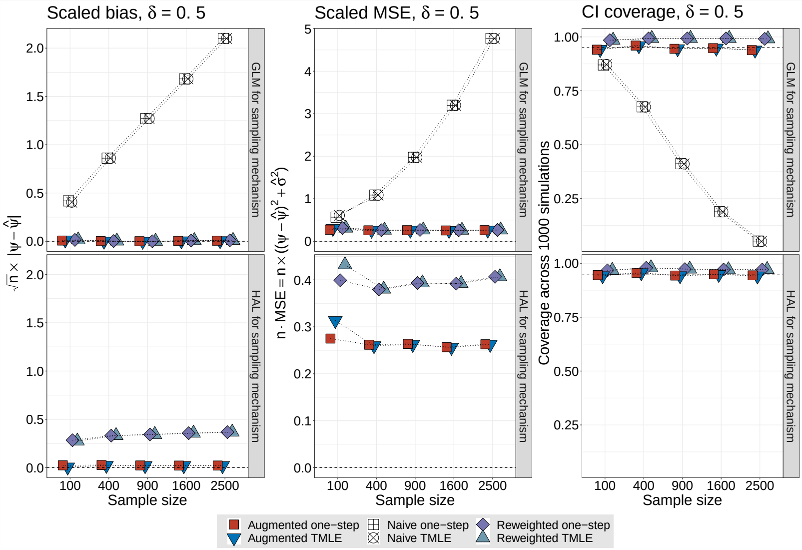

Figure 2: Relative performance of reweighted and augmented estimators.

A linear modeling perspective

- Briefly consider a simple data structure: $X = (Y, A)$; we seek to model the outcome Y as a function of A.
- To posit a linear model, consider $Y_i = \beta_0 + \beta_1 A_i + \epsilon_i$, with error $\epsilon_i \sim N(0, 1)$.
- Letting δ be a change in A, $Y_{A+\delta} - Y_A$ may be expressed

$$\begin{aligned} \mathbb{E}Y_{A+\delta} - \mathbb{E}Y_A &= [\beta_0 + \beta_1(\mathbb{E}A + \delta)] - [\beta_0 + \beta_1(\mathbb{E}A)] \\ &= \beta_0 - \beta_0 + \beta_1\mathbb{E}A - \beta_1\mathbb{E}A + \beta_1\delta \\ &= \beta_1\delta \end{aligned}$$

- Thus, a *unit shift* in A (i.e., $\delta = 1$) may be seen as inducing a change in the difference in outcomes of magnitude β_1.

- We extend this result to the mean counterfactual outcomes under the nonparametric model $\mathcal{M}$.
- Linear modeling analogy re: conversation with Alan on 22 August.

A causal inference perspective

- Consider a data structure: $(Y_a, a \in \mathcal{A})$.
- To posit a linear model, let $Y_a = \beta_0 + \beta_1 a + \epsilon_a$ for $a \in \mathcal{A}$, with error $\epsilon_a \sim N(0, \sigma_a^2)\ \forall a \in A$.
- For the counterfactual outcomes $(Y_{a'+\delta}, Y_{a'})$, their difference, $Y_{a'+\delta} - Y_{a'}$, for some $a' \in \mathcal{A}$, may be expressed

$$\begin{aligned}\mathbb{E}Y_{a'+\delta} - \mathbb{E}Y_{a'} &= [\beta_0 + \beta_1(a' + \delta) + \mathbb{E}\epsilon_{a'+\delta}] - [\beta_0 + \beta_1 a' + \mathbb{E}\epsilon_{a'}] \\ &= \beta_1 \delta\end{aligned}$$

- Thus, a *unit shift* for $a' \in A$ (i.e., $\delta = 1$) may be seen as inducing a change in the difference in the counterfactual outcomes of magnitude β_1.

- Note that this analysis is exactly what we're told we **cannot** do in linear models 101 — that is, the slope of a regression line cannot be interpreted as *causing* a change in the outcome.
- We extend this result to the mean counterfactual outcomes under the nonparametric model $\mathcal{M}$.
- Linear modeling analogy re: conversation with Alan on 22 August.
- Example updated to incorporate countercatuals re: conversation with David on 30 August

Slope in a semiparametric model

- Consider the stochastic intervention $g^{\star}(\cdot \mid L)$:

$$\mathbb{E}Y_{g^{\star}} = \int_L \int_a \mathbb{E}(Y \mid A = a, L) g(a - \delta \mid L) \cdot da \cdot dP_0(L)$$
$$= \int_L \int_z \mathbb{E}(Y \mid A = z + \delta, L) g(z \mid L) \cdot dz \cdot dP_0(L),$$

defning the change of variable $z = a - \delta$.

- For a semiparametric model, $\mathbb{E}(Y \mid A = z, L) = \beta z + \theta(L)$:

$$\mathbb{E}Y_{g^{\star}} - \mathbb{E}Y = \int_L \int_z [\mathbb{E}(Y \mid A = z + \delta, L) - \mathbb{E}(Y \mid A = z, L)] \\ g(z \mid L) \cdot dz \cdot dP_0(L)$$
$$= [\beta(z + \delta) + \theta(L)] - [\beta z + \theta(L)]$$
$$= \beta\delta$$

References

Breiman, L. (1996). Stacked regressions. *Machine Learning*, 24(1):49–64.

Díaz, I. and van der Laan, M. J. (2011). Super learner based conditional density estimation with application to marginal structural models. *The 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.

Dudoit, S. and van der Laan, M. J. (2005). Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology*, 2(2):131–154.

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.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

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.

Kennedy, E. H. (2018). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association*, (just-accepted).

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1229–1245.

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

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.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, 75(371):591–593.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331.

van der Laan, M. J., Dudoit, S., and Keles, S. (2004). Asymptotic optimality of likelihood-based cross-validation. *Statistical Applications in Genetics and Molecular Biology*, 3(1):1–23.

van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super Learner. *Statistical Applications in Genetics and Molecular Biology*, 6(1).

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1).

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, 5(2):241–259.

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.