Nonparametric sensitivity analysis for the survivor average causal effect: Towards sharp bounds for vaccine efficacy

Nima Hejazi

Tuesday, 24 May 2022

Division of Biostatistics
Department of Population Health Sciences
Weill Cornell Medicine

nshejazi
nhejazi
nimahejazi.org
with I. Díaz, P. Gilbert, & D. Benkeser
American Causal Inference Conference

The fights against HIV-1 and COVID-19

- The HIV-1 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 an antibody boost vaccine (Hammer et al. 2013).
- The COVID-19 pandemic:
  - ~~270M~~ ~~331M~~ 523M total cases estimated globally;
  - new variants emerging, with vaccine uptake plateauing;
  - COVID-19 Vaccine Prevention Network's (CoVPN) COVE trial evaluated the Moderna mRNA vaccine (Baden et al. 2021).

Key questions in evaluating vaccine efficacy

- How would [HIV-1, COVID-19] infection risk have differed had immunogenic responses been modulated higher or lower? (See Hejazi et al. 2020, Gilbert et al. 2021a;b)
- How can changes in vaccine-induced immunogenic response inform future [HIV-1, COVID-19] vaccine development?
- What is the effect on [HIV-1, COVID-19] viral load *in the infected* of a vaccine administered in an RCT?

Measuring viral load in vaccine efficacy trials

- Recruit participants, measuring baseline covariates (e.g., sex-at-birth, age), and randomize to vaccine vs. placebo.
- By trial's end, some participants infected with the disease in question, others not. Measure viral load in the infected.
- What is the vaccine's effect on infection? Randomization!
- What is the vaccine's effect on viral load? Randomization?

Formalizing the problem

- We recruit n units, $O_1, \ldots, O_n$, where $O = (W, A, Y, V)$:
  - Baseline covariates (W), e.g., sex-at-birth, race/ethnicity.
  - Treatment ($A \in \{0, 1\}$): placebo or vaccine (randomized).
  - Preliminary outcome ($Y \in \{0, 1\}$): infection by trial's end.
  - Outcome ($V \in \mathbb{R}$): viral load measured by quantitative assay.
- Structural causal model (with standard assumptions):

$$W = f_W(U_W); A = f_A(W, U_A);$$
$$Y = f_Y(W, A, U_Y); V = f_V(W, A, Y, U_V)$$

Evaluating the vaccine's effect: A first pass

- Randomization lets us learn $\mathbb{E}[Y(1) - Y(0)]$, using W for precision enhancement via covariate adjustment possibly.
- V is only meaningfully quantifiable in the infected, or effect on the infected is of interest (e.g., Gilbert et al. 2003).
- Y is a post-randomization event: learning $\mathbb{E}[V(1) - V(0)]$ could be impossible due to confounding by infection status.
- Can we exploit any shortcuts to quantifying this effect?

Using the average treatment effect

- Assume $V = 0$ when $Y = 0$: we can quantify $\mathbb{E}[V(1) - V(0)]$, but we still cannot rely on randomization.
- How? Use inverse weighting or doubly robust approaches.
- The catch: Analysis is *under-powered by assumption*. Setting $V = 0$ artificially biases ATE estimates towards the null.
- A small step forward, but we substantially risk missing effect of vaccination on viral load, even when large (rare infection).

Without randomization, what can we learn?

- Without assuming $V = 0$ when $Y = 0$, we only have V quantifiable in the group for whom $Y = 1$, the infected.
- Infection is a post-randomization event, so this group could be highly unbalanced on W — so we can't learn $\mathbb{E}[V(1) - V(0)]$.
- Perhaps the infected are meaningfully different from those that remain uninfected (e.g., at higher baseline risk)?
- How do we balance on a post-randomization event?

Principal stratification

- The potential outcomes of Y define useful subgroups:
  - $\{Y(1) = 1, Y(0) = 1\}$, the "always infected"
  - $\{Y(1) = 0, Y(0) = 0\}$, the "never infected"
  - $\{Y(1) = 0, Y(0) = 1\}$, the vaccine-protected
  - $\{Y(1) = 1, Y(0) = 0\}$, the vaccine-harmed
- Formalized as *principal strata* (Robins 1995, Balke and Pearl 1997, Rubin 1998, Frangakis and Rubin 2002).
- For convenience, let $\theta(y_1, y_0) \coloneqq \mathbb{P}(Y(1) = y_1, Y(0) = y_0)$.

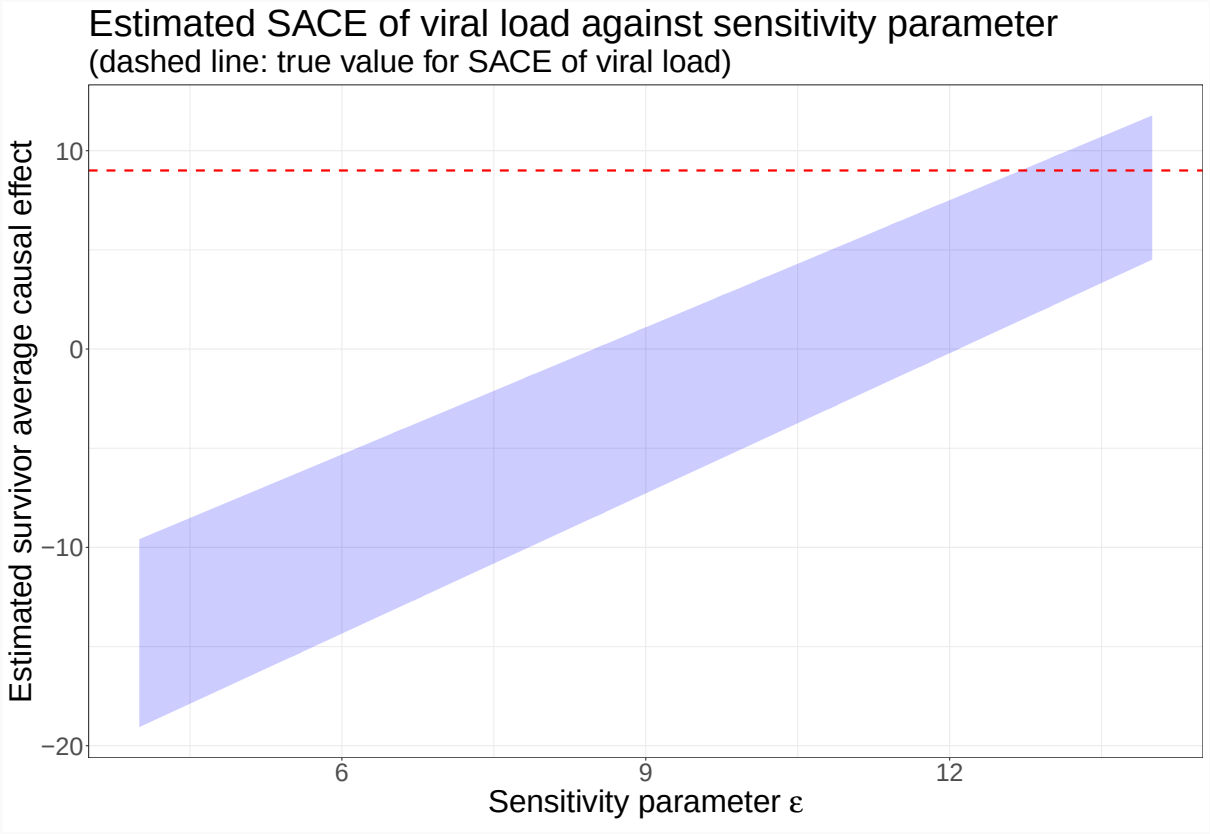
The survivor average causal effect (SACE)

- Harmless treatment: $\theta(1, 0) = \mathbb{P}(\{Y(1) = 1, Y(0) = 0\}) = 0$, common in principal stratification (e.g., Angrist et al. 1996).
- $\lambda_{\text{SACE}} \coloneqq \mathbb{E}[V(1) - V(0) \mid Y(1) = Y(0) = 1]$ which quantifies the vaccine's effect on viral load in the always-infected.
- $\psi = \sum_{y_1, y_0} \mathbb{E}[V(1) - V(0) \mid Y(1) = y_1, Y(0) = y_0]\theta(y_1, y_0)$, where this is simply the total effect $\psi = \mathbb{E}[V(1) - V(0)]$.
- Re-expressed in terms of the SACE, this total effect takes the form $\psi = \lambda_{\text{SACE}}\theta(1, 1) - \mathbb{E}[V(0) \mid Y(1) = 0, Y(0) = 1]\theta(0, 1)$.

Nonparametric partial identification of the SACE

- Now we can write, $\lambda_{\text{SACE}} = [\psi + \epsilon_0\theta(0,1)]/\theta(1,1)$ in terms of ψ and where $\epsilon_0 := \mathbb{E}[V(0) \mid Y(1) = 0, Y(0) = 1]$.
- $\{\theta(0,1), \theta(1,1), \epsilon_0\}$ cannot be point-identified — what now?
- But $\{\theta(0,1), \theta(1,1)\}$ can be set-identified, with sharp bounds via Huang et al. (2017; 2019) or Gabriel et al. (2020; 2022).
- ϵ_0 can be used as a sensitivity parameter.
  - ϵ_0: Viral load of the vaccine–protected under placebo.
  - $\mathbb{E}[V(0) \mid Y(1) = Y(0) = 1] - \mathbb{E}[V(0) \mid Y(0) = 1, Y(1) = 0]$ is the average benefit, comparing viral load in infected placebo recipients causally unprotected vs. protected by vaccination.

Illustrating the approach



The big picture and future work

1. Quantifying effects on viral load is critical to understanding better the efficacy profiles of novel vaccines.
2. The SACE aligns well with scientific aims of vaccine trials; we provide a nonparametric sensitivity-based characterization.
3. Understanding effects on viral load *thru immunogenic markers* is key: decompose the SACE into direct and indirect effects.
4. Remaining questions and ongoing work:
   - Compare existing (sharp) bounds, contributed independently by Huang et al. (2017; 2019) and Gabriel et al. (2020; 2022).
   - Incorporate prior nonparametric sensitivity analysis efforts (e.g., Díaz and van der Laan 2013, Ding and VanderWeele 2016).

References

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

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.

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176.

Díaz, I. and van der Laan, M. J. (2013). Sensitivity analysis for causal inference under unmeasured confounding and measurement error problems. *International Journal of Biostatistics*, 9(2):149–160.

Ding, P. and VanderWeele, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology*, 27(3):368.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Gabriel, E. E., Sachs, M. C., and Sjölander, A. (2020). Causal bounds for outcome-dependent sampling in observational studies. *Journal of the American Statistical Association*, pages 1–12.

Gabriel, E. E., Sachs, M. C., and Sjölander, A. (2022). Sharp nonparametric bounds for decomposition effects with two binary mediators. *Journal of the American Statistical Association*, (just-accepted):1–21.

Gilbert, P. B., Bosch, R. J., and Hudgens, M. G. (2003). Sensitivity analysis for the assessment of causal vaccine effects on viral load in HIV vaccine trials. *Biometrics*, 59(3):531–541.

Gilbert, P. B., Fong, Y., Benkeser, D., Andriesen, J., Borate, B., Carone, M., Carpp, L. N., Díaz, I., Fay, M. P., Fiore-Gartland, A., Hejazi, N. S., Huang, Y., Huang, Y., Hyrien, O., Janes, H. E., Juraska, M., Li, K., Luedtke, A., Nason, M., Randhawa, A. K., van der Laan, L., Williamson, B. D., Zhang, W., and Follmann, D. (2021a). CoVPN COVID-19 Vaccine Efficacy Trial Immune Correlates Statistical Analysis Plan.

Gilbert, P. B., Montefiori, D. C., McDermott, A. B., Fong, Y., Benkeser, D., Deng, W., Zhou, H., Houchens, C. R., Martins, K., Jayashankar, L., Castellino, F., Flach, B., Lin, B. C., O'Connell, S., McDanal, C., Eaton, A., Sarzotti-Kelsoe, M., Lu, Y., Yu, C., Borate, B., van der Laan, L. W., Hejazi, N. S., Huynh, C., Miller, J., El Sahly, H. M., Baden, L. R., Baron, M., De La Cruz, L., Gay, C., Kalams, S., Kelley, C. F., Kutner, M., Andrasik, M. P., Kublin, J. G., Corey, L., Neuzil, K. M., Carpp, L. N., Pajon, R., Follmann, D., Donis, R. O., Koup, R. A., and on behalf of the Immune Assays; Moderna, Inc.; Coronavirus Vaccine Prevention Network (CoVPN) / Coronavirus Efficacy (COVE); and United States Government (USG) / CoVPN Biostatistics Teams (2021b). Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science*, 375(6576):43–50.

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.

Hejazi, N. S., van der Laan, M. J., Janes, H. E., Gilbert, P. B., and Benkeser, D. C. (2020). Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*, 77(4):1241–1253.

Huang, E. J., Fang, E. X., Hanley, D. F., and Rosenblum, M. (2017). Inequality in treatment benefits: Can we determine if a new treatment benefits the many or the few? *Biostatistics*, 18(2):308–324.

Huang, E. J., Fang, E. X., Hanley, D. F., and Rosenblum, M. (2019). Constructing a confidence interval for the fraction who benefit from treatment, using randomized trial data. *Biometrics*, 75(4):1228–1239.

Robins, J. M. (1995). An analytic method for randomized trials with informative censoring: part 1. *Lifetime Data Analysis*, 1(3):241–254.

Rubin, D. B. (1998). More powerful randomization-based *p*-values in double-blind trials with non-compliance. *Statistics in Medicine*, 17(3):371–385.

Thank you! Questions?

`https://nimahejazi.org`

`https://twitter.com/nshejazi`

`https://github.com/nhejazi`

Work in progress — stay tuned!