# Treatment Effect Risk: Bounds and Inference

Nathan Kallus
kallus@cornell.edu
Netflix and Cornell University
New York, New York, USA

## ABSTRACT

Since the average treatment effect (ATE) measures the change in social welfare, even if positive, there is a risk of negative effect on, say, some 10% of the population. Assessing such risk is difficult, however, because any one individual treatment effect (ITE) is never observed so the 10% worst-affected cannot be identified, while distributional treatment effects only compare the first deciles within each treatment group, which does not correspond to any 10%-subpopulation. In this paper we consider how to nonetheless assess this important risk measure, formalized as the conditional value at risk (CVaR) of the ITE-distribution. We leverage the availability of pre-treatment covariates and characterize the tightest-possible upper and lower bounds on ITE-CVaR given by the covariate-conditional average treatment effect (CATE) function. We then proceed to study how to estimate these bounds efficiently from data and construct confidence intervals. This is challenging even in randomized experiments as it requires understanding the distribution of the unknown CATE function, which can be very complex if we use rich covariates so as to best control for heterogeneity. We develop a debiasing method that overcomes this and prove it enjoys favorable statistical properties even when CATE and other nuisances are estimated by black-box machine learning or even inconsistently. Studying a hypothetical change to French job-search counseling services, our bounds and inference demonstrate a small social benefit entails a negative impact on a substantial subpopulation.

## KEYWORDS

Program evaluation, Individual treatment effect, Conditional average treatment effect, Conditional value at risk, Partial identification, Debiased machine learning

**ACM Reference Format:**
Nathan Kallus. 2022. Treatment Effect Risk: Bounds and Inference. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3531146.3533087

This research was partially supported by NSF and Netflix.

*FAccT '22, June 21–24, 2022, Seoul, Republic of Korea*
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9352-2/22/06.
https://doi.org/10.1145/3531146.3533087