The Journal of Open Source Software

medoutcon: Nonparametric efficient causal mediation analysis with machine learning in R

Nima S. Hejazi[1], **Kara E. Rudolph**[2], **and Iván Díaz**[1]

1 Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine **2** Department of Epidemiology, Mailman School of Public Health, Columbia University

DOI: 10.21105/joss.03979

Software

- Review
- Repository
- Archive

Editor: Mikkel Meyer Andersen

Reviewers:

- @erikcs
- @rrrlw

Submitted: 10 November 2021
Published: 10 December 2021

Summary

Science is most often concerned with questions of *mechanism*. In myriad applications, only the portion of the causal effect of an exposure on an outcome through a particular pathway under study is of interest. The study of such path-specific, or mediation, effects has a rich history, first undertaken scientifically by Wright (1921) and Wright (1934). Today, the study of such effects has attracted a great deal of attention in statistics and causal inference, inspired by applications in disciplines ranging from epidemiology and vaccinology to psychology and economics. Examples include understanding the biological mechanisms by which vaccines causally alter infection risk (Benkeser et al., 2021; Hejazi et al., 2020), assessing the effect of novel pharmacological therapies on substance abuse disorder relapse (Hejazi et al., 2021; Rudolph et al., 2020), and evaluating the effects of housing vouchers on adolescent development (Rudolph et al., 2021). The `medoutcon` `R` package provides researchers in each of these disciplines, and in others, with the tools necessary to implement statistically efficient estimators of the *interventional* direct and indirect effects (Díaz et al., 2020) (for brevity, henceforth, (in)direct effects), a recently formulated set of causal effects robust to the presence of confounding of the mediator-outcome relationship by the exposure. In cases where such confounding is a nonissue, the interventional (in)direct effects (VanderWeele et al., 2014) reduce to the well-studied *natural* (in)direct effects (Pearl, 2001; Robins & Greenland, 1992), for which `medoutcon` provides efficient estimators similar to those of Zheng & van der Laan (2012). By readily incorporating the use of machine learning in the estimation of nuisance parameters (through integration with the `sl3` `R` package (Coyle, Hejazi, Malenica, Phillips, & Sofrygin, 2021) of the `tlverse` ecosystem (van der Laan et al., 2022)), `medoutcon` incorporates state-of-the-art non/semi-parametric estimation techniques, facilitating their adoption in a vast array of settings.

Statement of Need

While there is demonstrable interest in causal mediation analysis in a large variety of disciplines, thoughtfully implementing data analysis strategies based on recent developments in this area is challenging. Contributions in the causal inference and statistics literature largely fall into two key areas. Broadly, the study of identification outlines novel causal effect parameters with properties desirable in real-world settings (e.g., the interventional effects, which can be learned under mediator-outcome confounding) and untestable assumptions under which a statistical functional corresponds to a causal estimand of interest. A complementary line of study develops non/semi-parametric efficiency theory for the statistical functionals outlined in the causal identification literature, allowing for their robust estimation with modern techniques from machine learning. Neither concerns itself with opening the door to applying these estimators in real-world data analyses. Moreover, the implementation of open source software

The Journal of Open Source Software

for efficient estimators of causal effects is complex – for such a task, the data scientist must be knowledgeable of causal inference, semiparametric statistical theory, machine learning, and the intersection of these disciplines, and that is to forego mention of research software engineering best practices, including, for example, unit/regression testing and automated continuous integration. The `medoutcon` `R` package is a free, open source implementation of non/semi-parametric efficient estimators of the natural and interventional (in)direct effects, providing data scientists in research and in industry with access to state-of-the-art statistical methodology for causal mediation analysis. Its estimators have been interrogated in simulation studies and applied in real-world data analyses. To the best of our knowledge, no other `R` package provides similarly convenient access to multiply robust, non/semi-parametric efficient estimators of causal mediation effects with a flexible interface to accommodate machine learning of nuisance parameters.

Natural and Interventional Causal Mediation Effects

To evaluate the causal effects of an exposure on an outcome through mediating pathways, let's consider a dataset of n units, where the observed data on a single unit is assumed to have been generated by a nonparametric structural equation model (NPSEM) (Pearl, 2009):

$$W = f_W(U_W); A = f_A(W, U_A); Z = f_Z(W, A, U_Z);$$
$$M = f_M(W, A, Z, U_M); Y = f_Y(W, A, Z, M, U_Y),$$

where W are baseline (pre-exposure) covariates, $A \in \{0, 1\}$ is the (binary) exposure of interest, Z is an intermediate confounder of the mediator-outcome relationship and is affected by exposure A, M represents mediating variables, and Y is the outcome. This NPSEM admits an equivalent representation as a directed acyclic graph (or DAG), in which each variable is a node and dependencies are represented by directed paths between the nodes. The natural (in)direct effects cannot generally be identified (i.e., learned from the observed data) in the presence of intermediate confounding, so, for now, we make the simplifying assumption that the intermediate variable Z is absent. In this simple case, the population average treatment effect (ATE) – that is, the total effect of A on Y, comparing two exposure contrasts $\{a', a^\star\}$ – may be decomposed into the natural direct effect (NDE) and the natural indirect effect (NIE) as

$$\mathbb{E}[Y(a') - Y(a^\star)] = \underbrace{\mathbb{E}[Y(a', M(a')) - Y(a', M(a^\star))]}_{\text{Indirect effect (through } M)} + \underbrace{\mathbb{E}[Y(a', M(a^\star)) - Y(a^\star, M(a^\star))]}_{\text{Direct effect (not through } M)},$$

where the *counterfactual* variables $Y(\cdot)$ are *potential outcomes* (Hernán & Robins, 2021; Imbens & Rubin, 2015) – that is, $Y(a')$ is the value that the outcome would take when the exposure is set to level a', possibly contrary to fact. Similarly, $M(a^\star)$ is the value that the mediators would take when the exposure is set to level $a^\star$, as the result of an intervention, for example. The NIE captures the effect of the exposure A on Y through the mediating variables M while the NDE captures the effect of A on Y through all other pathways. Robins & Greenland (1992) and Pearl (2001) independently studied this decomposition within the potential outcomes and NPSEM frameworks, respectively. In both cases, the NDE and NIE are derived from the ATE by introducing a decomposition term that deterministically sets the values of the exposure and mediators to differing values by the application of *static* interventions. As regards estimation, Tchetgen Tchetgen & Shpitser (2012) and Zheng & van der Laan (2012) outlined non/semi-parametric efficiency theory for developing estimators of the NDE and NIE and proposed efficient estimators of these causal quantities.

The presence of intermediate confounders Z often cannot be ruled out in real-world data analysis scenarios. Such post-exposure variables, which are affected by A and affect both M and Y, complicate efforts to disentangle the effect of A on Y through paths involving

M and other paths. Recognizing the limitations of the natural effects in these settings, Didelez et al. (2006), Petersen et al. (2006), VanderWeele et al. (2014), and Rudolph et al. (2017), among others, contributed to the development of the interventional (in)direct effects. Unlike the decomposition strategy that delineates the NDE and NIE, these effects require a more sophisticated approach to identification, relying upon *stochastic* interventions on the mediator(s), which require random draws from the mediators post-intervention distribution rather than the setting of fixed counterfactual values. Specifically, for the two exposure contrasts $\{a', a^\star\}$, the effect of A on Y can be defined as the difference in expected outcome in the hypothetical worlds in which $(A, M) = (a', G_{a'})$ versus $(A, M) = (a^\star, G_{a^\star})$. Here, G_a denotes a random draw from the conditional distribution of M_a conditional on W, as defined by a stochastic intervention. The direct and indirect effects are defined as follows

$$\mathbb{E}[Y(a', G_{a'}) - Y(a^\star, G_{a^\star})] = \underbrace{\mathbb{E}[Y(a', G_{a'}) - Y(a', G_{a^\star})]}_{\text{Indirect effect (through } M)} + \underbrace{\mathbb{E}[Y(a', G_{a^\star}) - Y(a^\star, G_{a^\star})]}_{\text{Direct effect (not through } M)}.$$

Like the NDE, this interventional direct effect measures the effects through all paths avoiding the mediating variables. Analogous to the NIE, the interventional indirect effect measures the effect through paths involving the mediators. Note, however, that natural and interventional mediation effects have different interpretations. That is, the interventional indirect effect measures the effect of fixing the exposure at a' while setting the mediator to a random draw $G_{a^\star}$ (i.e., under an intervention setting the exposure to $a^\star$) versus a random draw $G_{a'}$ (i.e., after setting the exposure to a'), given covariates W. Intuitively, the interventional effects remain identifiable under intermediate confounding since the stochastic intervention on the mediators breaks the relationship between Z and M. Prior to the work of Díaz et al. (2020), and contemporaneous developments by Benkeser & Ran (2021), non/semi-parametric efficiency theory for the interventional (in)direct effects was unavailable. Recently, a novel family of interventional effects, accommodating flexible stochastic interventions on the exposure, have been formulated (Hejazi et al., 2021).

`medoutcon`'s Scope

Development of the `medoutcon` package began as a software accompaniment to the theoretical developments of Díaz et al. (2020) – where the investigations of these authors outlined efficient estimators of the interventional (in)direct effects, `medoutcon` implements these efficient estimators. Implemented in the `R` language and environment for statistical computing (R Core Team, 2021), `medoutcon` aims to provide a simple application user interface (API) for convenience in a variety of data analytic applications. Specifically, `medoutcon` – via a single, user-facing eponymous function `medoutcon()` – provides access to both one-step and targeted minimum loss (TML) estimators of these causal (in)direct effects. State-of-the-art machine learning algorithms, including ensemble modeling (van der Laan et al., 2007), may readily be used for the estimation of relevant nuisance parameters, through a design that tightly couples `medoutcon` with the `sl3` `R` package (Coyle, Hejazi, Malenica, Phillips, & Sofrygin, 2021). Cross-fitting is automatically incorporated, via the `origami` `R` package (Coyle, Hejazi, Malenica, & Phillips, 2021; Coyle & Hejazi, 2018), in computing the efficient estimators, allowing for some common but restrictive regularity conditions to be relaxed (Bickel et al., 1993; Chernozhukov et al., 2017; Zheng & van der Laan, 2011).

Beyond implementing the interventional (in)direct effects, `medoutcon` additionally allows for the natural (in)direct effects to be estimated when intermediate confounders are omitted from the call to the `medoutcon()` function (i.e., by setting `Z = NULL`). This feature is based on a correspondence between the identifying statistical functionals of the natural and interventional (in)direct effects in the absence of intermediate confounding. In this simplified case, the efficient estimators of the interventional (in)direct effects formulated by Díaz et al. (2020) are analogous to the efficient estimators of the natural (in)direct effects formulated

by Zheng & van der Laan (2012). By supporting this case, `medoutcon` serves as a one-stop tool for estimating these classical and popular causal mediation effects, allowing for practicing data scientists and applied statisticians to deploy cutting-edge estimators of the natural and interventional (in)direct effects through a unified API.

Availability

The `medoutcon` package is publicly available via GitHub, with plans for submission to the Comprehensive `R` Archive Network, pending the inclusion of its dependencies (`sl3`, in particular) in that repository. Use of the `medoutcon` package has been extensively documented in the package's `README`, a vignette, and its documentation website. Ongoing development of the package incorporates research and data science software engineering best practices, including a suite of unit tests and automated continuous integration checking.

Acknowledgments

NSH's contributions to this work were supported in part by a grant from the National Science Foundation (award number DMS 2102840).

References

Benkeser, D., Díaz, I., & Ran, J. (2021). Inference for natural mediation effects under case-cohort sampling with applications in identifying COVID-19 vaccine correlates of protection. *arXiv Preprint arXiv:2103.02643*. https://arxiv.org/abs/2103.02643

Benkeser, D., & Ran, J. (2021). Nonparametric inference for interventional effects with multiple mediators. *Journal of Causal Inference*. https://doi.org/10.1515/jci-2020-0018

Bickel, P. J., Klaassen, C. A., Ritov, Y., & Wellner, J. A. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press Baltimore.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., & Newey, W. (2017). Double/debiased/Neyman machine learning of treatment effects. *American Economic Review*, *107*(5), 261–265. https://doi.org/10.1257/aer.p20171038

Coyle, J. R., & Hejazi, N. S. (2018). Origami: A generalized framework for cross-validation in r. *Journal of Open Source Software*, *3*(21). https://doi.org/10.21105/joss.00512

Coyle, J. R., Hejazi, N. S., Malenica, I., & Phillips, R. V. (2021). *origami: Generalized framework for cross-validation* (Version 1.0.5) [Computer software]. https://doi.org/10.5281/zenodo.835602

Coyle, J. R., Hejazi, N. S., Malenica, I., Phillips, R. V., & Sofrygin, O. (2021). *sl3: Modern pipelines for machine learning and Super Learning* (Version 1.4.3) [Computer software]. https://doi.org/10.5281/zenodo.1342293

Didelez, V., Dawid, P., & Geneletti, S. (2006). Direct and indirect effects of sequential treatments. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, 138–146.

Díaz, I., Hejazi, N. S., Rudolph, K. E., & van der Laan, M. J. (2020). Non-parametric efficient causal mediation with intermediate confounders. *Biometrika*. https://doi.org/10.1093/biomet/asaa085

Hejazi, N. S., Rudolph, K. E., van der Laan, M. J., & Díaz, I. (2021). Nonparametric causal mediation analysis for stochastic interventional (in)direct effects. *Revision Invited at Biostatistics*.

Hejazi, N. S., van der Laan, M. J., Janes, H. E., Gilbert, P. B., & Benkeser, D. C. (2020). Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*. https://doi.org/10.1111/biom.13375

Hernán, M. A., & Robins, J. M. (2021). *Causal Inference: What If*. CRC Boca Raton, FL.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Pearl, J. (2001). Direct and indirect effects. *arXiv Preprint arXiv:1301.2300*.

Pearl, J. (2009). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Petersen, M. L., Sinisi, S. E., & van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology*, 276–284. https://doi.org/10.1097/01.ede.0000208475.99429.2d

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://R-project.org/

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 143–155. https://doi.org/10.1097/00001648-199203000-00013

Rudolph, K. E., Díaz, I., Hejazi, N. S., van der Laan, M. J., Luo, S. X., Shulman, M., Campbell, A., Rotrosen, J., & Nunes, E. V. (2020). Explaining differential effects on opioid use disorder treatment using a novel causal approach incorporating mediating and intermediate variables. *Addiction*. https://doi.org/10.1111/add.15377

Rudolph, K. E., Gimbrone, C., & Díaz, I. (2021). Helped into harm: Mediation of a housing voucher intervention on mental health and substance use in boys. *Epidemiology*, *32*(3), 336–346.

Rudolph, K. E., Sofrygin, O., Zheng, W., & van der Laan, M. J. (2017). Robust and flexible estimation of stochastic mediation effects: A proposed method and example in a randomized trial setting. *Epidemiologic Methods*, *7*(1). https://doi.org/10.1515/em-2017-0007

Tchetgen Tchetgen, E. J., & Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: Efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of Statistics*, *40*(3), 1816–1845. https://doi.org/10.1214/12-AOS990

van der Laan, M. J., Coyle, J. R., Hejazi, N. S., Malenica, I., Phillips, R. V., & Hubbard, A. E. (2022). *Targeted Learning in R: Causal Data Science with the `tlverse` Software Ecosystem*. CRC Press.

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

VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology (Cambridge, Mass.)*, *25*(2), 300. https://doi.org/10.1097/ede.0000000000000034

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*(7), 557–585.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, *5*(3), 161–215. https://doi.org/10.1007/978-3-319-59626-6_5

Zheng, W., & van der Laan, M. J. (2011). Cross-validated targeted minimum-loss-based estimation. In *Targeted learning* (pp. 459–474). Springer. https://doi.org/10.1007/978-1-4419-9782-1_27

Zheng, W., & van der Laan, M. J. (2012). Targeted maximum likelihood estimation of natural direct effects. *International Journal of Biostatistics*, *8*(1). https://doi.org/10.2202/1557-4679.1361