# Partial identification and efficient estimation for the stratum-specific probability of benefit with thresholds on a continuous outcome

**Johan de Aguas**[1,3] **Sebastian Krumscheid**[2] **Johan Pensar**[1] **Guido Biele**[3]

[1]Department of Mathematics, University of Oslo, Oslo, Norway
[2]Scientific Computing Center, Karlsruhe Institute of Technology, Karlsruhe, Germany
[3]Department of Child Health and Development, Norwegian Institute of Public Health, Oslo, Norway

## Abstract

We define the probability of benefit for a scenario involving a binary exposure, a continuous outcome, and a partition of the outcome support with $K$ fixed thresholds. As with other counterfactual queries, this parameter is often not $g$-identifiable, and we show that monotonicity assumption is not sufficient when $K > 1$. We introduce a partial identification strategy by adapting existing bounds. Conducting asymptotic inference and uncertainty quantification for estimates of these bounds is challenging due to potential nonregularity and the lack of differentiability of the involved functionals. Moreover, results might be sensitive to model specification. To address this, we reformulate the problem in terms of individualized rules, adapting the available online one-step estimator with stabilizing weights. We show the connection with solutions based on conservative optimal transport and illustrate the advantages over surrogate bounds derived from smooth approximations. We present an application aimed at estimating the probability of benefit from pharmacological treatment for ADHD upon school performance using observational data.

## 1 INTRODUCTION

One important counterfactual parameter is the *probability of benefit*, which gauges the extent to which an exposure acts as a *necessary and sufficient cause* for a desired outcome event [Pearl, 1999]. It quantifies the proportion of *responders* in the population, referring to units who would have benefited *if and only if* had they been treated. This parameter is typically not $g$-identifiable, meaning it cannot be determined from any combination of observational data and experiments [Robins and Greenland, 1989, Pearl, 1999]. While $g$-identification may be attainable under a monotonicity assumption [Balke and Pearl, 1997], such condition is not always realistic, as it presupposes the absence of unintended effects from the exposure. An alternative approach is to pursue *partial identification* with bounds, which can be computed from population-level observational and experimental distributions, or solely from the former under the assumption of conditional ignorability [Tian and Pearl, 2000]. However, conducting asymptotic inference and uncertainty quantification for estimates of these bounds is challenging due to the potential nonregularity and lack of smoothness of the involved functionals. This difficulty stems from impossibility results for targets that fail to be *pathwise differentiable* at the true distribution [Hirano and Porter, 2012, Dümbgen, 1993, Fang and Santos, 2019, Kitagawa et al., 2020].

The studied scenario involves a categorical pre-exposure covariate $X \in \mathcal{X}$, a binary exposure $A \in \{0, 1\}$, and an absolutely continuous outcome $Y \in \mathcal{Y} \subseteq \mathbb{R}$, with partitioned support $\inf \mathcal{Y} = c_0 < c_1 < \cdots < c_K < \sup \mathcal{Y}$. We define the $x$-*specific probability of benefit* $\mathrm{PB}(x)$ for stratum $X = x$, with $\mathbb{P}(X = x) > 0$, as the joint probability of attaining an outcome value under placebo $Y^0$ within any given $c$-interval and, counterfactually, an outcome value under exposure $Y^1$ within any higher $c$-interval:

$$\mathrm{PB}(x) = \sum_{k=1}^{K} \mathbb{P}(Y^0 \in (c_{k-1}, c_k], Y^1 > c_k \mid X = x) \quad (1)$$

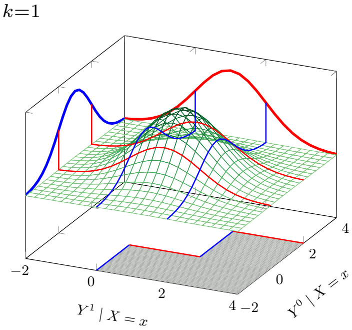


Figure 1: $K = 2, c_1 = 0, c_2 = 2$: $\mathrm{PB}(x)$ is the volume under the joint PDF of $(Y^0, Y^1 \mid X = x)$ enclosed above the gray area.

*Accepted for 9th Causal Inference Workshop at* the 40th Conference on Uncertainty in Artificial Intelligence 2024.

## 2 PARTIAL IDENTIFICATION

**Monotonicity:** A potential outcome $Y^A$ is said to be *monotonic* in relation to the exposure, at stratum $X = x$, if $\mathbb{P}(Y^0 \leq Y^1 \mid X = x) \in \{0, 1\}$.

If the exposure is known to be nondeleterious for the stratum, then monotonicity implies $\mathbb{P}(Y^0 > c_k, Y^1 \in I_k \mid X = x) = 0$ holds for all $k \in [K]$ [Vlontzos et al., 2023]. When $K = 1$, this condition is sufficient for $g$-identification of $\mathrm{PNS}(x)$ with intervention nodes on $\mathcal{E} = \{A\}$, yielding $\mathrm{PNS}(x) = \mathbb{P}(Y^1 > c_1 \mid X = x) - \mathbb{P}(Y^0 > c_1 \mid X = x)$ [Tian and Pearl, 2000, Mueller and Pearl, 2023].

**Proposition 1** *For $K > 1$, monotonicity is not sufficient for $g$-identification of $\mathrm{PB}(x)$ with $\mathcal{E} = \{A\}$.*

**Proposition 2** *Let $Z \in \mathcal{Z}$ be a set of pre-exposure variables such that conditional ignorability holds with $Z \cup X$, i.e., $Y^a \perp\!\!\!\perp A \mid Z, X$, $\forall a \in \{0, 1\}$. Then, bounds for $\mathrm{PB}(x)$, $\Lambda(x) \leq \mathrm{PB}(x) \leq \Upsilon(x)$, are given by:*

$$\Lambda(x) = \sum_{k=1}^{K} \mathbb{E}_{Z|X=x} \max\{0; R_k(Z, x, 0) + S_k(Z, x, 1) - 1\}$$
$$\Upsilon(x) = \sum_{k=1}^{K} \mathbb{E}_{Z|X=x} \min\{R_k(Z, x, 0);\ S_k(Z, x, 1)\}, \quad (2)$$

where $S_k(z, x, a)$ and $R_k(z, x, a)$ denote, respectively, the probability of passing threshold $c_k$ and the probability of staying at $c$-interval $I_k$ under treatment $A = a$ for stratum $(Z = z, X = x)$. This is:

$$S_k(z, x, a) = \mathbb{P}(Y > c_k \mid Z = z, X = x, A = a), \quad (3)$$
$$R_k(z, x, a) = S_{k-1}(z, x, a) - S_k(z, x, a). \quad (4)$$

These bounds are based on those derived by Li and Pearl [2024], Mueller et al. [2022] for the general categorical case, and are intimately related to the *Fréchet-Hoeffding* sharp bounds for joint events [Hoeffding, 1940, Fréchet, 1951].

## 3 TARGET ESTIMAND, ESTIMATORS AND STATISTICAL INFERENCE

Consider the following reformulation of the problem in terms of individualized rules $\lambda_k, \upsilon_k : \mathcal{Z} \times \mathcal{X} \to \{0, 1\}$:

$$\lambda_k(z, x) = \mathbb{I}[R_k(z, x, 0) + S_k(z, x, 1) - 1 > 0], \quad (5)$$
$$\upsilon_k(z, x) = \mathbb{I}[R_k(z, x, 0) - S_k(z, x, 1) > 0]. \quad (6)$$

Intuitively, these rules pick the *index* of the solution terms in the respective optimization problems at the individual level. For instance, $\lambda_k(z, x) = 1$ indicates that the second term in $\max\{0;\ R_k(z, x, 0) + S_k(z, x, 1) - 1\}$ is the maximum of the two. These rules are well-defined and nonambiguous, even when the two terms being compared equate. Their *values* (negative in case of $\Upsilon_k$) are given by:

$$\begin{aligned} \Lambda_k(z, x, \lambda_k) &= \lambda_k(z, x) \cdot [R_k(z, x, 0) + S_k(z, x, 1) - 1], \\ \Upsilon_k(z, x, \upsilon_k) &= \upsilon_k(z, x) \cdot [R_k(z, x, 0) - S_k(z, x, 1)] \\ &\quad - R_k(z, x, 0). \end{aligned} \quad (7)$$

The target estimand comprises the bound functionals as joint components, for $x$ with $\mathbb{P}(X = x) > 0$:

$$\Psi[P_0](x) = (\Lambda(x, \lambda);\ \Upsilon(x, \upsilon))^\top \in \mathbb{R}^2, \quad (8)$$

Under some positivity and boundedness conditions, $\Lambda(x, \lambda^*)$ and $\Upsilon(x, \upsilon^*)$ are pathwise differentiable at $P_0$ for **fixed** rules $\lambda^*, \upsilon^*$. Yet, in the original problem, the rules are not fixed but they are to be estimated. *Exceptional laws*, might induce ambiguities that make the inference problem nonregular [Robins, 2004]. If, for at least one $k \in [K]$, either the two arguments of max or of min in equation (2) are equal, the problem becomes nonregular. Under an exceptional law, the limiting distribution is nonstandard, and no locally unbiased estimator exists [Hirano and Porter, 2012]. Moreover, the naïve plug-in estimator of bound functionals is known to be highly sensitive to model misspecification, particularly when covariates are continuous or discrete with high-dimensional support, unless parametric, smoothness, or sparsity assumptions are imposed [Ji et al., 2023].

We adapted the methodology proposed by Luedtke and van der Laan [2016, 2018], extending it to address multiple sources of nonregularity, employing stabilizing matrices instead of scalar weights for joint inference, and focusing on stratum-specific queries. The proposed procedure mitigates plug-in bias by utilizing an AIPW-based estimator of cumulative intervention distributions. Initial estimators are developed for a semiparametric heteroskedastic regression model. We demonstrate connections to and integrate ideas from *structural nested distribution models* [Vansteelandt and Joffe, 2014], as well as conservative solutions based on *optimal transport* in both the primal [Balakrishnan et al., 2023] and dual formulations [Ji et al., 2023].

## 4 APPLICATION AND DISCUSSION

In simulation studies, the OOSE with stabilizing matrices achieved nominal coverage for the bounds individually. The proposed method consistently outperformed the plug-in and GELU-smoothed approaches in MSE.

We employ developed procedures to estimate the probability of benefit from pharmacological treatment for ADHD, using official thresholds between mastery levels 1, 2, and 3 in the grade 8 national numeracy test. The upper bound results indicates that ADHD stimulant medication may benefit a minority of the population.

The estimator presented in this paper could be applied to other bounds in causal inference, including assumption-free bounds for causal effects [Pearl, 1999], instrumental variables [Balke and Pearl, 1997], sensitivity analysis [Diaz et al., 2018], and latent confounding [Ding and VanderWeele, 2016, VanderWeele and Ding, 2017]. We intend to explore scope of application and challenges in future work.

## References


Sivaraman Balakrishnan, Edward Kennedy, and Larry Wasserman. Conservative inference for counterfactuals. *arXiv preprint arXiv:2310.12757*, 2023.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176, 1997. ISSN 01621459. URL http://www.jstor.org/stable/2965583.

Ivan Diaz, Alexander R. Luedtke, and Mark J van der Laan. *Sensitivity Analysis*, pages 511–522. Springer International Publishing, Cham, 2018. ISBN 978-3-319-65303-7. doi: 10.1007/978-3-319-65304-4_27. URL https://doi.org/10.1007/978-3-319-65304-4_27.

Peng Ding and Tyler J VanderWeele. Sensitivity analysis without assumptions. *Epidemiology*, 27(3):368–377, 2016.

Lutz Dümbgen. On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*, 95:125–140, 1993.

Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *The Review of Economic Studies*, 86(1 (306)):377–412, 2019.

Maurice Fréchet. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon. Série A: Sciences mathématiques et astronomie*, 14:53–77, 1951.

Keisuke Hirano and Jack R. Porter. Impossibility results for nondifferentiable functionals. *Econometrica*, 80(4):1769–1790, 2012. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/23271416.

Wassily Hoeffding. *Masstabinvariante Korrelationstheorie*, volume 5 of *Schriften des Mathematischen Instituts und des Institutes für Angewandte Mathematik der Universität Berlin*. Universität Berlin, Berlin, Germany, 1940.

Wenlong Ji, Lihua Lei, and Asher Spector. Model-agnostic covariate-assisted inference on partially identified causal effects, 2023. URL https://arxiv.org/abs/2310.08115.

Toru Kitagawa, José Luis Montiel Olea, Jonathan Payne, and Amilcar Velez. Posterior distribution of nondifferentiable functions. *Journal of Econometrics*, 217(1):161–175, 2020. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2019.10.009. URL https://www.sciencedirect.com/science/article/pii/S0304407620300014.

Ang Li and Judea Pearl. Probabilities of causation with nonbinary treatment and effect. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 20465–20472, 2024.

Alexander R Luedtke and Mark J van der Laan. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics*, 44(2):713, 2016.

Alexander R. Luedtke and Mark J. van der Laan. Parametric-rate inference for one-sided differentiable parameters. *Journal of the American Statistical Association*, 113(522): 780–788, 2018. doi: 10.1080/01621459.2017.1285777. URL https://doi.org/10.1080/01621459.2017.1285777. PMID: 30078921.

S. Mueller and J. Pearl. Monotonicity: Detection, refutation, and ramification. Technical Report R-529, Department of Computer Science, University of California, Los Angeles, CA, 2023. URL http://ftp.cs.ucla.edu/pub/stat_ser/r529.pdf.

Scott Mueller, Ang Li, and Judea Pearl. Causes of effects: Learning individual responses from population data. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 2712–2718. International Joint Conferences on Artificial Intelligence Organization, 7 2022. doi: 10.24963/ijcai.2022/376. URL https://doi.org/10.24963/ijcai.2022/376. Main Track.

Judea Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, 121(1):93–149, 1999. doi: 10.1023/A:1005233831499. URL https://doi.org/10.1023/A:1005233831499.

James M Robins. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium in Biostatistics*, Lecture Notes in Statistics, pages 179–326. Springer, New York, 2004.

James M. Robins and Sander Greenland. The probability of causation under a stochastic model for individual risk. *Biometrics*, 45(4):1125–1138, 1989. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2531765.

Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1):287–313, 2000. doi: 10.1023/A:1018912507879. URL https://doi.org/10.1023/A:1018912507879.

Tyler J VanderWeele and Peng Ding. Sensitivity analysis in observational research: introducing the e-value. *Annals of internal medicine*, 167(4):268–274, 2017.

Stijn Vansteelandt and Marshall Joffe. Structural Nested Models and G-estimation: The Partially Realized Promise. *Statistical Science*, 29(4):707 – 731, 2014. doi: 10.1214/14-STS493. URL https://doi.org/10.1214/14-STS493.

Athanasios Vlontzos, Bernhard Kainz, and Ciarán M. Gilligan-Lee. Estimating categorical counterfactuals via deep twin networks. *Nature Machine Intelligence*, 5(2):159–168, 2023. doi: 10.1038/s42256-023-00611-x. URL https://doi.org/10.1038/s42256-023-00611-x.