TECHNICAL REPORT
R-529
August 2023

Monotonicity: Detection, Refutation, and Ramification

Scott Mueller and Judea Pearl

Abstract

The assumption of monotonicity, namely that outputs cannot decrease when inputs increase, is critical for many reasoning tasks, including unit selection, A/B testing, and quasi-experimental econometrics. It is also vital for identifying Probabilities of Causation, which, in turn, enable the estimation of individual-level behavior. This paper demonstrates how monotonicity can be detected (or refuted) using observational, experimental, or combined data. Using such data, we pinpoint regions where monotonicity is definitively violated, where it unequivocally holds, and where its status remains undetermined. We further explore the consequences of monotonicity violations, especially when a maximum percentage of possible violation is specified. Finally, we illustrate applications for personalized decision-making.

1 Introduction

Many reasoning tasks in healthcare, marketing, and economics are plagued with indeterminacies in the sense that point estimates of some probabilities cannot be obtained even with infinite data. Instead, ranges of values can be derived, but these are often too wide to be useful.

A common thread among these tasks is that indeterminacies are alleviated or eliminated when monotonicity is assumed (i.e. that outputs can never decrease when inputs increase). For example, that no patient can be harmed by a certain treatment, or that no customer will churn when offered an incentive. A formal definition of this notion will be given in Section 2 together with formulas that connect this to the observed data.

To illustrate the role of monotonicity, we first discuss the problem of unit selection [Li and Pearl 2019]. Here the goal is to maximize the gain f associated with a set of units (e.g. patients, customers, or voters) each of them may either benefit from, be harmed by, or remain unaffected by an action under consideration (e.g. treatment, advertisement, or policy). The overall gain, $f(\beta, \gamma, \theta, \delta)$, depends on four parameters: the gain of selecting a unit benefiting from treatment (β), the gain of selecting a unit always having a positive outcome regardless of treatment (γ), the gain of selecting a unit always having a negative outcome regardless of treatment (θ), and the gain of selecting a unit

harmed by treatment (δ). Li showed that when monotonicity does not hold, the overall gain f cannot be point estimated from experimental data alone, as practiced in A/B testing. Moreover, A/B testing, which has been the mainstay of marketing, product development, and other business optimizations, may be grossly sub-optimal, leading to regrettable decisions. Fortunately, the assumption of monotonicity renders optimizations based on A/B testing equivalent to optimizing $f(\beta, \gamma, \theta, \delta)$ over its four parameters. Given data from several sources our paper identifies when this equivalence holds.

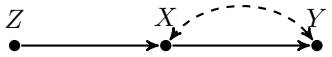

Figure 1: Typical structure for IV methods where Z is an instrument for the relationship between X and Y, shown to be marred by unobserved confounders (bidirectional arrow).

A second task demanding the assumption of monotonicity is Instrumental Variable (IV) analysis[1]. The purpose of IV analysis is to estimate the Average Treatment Effect (ATE) in situations where unobserved confounders exist between treatment X and outcome Y, as shown in Figure 1. When monotonocity can be assumed between the instrument, Z, and the treatment variable, X, the ATE can be identified in certain subpopulations, called "compliers"[2], and is given by:

$$\text{LATE} = E[Y_x - Y_{x'}|\text{complier}] = \frac{E[Y|z] - E[Y|z']}{E[X|z] - E[X|z']} \tag{1}$$

where Y_x is the value Y would have had X been x (treatment) and $Y_{x'}$ is the value Y would have had X been x' (non-treatment). Naturally, this is vital in disciplines where confounding is difficult to deal with, such as experimental econometrics [Imbens and Angrist 1994] and social sciences [Morgan and Winship 2014].

Monotonicity is required to assure the validity of Eq. (1). Absent monotonity, the denominator of (1) may blow up LATE, which further distances LATE from ATE. For a formal definition of IV and its various extensions using graphs see [Pearl 2009, Chapter 7] and [Pearl 2011].

A third task where monotonicity plays an important role is Causes of Effects (CoE) analysis, which aims to estimate the probability that one event is a "cause" of an observed outcome. Examples are assigning credit and blame in legal situations, medical diagnosis, and system troubleshooting. These applications invoke counterfactual reasoning and therefore the desired probabilities cannot be determined from either experimental or observational data. Counterfactual probabilities in common use are Probability of Necessity (PN), Probably

[1] IV analysis is also possible if effect homogeneity holds instead of monotonicity. However, [Hernán and Robins 2020] note that, "homogeneity is often an implausible condition," whereas, "monotonicity [appears] credible in many settings."

[2] A unit is called a complier if treatment is taken if and only if it is assigned to that unit.

of Sufficiency (PS), and Probability of Necessity and Sufficiency (PNS). In medical applications, for example, PNS is the fraction of patients who would survive with treatment *and* die in the absence of treatment [Pearl 1999]. PNS is essential in personalized medicine and personalized decision making [Mueller and Pearl 2022] because it measures the benefit and harm at the individual level.

Tian and Pearl [Tian and Pearl 2000] derived tight bounds on Probabilities of Causation on the basis of experimental and observational data. Mueller, Li, and Pearl [Mueller, Li, and Pearl 2022] further narrowed those bounds by appealing to the causal structure when such is available. These bounds are often too loose to be useful. If monotonicity can be assumed, however, the bounds collapse to a point [Pearl 1999] based on experimental data alone, even without considering the causal structure. If an identifiable causal structure can additionally be assumed on top of monotonicity then PN and PNS are identified with just observational data.

Given its ubiquity in interpreting experimental studies, the need arises to determine when monotonicity is violated, when it can be presumed to hold, and when it definitely holds. In some cases monotonicity is self-evident, for example, in advertising a new product. The control group, not given the information about the product, has no way of purchasing it. Monotonicity must hold because $P(Y_{\text{no ad}} = \text{purchase}) = 0$. In general, however, monotonicity cannot be assured a priori. In medicine, for example, a person might have a 5% chance of being harmed by treatment and a 10% chance of benefiting from it, which may result in a lawsuit if an autopsy proves the former.

This paper shows how data can be assessed for monotonicity. A *necessary test* indicates when monotonicity is possible and a *sufficiency test* indicates when monotonicity is assured (Section 2). An accompanying interactive plot visualizes how necessity and sufficiency depend on experimental and observational data available (Section 3).

2 Monotonicity Tests, Sufficiency and Necessity

Let us denote the variables $X \in \{x, x'\}$ and $Y \in \{y, y'\}$ as binary treatment and recovery, respectively. The values x and x' may represent treatment and no treatment, and the values y and y' may represent recovery and no recovery. We will further use y_x to denote the counterfactual sentence, "Variable Y would have the value y, had X been x." Extensions to multi-valued variables are straightforward [Pearl 2009].

Using these binary variables, monotonicity is defined as,

$$P(y'_x, y_{x'}) \stackrel{\text{def}}{=} P(\text{harm}) = 0. \tag{2}$$

A properly conducted Randomized Controlled Trial (RCT) yields unbiased estimates of $P(y_x)$ and $P(y_{x'})$, from which we can obtain the Average Treatment Effect (ATE), defined as the difference:

$$\text{ATE} \stackrel{\text{def}}{=} P(y_x) - P(y_{x'}). \tag{3}$$

In contrast, observational studies provide estimates of the joint distribution $P(X, Y)$, from which we can obtain $P(x)$, $P(y)$, $P(y|x)$, and $P(y|x')$. Note that an RCT does not directly inform us about $P(\text{harm})$, nor about the other three response types:

$$\text{PNS} \stackrel{\text{def}}{=} P(\text{benefit}) \stackrel{\text{def}}{=} P(y_x, y'_{x'}), \tag{4}$$

$$P(\text{immune}) \stackrel{\text{def}}{=} P(y_x, y_{x'}), \tag{5}$$

$$P(\text{doomed}) \stackrel{\text{def}}{=} P(y'_x, y'_{x'}). \tag{6}$$

As a consequence, in contrary to a prevailing myth, ATE does not represent the proportion of people benefiting from treatment. Note also that the four probabilities above must sum to 1 and that ATE is related to $P(\text{harm})$ and PNS[3] via:

$$P(\text{harm}) = \text{PNS} - \text{ATE}. \tag{7}$$

Eq. (7) tells us immediately that under monotonicity, PNS coincides with ATE, or, in other words, ATE constitutes a point estimate of PNS. More generally, it allows us to compute $P(\text{harm})$ from PNS and ATE, which we will use to define the level of monotonicity violation.

Given these definitions, the question of whether monotonicity is testable can be answered by examining the bounds on $P(\text{harm})$ and asking what conditions would guarantee an upper bound of 0. Given both observational and experimental studies, the bounds on the probability of harm, derived from Eq. (7) and [Tian and Pearl 2000] (which derives tight bounds on PNS), are:

$$\max \left\{ \begin{array}{c} 0, \\ P(y_{x'}) - P(y_x), \\ P(y) - P(y_x), \\ P(y_{x'}) - P(y) \end{array} \right\} \leqslant P(\text{harm}) \leqslant \min \left\{ \begin{array}{c} P(y_{x'}), \\ P(y'_x), \\ P(x, y') + P(x', y), \\ P(y_{x'}) - P(y_x) + \\ P(x, y) + P(x', y') \end{array} \right\}. \tag{8}$$

We see that, when $P(y_x) \geqslant P(y_{x'})$ (or ATE is non-negative), a sufficient condition for monotonicity to hold is that at least one of the arguments to the min function be 0. We can summarize this in a theorem.

Theorem 1. *(Monotonicity Sufficiency Test) Y is monotonic relative to X if*

$$P(y_{x'}) = 0, \ \textit{or} \tag{9}$$

$$P(y_x) = 1, \ \textit{or} \tag{10}$$

$$P(x, y') = P(x', y) = 0, \ \textit{or} \tag{11}$$

$$P(y_x) - P(y_{x'}) = P(x, y) + P(x', y'). \tag{12}$$

[3]Eq. (7) can be obtained by expanding ATE, subtracting $P(y_{x'}) = P(y_x, y_{x'}) + P(y'_x, y_{x'})$ from $P(y_x) = P(y_x, y_{x'}) + P(y_x, y'_{x'})$ to obtain $\text{ATE} = P(y_x, y'_{x'}) - P(y'_x, y_{x'}) = \text{PNS} - P(\text{harm})$.

Note that the left side of Eq. (12) is the ATE. When $P(y_x) < P(y_{x'})$ (ATE is negative), monotonicity must fail because Eq. (7) shows that P(harm) must turn positive.

Unfortunately, conditions (9), (10), (11), and (12) are in the form of equalities and therefore can only materialize in rare cases. In contrast, lack of monotonicity is easier to verify. For this purpose we devise a necessary test for monotonicity, which identifies the requirements for monotonicity to be possible. This test is more informative and is derived by checking if all arguments to the max function in the lower bound of P(harm) are non-positive:

$$\begin{aligned} P(y_{x'}) &\leqslant P(y_x), \text{ and} \\ P(y) &\leqslant P(y_x), \text{ and} \\ P(y_{x'}) &\leqslant P(y). \end{aligned}$$

This can be put into a more succinct form [Pearl 2009, p. 294], as shown in Theorem 2.

Theorem 2. *(Monotonicity Necessity Test) Y is monotonic relative to X only if*

$$P(y_x) \geqslant P(y) \geqslant P(y_{x'}). \tag{13}$$

This is useful for two reasons. First, it can quickly eliminate the possibility of monotonicity by checking for three simple parameters in the data. Second, non-monotonicity implies the existence of subpopulations whose reaction to treatment is substantially different, which, in turn, informs us where the mechanism responsible for that variability could be.

3 Interactive Plot

In a linked website[4], we provide an interactive plot that visualizes regions of data which are necessary or sufficient for monotonicity, as we navigate the terrain of experimental and observational data available. Figure 2 provides a snapshot of this interactive plot. The white regions represent conditions that are required for monotonicity to hold. In other words, finding data outside this region implies the existence of units which can be harmed by treatment. Figure 2 shows this "necessary" region which, in the absence of observational data, is lying below the dashed diagonal line ($P(y_x) \geqslant P(y_{x'})$). Colored bands indicate data regions where monotonicity definitely does not hold and the color in each band indicates the minimum fraction of violation realizable in that band.

Figure 3 shows how the "necessary region" changes when observational data are added. For example, having obtained the additional information of $P(x) = P(y|x) = P(y|x') = 0.5$, the necessary region shrinks to $0.5 \leqslant P(y_x) \leqslant 0.75$ and $0.25 \leqslant P(y_{x'}) \leqslant 0.5$. The transparent gray region indicates areas where $P(y_x)$

[4] Available online at `https://lbmaps.web.app/mns.html`

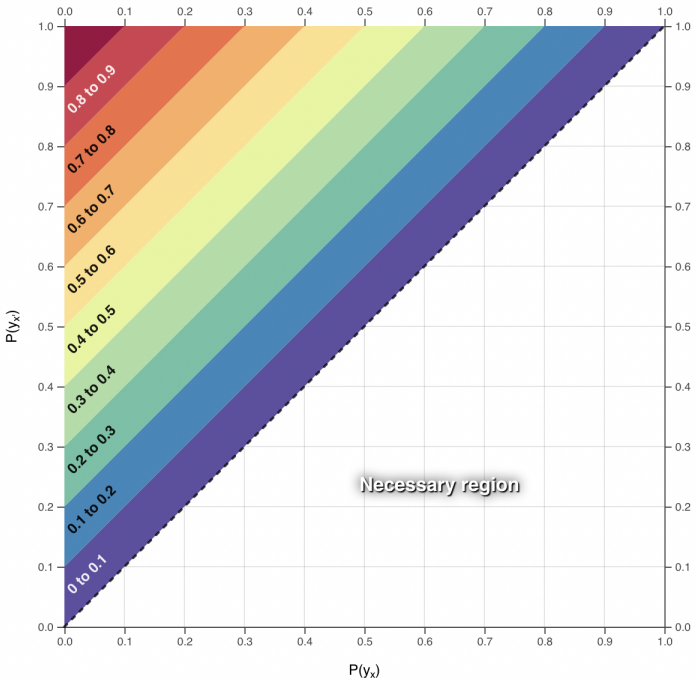

Figure 2: Assuming no observational data, it is necessary for $(P(y_x), P(y_{x'}))$ to be in the white region for monotonicity to hold. The color bands represent the minimum degree to which monotonicity is violated for each $(P(y_x), P(y_{x'}))$ combination.

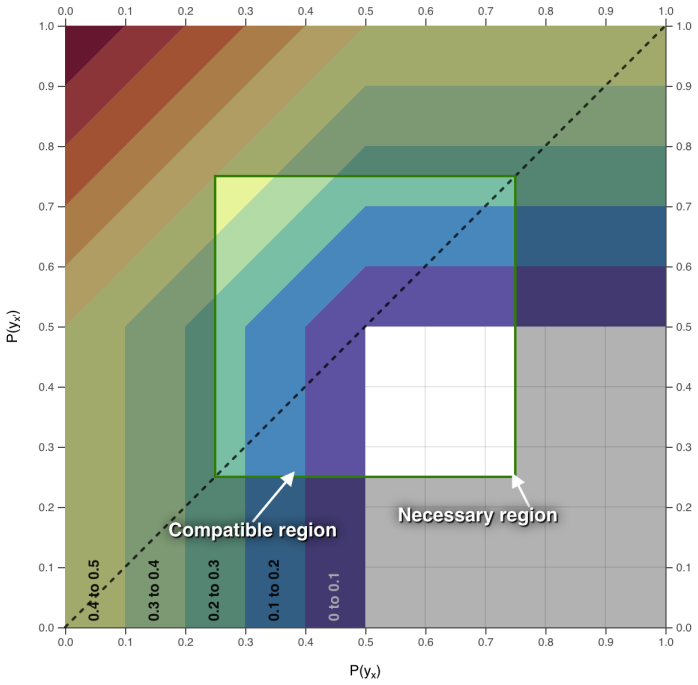

Figure 3: Chart showing the impact of observational data on minimum probability of harm. The square in the middle, labeled "Compatible region", indicates values of $P(y_x)$ and $P(y_{x'})$ which are compatible with the observational data $P(x) = P(y|x) = P(y|x') = 0.5$. Incompatibility implies experimental imperfections. The white square, labeled "Necessary region", indicates where monotonicity may hold. The colors in each band indicate the minimum probability of harm (Eq. (2)).

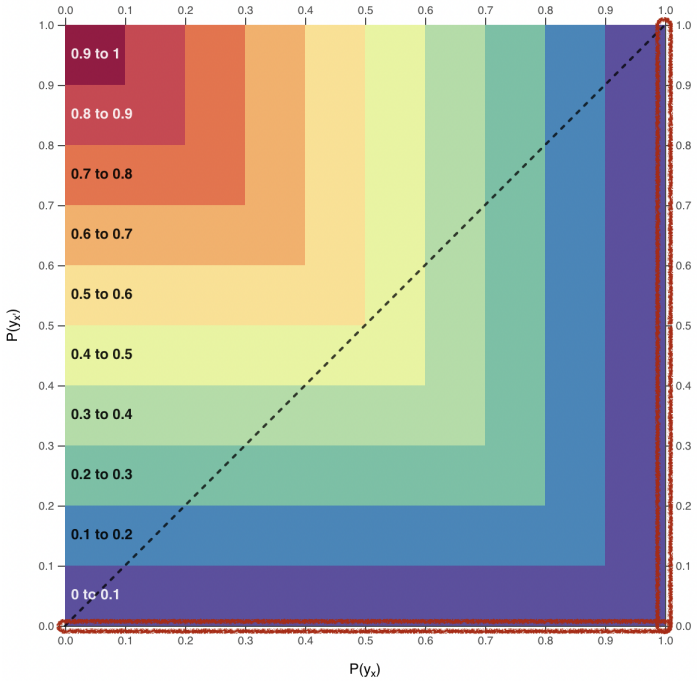

Figure 4: Chart showing maximum probability of harm with no observational data. To guarantee monotonicity ($P(\text{harm}) = 0$), $(P(y_x), P(y_{x'}))$ must be on the bottom or right edge of the chart (marked in red).

and $P(y_{x'})$ are incompatible with the observational data. This could happen, for example, when the population recruited for the experiments is totally different than the one used in the observational study, perhaps due to selection bias. Techniques for detection [Mueller 2023] and overcoming selection bias, under certain conditions, are reported in [Bareinboim, Tian, and Pearl 2014].

In comparison, to identify the vanishingly small regions of sufficiency (where monotonicity must hold), we have to look at the lines marked in red in Figure 4. The rarity of this condition is clear since the region is confined to the edges of the purple band ($P(y_{x'}) = 0$ or $P(y_x) = 1$). With observational data of $P(x) = P(y|x) = P(y|x') = 0.5$, this region shrinks to a single point at the bottom right of the possible region, as seen in Figure 5.

4 ϵ-Bounds on Benefit

As demonstrated with the edges of the plots above and Theorem 1, proving monotonicity from experimental and observational data is uncommon. However, allowing for some units or individuals to be harmed, or only part of the population to be monotonic, may result in a more informed distribution of the

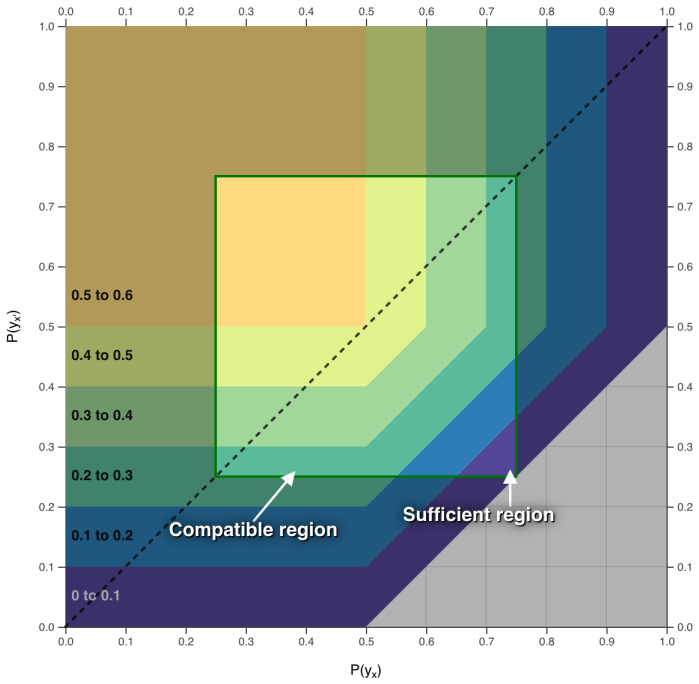

Figure 5: Chart showing the impact of observational data on maximum probability of harm. $(P(y_x), P(y_{x'}))$ is only possible in the center square region if $P(x) = P(y|x) = P(y|x') = 0.5$ and it is sufficient for monotonicity at only one point, the bottom right corner of this compatible region.

beneficiaries, which can be utilized in policy making. We call the new bounds induced by such allowance ϵ-bounds. Theorems 3 and 4 reflect this less restrictive form:

Theorem 3. *(Sufficiency Test for ϵ-Limited Harm) If the following conditions hold, then ϵ must be the maximum proportion of units harmed:*

$$P(y_{x'}) \leqslant \epsilon, \ \textit{or} \tag{14}$$
$$P(y_x) \geqslant 1 - \epsilon, \ \textit{or} \tag{15}$$
$$P(x, y') + P(x', y) \leqslant \epsilon, \ \textit{or} \tag{16}$$
$$P(y_x) - P(y_{x'}) \geqslant P(x, y) + P(x', y') - \epsilon. \tag{17}$$

Theorem 4. *(Necessary Test for ϵ-Limited Harm) If ϵ is the maximum proportion of units harmed, then the following conditions must hold:*

$$P(y_{x'}) \leqslant P(y_x) + \epsilon, \ \textit{and}$$
$$P(y) \leqslant P(y_x) + \epsilon, \ \textit{and}$$
$$P(y_{x'}) \leqslant P(y) + \epsilon.$$

The proof of Theorems 3 and 4 follow directly from the bounds expressed in Eq. (8).

4.1 ϵ-Bounds on PNS

Assumptions of ϵ-limited harm can be used to narrow the bounds on the PNS. From Eq. (7), PNS can be expressed in terms of ATE and P(harm):

$$\text{PNS} = \text{ATE} + P(\text{harm}). \tag{18}$$

Assuming the inequality $0 \leqslant P(\text{harm}) \leqslant \epsilon$ gives:

$$\begin{gathered} \text{ATE} \leqslant \text{PNS} \leqslant \text{ATE} + \epsilon, \\ P(y_x) - P(y_{x'}) \leqslant \text{PNS} \leqslant P(y_x) - P(y_{x'}) + \epsilon. \end{gathered} \tag{19}$$

Tian and Pearl (2000) derived tight bounds on PNS [Tian and Pearl 2000]. These bounds resemble those in Eq. (8), with x and x' swapped, and are expressed as:

$$\max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ P(y_x) - P(y), \\ P(y) - P(y_{x'}) \end{array} \right\} \leqslant \text{PNS} \leqslant \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ P(x, y) + P(x', y'), \\ P(y_x) - P(y_{x'}) + \\ P(x, y') + P(x', y) \end{array} \right\}. \tag{20}$$

The lower bound of PNS in (20) already includes ATE as an argument to its max function, so Eq. (19) cannot help with the lower bound. However, Eq. (19) can potentially lower the upper bound of PNS in (20) by adding the right side of (19) as an argument to the min function in (20):

$$\max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ P(y_x) - P(y), \\ P(y) - P(y_{x'}) \end{array} \right\} \leqslant \text{PNS} \leqslant \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ P(x,y) + P(x',y'), \\ P(y_x) - P(y_{x'}) + \\ \quad P(x,y') + P(x',y), \\ P(y_x) - P(y_{x'}) + \epsilon \end{array} \right\}. \quad (21)$$

Note that any reduced bounds on PNS due to limited harm must come from domain-specific knowledge outside the experimental and observational data[5].

5 Examples

The following three examples demonstrate the value of refuting and confirming monotonicity in a mental health context.

5.1 Harmful Effects

A man is suing a pharmaceutical company claiming that he remained depressed *because* of their anti-depressant drug. While the company claims that the drug cannot worsen depression, the plaintiff asserts that he would've been cured on his own but the drug prolonged his depression. He presents the following data from an independent third party's observational study. Does he have a case?

$$P(x) = 0.6, \quad (22)$$
$$P(y|x) = 0.3, \quad (23)$$
$$P(y|x') = 0.5. \quad (24)$$

This data shows that 30% of people choosing to take the drug recovered, while 50% of people choosing not to take the drug recovered. The pharmaceutical company conducts a Randomized Controlled Trial (RCT), suggesting that the reason people choosing their drug fared worse was because their willingness to incur the large expense of the drug was due to more severe depression where psychotherapy was ineffective. The RCT results showed the drug to be 11% effective in all categories measured, demonstrating that the above observational results were, in fact, biased due to confounding:

$$P(y_x) = 0.58,$$
$$P(y_{x'}) = 0.47.$$

[5]The ϵ obtained from data will merely replicate Tian-Pearl bounds. Applying argmin_ϵ to Theorem 3 will provide a ϵ that will not narrow PNS bounds beyond the Tian-Pearl bounds of (20).

The pharmaceutical company's experts further state that none of the drug's chemical mechanisms would make it possible for depression to be extended due to the drug itself. Our analysis, however, gives a different story.

Since $P(y) = P(y|x) \cdot P(x) + P(y|x') \cdot P(x') = 0.38$, the data fails the Monotonicity Necessity Test of Theorem 2. Specifically, $0.38 = P(y) \not\geqslant P(y_{x'}) = 0.47$. After plugging the data into Eq. (8), the lower bound is $P(y_{x'}) - P(y) = 0.47 - 0.38 = 0.09$. Therefore, $P(\text{harm}) \geqslant 9\%$, contrary to the company's claim.

The interactive plot confirms this. After check-marking "Necessary" and "Observational data" and adjusting the probability sliders to match probabilities (22), (23), and (24), the coordinates $(P(y_x), P(y_{x'})) = (0.58, 0.47)$ point to the upper part of the purple band. Therefore, the pharmaceutical company is wrong and there is a risk of people staying depressed due to their drug.

The man has a strong claim in his lawsuit. Furthermore, market research may determine that, among potential customers, half of them would not purchase if they knew some people would remain depressed *because of* the drug. The anxiety induced by this knowledge could also make the drug less effective.

5.2 Confirming No Harm Claim

An ethical pharmaceutical company wants to proclaim that their drug to combat depression does not harm users. They believe that none of their users would simultaneously be cured without the drug *and* remain with depression after using the drug. However, they want to be responsible and confirm this before announcing anything. An RCT and observational study is conducted for this purpose, yielding the following:

$$\begin{aligned} P(x) &= 0.55, \\ P(y|x) &= 0.4, \\ P(y|x') &= 0.6, \\ P(y_x) &= 0.67, \\ P(y_{x'}) &= 0.27. \end{aligned}$$

One of the conditions of the sufficient test, Eq. (12), is true:

$$\begin{aligned} P(x, y) + P(x', y') &= P(y|x) \cdot P(x) + P(y'|x') \cdot P(x') \\ &= P(y|x) \cdot P(x) + [1 - P(y|x')] \cdot [1 - P(x)] \\ &= 0.4 \\ &= \text{ATE} = P(y_x) - P(y_{x'}). \end{aligned}$$

Therefore, the pharmaceutical company can assure their customers of monotonicity. The interactive plot confirms this sufficient condition with the coordinates $(P(y_x), P(y_{x'})) = (0.67, 0.27)$ pointing to the bottom right of the purple band.

5.3 Improved Probability of Benefit

Remaining on our depression-drug theme, a third pharmaceutical company wishes to market their anti-depression drug as having a minimum 50% efficacy level for curing depression. They define efficacy as the proportion of people benefiting. The RCT they conducted for FDA approval yielded the following results:

$$\begin{aligned} P(y_x) &= 0.55, \\ P(y_{x'}) &= 0.46. \end{aligned}$$

With only a paltry difference between experimental probabilities, $P(y_x)$ and $P(y_{x'})$, the ATE is $0.55 - 0.46 = 0.09$, which naively suggests low efficacy. Far below the hoped-for 50%. Even though this *average* treatment effect is low, the proportion of *individuals* benefiting may still be high. This is apparent when combing the RCT results above with the following observational study results:

$$\begin{aligned} P(x) &= 0.35, \\ P(y|x) &= 0.95, \\ P(y|x') &= 0.7, \\ P(y) &= 0.95 \cdot 0.35 + 0.7 \cdot 0.65 = 0.7875. \end{aligned}$$

It appears as though individuals are good at assessing whether they should consume this drug. Both the group choosing the drug and the group avoiding the drug fared better than both the treatment and control arms of the RCT. We can now compute bounds on PNS using Eq. (20), which represents the proportion of individuals benefiting from this drug:

$$\begin{aligned} \max\{0, 0.09, -0.2375, 0.3275\} \leqslant \text{PNS} \leqslant \min\{0.55, 0.54, 0.5275, 0.5625\}, \\ 0.3275 \leqslant \text{PNS} \leqslant 0.5275. \end{aligned}$$

These results do not allow the pharmaceutical company to lay a legitimate claim to minimum 50% efficacy. However, they can claim *potentially up to* 52.75% efficacy. This is a vague claim, but may sway some hopeful depression sufferers to buy the drug.

Psychiatrists report that they believe many of their depressed patients are not getting better *because of* the drug, despite their belief that the drug is effective for an abundance of their other patients. The pharmaceutical company investigates and determines that the molecular mechanism does allow for some patients to be harmed by the drug. However, this mechanism only allows for a maximum of 24% of depressed people to be harmed. While this 0.24-limited harm is not ideal, the ATE is still positive, so psychiatrists and patients are largely amenable to continuing with the medication.

Unfortunately for the pharmaceutical company, 0.24-limited harm affects the PNS bounds. Using Eq. (21):

$$\begin{aligned} 0.3275 \leqslant \text{PNS}_\epsilon \leqslant \min\{0.5275, 0.09 + 0.24\}, \\ 0.3275 \leqslant \text{PNS}_\epsilon \leqslant 0.33, \end{aligned}$$

where PNS_ϵ is the probability of benefit incorporating ϵ-limited harm. PNS is now shrunken to nearly a point estimate. The pharmaceutical company can no longer claim even a possibility of 50% efficacy.

6 Conclusion

Many reasoning tasks, such as unit selection, A/B testing, quasi-experimental econometrics, and, more generally, identification of Probabilities of Causation, benefit substantially from an assumption of monotonicity. In this paper, we have shown how monotonicity can be detected (or refuted) from observational, experimental, or combined data. We then identify when monotonicity is definitely violated, when it definitely holds, and when it is undetermined. We further show the consequences of monotonicity violations when the degree of violation is limited. Examples taken from healthcare were discussed.

References

Bareinboim, Elias, Jin Tian, and Judea Pearl (2014). "Recovering from Selection Bias in Causal and Statistical Inference". In: *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence*. Ed. by Carla E. Brodley and Peter Stone, pp. 2410–2416. URL: http://ftp.cs.ucla.edu/pub/stat_ser/r425.pdf.

Hernán, Miguel and James Robins (2020). *Causal Inference: What If*. 1st edition. Boca Raton: Chapman & Hall/CRC. URL: https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/ (visited on 09/18/2022).

Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62.2. Publisher: [Wiley, Econometric Society], pp. 467–475. ISSN: 0012-9682. DOI: 10.2307/2951620. URL: https://www.jstor.org/stable/2951620 (visited on 10/13/2022).

Li, Ang and Judea Pearl (2019). "Unit selection based on counterfactual logic". In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*. AAAI Press, pp. 1793–1799.

Morgan, Stephen L. and Christopher Winship (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. Analytical Methods for Social Research. Cambridge: Cambridge University Press. ISBN: 978-1-107-06507-9. DOI: 10.1017/CBO9781107587991. URL: https://www.cambridge.org/core/books/counterfactuals-and-causal-inference/5CC81E6DF63C5E5A8B88F79D45E1D1B7 (visited on 08/09/2023).

Mueller, Scott, Ang Li, and Judea Pearl (July 2022). "Causes of Effects: Learning Individual Responses from Population Data". In: *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. Thirty-First International Joint Conference on Artificial Intelligence {IJCAI-22}. Vienna, Austria: International Joint Conferences on Artificial Intelligence

Organization, pp. 2712–2718. ISBN: 978-1-956792-00-3. DOI: 10.24963/ijcai.2022/376. URL: https://www.ijcai.org/proceedings/2022/376 (visited on 09/19/2022).

Mueller, Scott and Judea Pearl (2022). "Personalized Decision Making – A Conceptual Introduction". In: URL: http://ftp.cs.ucla.edu/pub/stat_ser/r513.pdf.

Pearl, Judea (1999). "Probabilities of Causation: Three Counterfactual Interpretations and their identification". In: *Synthese* 121, pp. 93–149. URL: https://ftp.cs.ucla.edu/pub/stat_ser/r260-reprint.pdf.

— (2009). *Causality*. Second. Cambridge University Press.

— (Mar. 30, 2011). "Principal Stratification – a Goal or a Tool?" In: *The International Journal of Biostatistics* 7.1. Publisher: De Gruyter, pp. 1–13. ISSN: 1557-4679. DOI: 10.2202/1557-4679.1322. URL: https://www.degruyter.com/document/doi/10.2202/1557-4679.1322/html (visited on 08/08/2023).

Tian, Jin and Judea Pearl (2000). "Probabilities of causation: Bounds and identification". In: *Annals of Mathematics and Artificial Intelligence* 28.1-4, pp. 287–313. URL: http://ftp.cs.ucla.edu/pub/stat_ser/r271-A.pdf.