

Identification of Probabilities of Causation: from Recursive to Closed-Form Bounds

Xin Shu, Shuai Wang and Ang Li

Department of Computer Science, Florida State University

{xs24a, sw23v}@fsu.edu, angli@cs.fsu.edu

Abstract

Probabilities of causation (PoCs) are fundamental quantities for counterfactual analysis and personalized decision making. However, existing analytical results are largely confined to binary settings. This paper extends PoCs to multi-valued treatments and outcomes by deriving closed form bounds for a representative family of discrete PoCs within Structural Causal Models, using standard experimental and observational distributions. We introduce the notion of equivalence classes of PoCs, which reduces arbitrary discrete PoCs to this family, and establish a replaceability principle that transfers bounds across value permutations. For the resulting bounds, we prove soundness in all dimensions and empirically verify tightness in low dimensional cases via Balke’s linear programming method; we further conjecture that this tightness extends to all dimensions. Simulations indicate that our closed form bounds consistently tighten recent recursive bounds while remaining simpler to compute. Finally, we illustrate the practical relevance of our results through toy examples.

1 Introduction

In fields such as healthcare, finance, and energy, probabilities of causation can enable effective, accurate, and explainable decision-making in complex scenarios. This causality-based approach addresses key challenges, including selection bias, counterfactual reasoning, individualized decision-making, and proper credit assignment—ensuring that outcomes are attributed to the treatments or factors that actually caused them. For example, Li and Pearl [2019] proposed a unit selection framework that uses a linear combination of probabilities of causation to identify individuals who are most likely to respond as desired, while also evaluating the associated rewards and costs. Similarly, Mueller and Pearl [2022] demonstrated how probabilities of causation can guide personalized medical decisions in alignment with the ethical principle of “do no harm.”

The development of probabilities of causation began in the early 2000s, when Pearl [1999] first introduced three key binary measures: the probability of necessity (PN), the proba-

bility of sufficiency (PS), and the probability of necessity and sufficiency (PNS) in the Structural Causal Models (SCMs) [Galles and Pearl, 1998; Halpern, 2000]. Tian and Pearl [2000] then derived tight bounds for these quantities using Balke’s linear programming method [Balke, 1995]. Decades later, Li and Pearl [2019; 2022b] formally proved the validity of these bounds. Building on this foundation, Muller et al. [2021], along with Dawid et al. [2017], further narrowed the bounds by incorporating covariate information and leveraging underlying causal structures.

The aforementioned studies are all restricted to binary treatments and outcomes, which poses significant limitations for real-world applications. In contrast, extending probabilities of causation to the non-binary case enables a more detailed and informative assessment. For example, in a binary setting, physicians can only determine whether a medication is effective or not, leading to a simple yes-or-no prescription decision. However, with multi-valued treatments x_i (e.g., different dosage levels, where $1 \leq i \leq n$ and $n \in \mathbb{N}$), doctors can evaluate which dosage level yields the optimal therapeutic effect for a particular patient. This finer-grained understanding supports more personalized and effective treatment decisions.

Efforts to address non-binary cases have continued in recent work. Zhang et al. [2022], as well as Li and Pearl [2022a], proposed numerical methods for computing non-binary probabilities of causation using nonlinear programming techniques.

More recently, Li and Pearl [2024] introduced recursive-form but un-tight theoretical bounds for non-binary probabilities. Relatedly, de Aguas et al. [2025] studied probabilities of tiered benefit and harm for ordinal outcomes, focusing on cross-threshold causal effects and semiparametric inference. Despite these advances, existing approaches still face important limitations: numerical methods can become computationally demanding as the numbers of treatment and outcome values grow, while recursive analytical bounds may lose sharpness through repeated bounding steps. This lack of sharpness can directly affect downstream decision-making, since bounds that straddle a meaningful decision threshold may leave the relevant conclusion unresolved.

To address these limitations, we derive closed-form, non-recursive bounds for arbitrary non-binary probabilities of causation in general SCMs, using standard experimental and

observational distributions. Our approach first identifies a representative family of non-binary PoCs that is sufficient to express arbitrary discrete PoC queries, and then derives bounds for this family through equivalence classes and a replaceability principle. We prove the soundness of the proposed bounds in all dimensions and empirically verify their tightness in low-dimensional cases, specifically when the treatment and outcome take three or four possible values, using Balke’s linear programming framework [Balke, 1995]. While a complete proof of tightness in the general n -dimensional case remains open.

1.1 Key Contributions

- We derive closed-form bounds for multi-valued PoCs in SCMs from only experimental and observational data.
- We introduce equivalence classes and a replaceability principle to reduce arbitrary PoC queries and transfer bounds across value permutations.
- We prove soundness for all dimensions, verify tightness in low dimensions, and show consistent tightening over Li–Pearl method with improved computational simplicity.

2 Preliminaries

We begin by reviewing the basic definitions of probabilities of causation and the associated notation in Structural Causal Models (SCMs) [Galles and Pearl, 1998; Halpern, 2000]. Readers already familiar with SCMs may choose to skip this section.

Counterfactuals are well defined within the SCM framework. A counterfactual statement such as “Variable Y would have the value y had X been x ” is denoted as $Y_x = y$. Here, the subscript x refers to a hypothetical event, which is well-defined in SCMs. For brevity, we will use $y_x, y_{x'}, y'_x$, and $y'_{x'}$ throughout the paper to represent the events $Y_x = y, Y_{x'} = y, Y_x = y'$, and $Y_{x'} = y'$, respectively. Unless otherwise specified, we assume that experimental data are available in the form of causal effects (e.g., $P(y_x)$), while observational data are provided as joint probabilities (e.g., $P(x, y)$). In our notation, X denotes the treatment variable and Y denotes the outcome or effect.

Let X and Y be two binary variables in a causal model M , where x denotes the proposition $X = \text{true}$, x' denotes $X = \text{false}$, y denotes $Y = \text{true}$, and y' denotes $Y = \text{false}$. The three basic binary probabilities of causation—PN, PS, and PNS—are defined as follows [Pearl, 1999]:

Definition 1 (Probabilities of Causation). [Pearl, 1999]

$$PN \triangleq P(y'_{x'} | x, y), PS \triangleq P(y_x | y', x'), PNS \triangleq P(y_x, y'_{x'})$$

PNS stands for the probability that y would respond to x both ways, and therefore measures both the sufficiency and necessity of x to produce y .

Tian and Pearl [2000] then derived tight bounds for the PN, PS, and PNS using Balke’s linear programming approach [Balke, 1995]. These bounds were later given a formal theoretical proof by Li and Pearl [2019; 2022b].

The tight bounds for binary PNS, PN, and PS, are given as follows:

$$\max \left\{ \begin{array}{c} 0, \\ P(y_x) - P(y_{x'}), \\ P(y) - P(y_{x'}), \\ P(y_x) - P(y) \end{array} \right\} \leq \text{PNS}$$

$$\text{PNS} \leq \min \left\{ \begin{array}{c} P(y_x), \\ P(y'_{x'}), \\ P(x, y) + P(x', y'), \\ P(y_x) - P(y_{x'}) + P(x, y') + P(x', y) \end{array} \right\}$$

$$\max \left\{ \begin{array}{c} 0, \\ \frac{P(y) - P(y_{x'})}{P(x, y)} \end{array} \right\} \leq \text{PN}$$

$$\text{PN} \leq \min \left\{ \begin{array}{c} 1, \\ \frac{P(y'_{x'}) - P(x', y')}{P(x, y)} \end{array} \right\}$$

$$\max \left\{ \begin{array}{c} 0, \\ \frac{P(y') - P(y'_x)}{P(x', y')} \end{array} \right\} \leq \text{PS}$$

$$\text{PS} \leq \min \left\{ \begin{array}{c} 1, \\ \frac{P(y_x) - P(x, y)}{P(x', y')} \end{array} \right\}$$

Regarding non-binary probabilities of causation, Li and Pearl [2024] recently provided eight theorems establishing theoretical (though not tight) recursive-form bounds for cases where both the treatment and the effect are non-binary. Specifically, when the treatment variable X takes m values and the outcome variable Y takes n values, the following set of non-binary probabilities of causation are defined and bounded. (We omit the detailed bounds here; interested readers may refer to [Li and Pearl, 2024] for full derivations.)

Probability of preservation(i, j):

$$P(y_{i x_j}, y_i), \quad s.t., 1 \leq i \leq n, 1 \leq j \leq m$$

Probability of replacement(i, j, k):

$$P(y_{i x_j}, y_k), \quad s.t., 1 \leq i, k \leq n, 1 \leq j \leq m, i \neq k$$

Probability of substitute(i, j, k):

$$P(y_{i x_j}, x_k), \quad s.t., 1 \leq i \leq n, 1 \leq j, k \leq m, j \neq k$$

Probability of necessity(i, j, k, p):

$$P(y_{i x_j}, y_k, x_p), \quad s.t., 1 \leq i, k \leq n, 1 \leq j, p \leq m, j \neq p$$

Probability of necessity and sufficiency(k):

$$P(y_{i_1 x_{j_1}}, \dots, y_{i_k x_{j_k}}), \quad s.t., 1 \leq i_1, \dots, i_k \leq n,$$

$$1 \leq j_1, \dots, j_k \leq m, j_1 \neq \dots \neq j_k$$

Probability of substitute(k, p):

$$P(y_{i_1 x_{j_1}}, \dots, y_{i_k x_{j_k}}, x_p), \quad s.t., 1 \leq i_1, \dots, i_k \leq n,$$

$$1 \leq j_1, \dots, j_k, p \leq m, j_1 \neq \dots \neq j_k \neq p$$

Probability of replacement(k,q):

$$P(y_{i_1 x_{j_1}}, \dots, y_{i_k x_{j_k}}, y_q), \quad s.t., 1 \leq i_1, \dots, i_k, q \leq n,$$

$$1 \leq j_1, \dots, j_k \leq m, j_1 \neq \dots \neq j_k$$

Probability of necessity(k, p, q):

$$P(y_{i_1 x_{j_1}}, \dots, y_{i_k x_{j_k}}, x_p, y_q), \quad s.t., 1 \leq i_1, \dots, i_k, q \leq n,$$

$$1 \leq j_1, \dots, j_k, p \leq m, j_1 \neq \dots \neq j_k \neq p.$$

In this paper, we will provide tighter bounds for a subset of the modified probabilities defined above and demonstrate that these selected probabilities are sufficient to represent all discrete probabilities of causation within SCMs.

3 Main Results

In this section, we first simplify the definitions of the Probability of Necessity and Sufficiency(k), Probability of Substitute(k,p), Probability of Replacement(k,q), and Probability of Necessity(k,p,q) defined by Li and Pearl [2024] into their most concise forms, while retaining their ability to represent all discrete probabilities of causation in Structural Causal Models (SCMs).

3.1 Probability of necessity and sufficiency(k)

The Probability of Necessity and Sufficiency(k) is simplified to the form $P(y_{1x_1}, \dots, y_{kx_k})$, where $|X| = |Y| = n$, the cardinalities of the state spaces of X and Y are both n , and the values of X and Y are arranged in increasing order within the probability expression. This simplification yields the most straightforward closed-form representation. We will later demonstrate that this simplification preserves generality. The closed-form bounds for the modified Probability of Necessity and Sufficiency(k) are presented in the following theorem. All theorem proofs are provided in the appendix.

Theorem 2 (Probability of necessity and sufficiency(k) (PNS(k))). *Suppose variable X has n values $x_1, \dots, x_n$ and Y has n values $y_1, \dots, y_n$, $k \leq n$, then the probability of necessity and sufficiency(k) $P(y_{1x_1}, \dots, y_{kx_k})$ is bounded as following:*

$$\max \left\{ \begin{array}{l} 0, \\ \sum_{j=1}^k P(y_{jx_j}) - k + 1, \\ \sum_{\substack{1 \leq j \leq k \\ j \neq i}} [P(y_{jx_j}) + P(x_j) - P(x_j, y_j)] + \\ \quad + P(x_i, y_i) - k + 1, \quad i \in \{1, \dots, k\} \end{array} \right\} \leq P(y_{1x_1}, \dots, y_{kx_k})$$

$$\min \left\{ \begin{array}{l} \sum_{j=1}^k P(x_j, y_j) + \sum_{j=k+1}^n P(x_j), \\ P(y_{jx_j}), \quad j \in \{1, \dots, k\}, \\ \frac{1}{m} \left[\sum_{j=0}^m P(y_{t_j x_{t_j}}) - P(x_{t_j}, y_{t_j}) \right], \\ m \in \{1, \dots, k-1\}, t_j \in \{1, \dots, k\} \end{array} \right\}$$

$$\geq P(y_{1x_1}, \dots, y_{kx_k})$$

Note that PNS(k) is a nonbinary, higher-order generalization of the original PNS. For example, PNS(2)—which reduces to the binary PNS—represents the probability that y_1 would respond to x_1 and y_2 would respond to x_2 , formally written as $P(y_{1x_1}, y_{2x_2})$. Bounds for similar probabilities, such as $P(y_{2x_1}, y_{1x_2})$, can also be derived by leveraging the replaceability property of the probabilities of causation, as described in Theorem 7.

The proof of soundness, specifying when the bounds are attained with equality, is provided in the appendix. We will also present toy examples to illustrate how our results enhance decision-making capabilities. While some may argue that the previous definition of PNS(k) is more general, we will later demonstrate their equivalence in expressive power in Theorem 6.

3.2 Probability of substitute(k, p)

Similar to the Probability of Necessity and Sufficiency(k), the Probability of Substitute(k, p) is simplified as stated in the following theorem:

Theorem 3 (Probability of substitute(k, p) (PSub(k, p))). *Suppose variable X has n values $x_1, \dots, x_n$ and Y has n values $y_1, \dots, y_n$, $k \leq n$, then the probability $P(y_{1x_1}, \dots, y_{kx_k}, x_p)$, s.t., $p \neq j$ for $1 \leq j \leq k$ is bounded as following:*

$$\max \left\{ \begin{array}{l} 0, \\ \sum_{j=1}^k [P(y_{jx_j}) + P(x_j) - P(x_j, y_j)] \\ \quad + P(x_p) - k \\ \leq P(y_{1x_1}, \dots, y_{kx_k}, x_p) \end{array} \right\}$$

$$\min \left\{ \begin{array}{l} P(x_p), \\ P(y_{jx_j}) - P(x_j, y_j), \quad j \in \{1, \dots, k\} \\ \geq P(y_{1x_1}, \dots, y_{kx_k}, x_p) \end{array} \right\}$$

3.3 Probability of replacement(k, q)

Then, the Probability of replacement(k, q) is simplified as stated in the following theorem:

Theorem 4 (Probability of replacement(k, q) (PRep(k, q))). *Suppose variable X has n values $x_1, \dots, x_n$ and Y has n values $y_1, \dots, y_n$, $k \leq n$, then the probability $P(y_{1x_1}, \dots, y_{kx_k}, y_q)$ is bounded as following:*

$$\max \left\{ \begin{array}{l} 0, \\ \sum_{j=1}^k [P(y_{jx_j}) + P(x_j) - P(x_j, y_j)] + \\ \quad + \sum_{\substack{k+1 \leq j \leq n \\ j \neq q}} P(x_j, y_q) + P(x_q, y_q) - k, \\ \sum_{\substack{1 \leq j \leq k \\ j \neq q}} [P(y_{jx_j}) + P(x_j) - P(x_j, y_j)] + \\ \quad + P(x_q, y_q) - (k-1), \quad \text{if } q \in \{1, \dots, k\} \end{array} \right\}$$

$$\begin{aligned} &\leq P(y_{1x_1}, \dots, y_{kx_k}, y_q) \\ \min &\left\{ \begin{array}{ll} P(y_{qx_q}), & \text{if } q \in \{1, \dots, k\}, \\ P(x_q, y_q) + \sum_{\substack{k+1 \leq j \leq n \\ j \neq q}} P(x_j, y_q), & \\ P(y_{jx_j}) - P(x_j, y_j), & j \in \{1, \dots, k\}, j \neq q \end{array} \right\} \\ &\geq P(y_{1x_1}, \dots, y_{kx_k}, y_q) \end{aligned}$$

3.4 Probability of necessity(k, p, q)

Finally, the Probability of necessity(k, p, q) is simplified as stated in the following theorem:

Theorem 5 (Probability of necessity(k, p, q) (PN(k, p, q))). Suppose variable X has n values $x_1, \dots, x_n$ and Y has n values $y_1, \dots, y_n$, $k \leq n$, then the probability $P(y_{1x_1}, \dots, y_{kx_k}, x_p, y_q)$, s.t., $p \neq j$ for $1 \leq j \leq k$ is bounded as following:

$$\begin{aligned} \max &\left\{ \begin{array}{l} 0, \\ \sum_{j=1}^k [P(y_{jx_j}) + P(x_j) - P(x_j, y_j)] + \\ \quad + P(x_p, y_q) - k \\ \leq P(y_{1x_1}, \dots, y_{kx_k}, x_p, y_q) \end{array} \right\} \\ \min &\left\{ \begin{array}{l} P(x_p, y_q), \\ P(y_{jx_j}) - P(x_j, y_j), \quad j \in \{1, \dots, k\} \\ \geq P(y_{1x_1}, \dots, y_{kx_k}, x_p, y_q) \end{array} \right\} \end{aligned}$$

3.5 Equivalence Class in Probability of Causation

We now proceed to demonstrate the generality of the above modified probabilities of causation through the next two theorems.

Theorem 6 (Equivalence classes in probabilities of causation). Suppose variable X has n values $x_1, \dots, x_n$, Y has m values $y_1, \dots, y_m$:

- **Case 1:** Let Y' have n values $y'_1, \dots, y'_n$. W.L.O.G., let $k \leq m < n$.

Then the bounds of the probability, $P(y_{1x_1}, \dots, y_{kx_k})$, is exactly the same as the bounds of the probability, $P(y'_{1x_1}, \dots, y'_{kx_k})$, where

$$\begin{aligned} P(y'_{lx_j}) &= 0, P(y'_l) = 0, \\ \text{for } m+1 \leq l \leq n, 1 \leq j \leq n, \end{aligned}$$

and,

$$\begin{aligned} P(y'_{lx_j}) &= P(y_{lx_j}), P(x_j, y'_l) = P(x_j, y_l), \\ \text{for } 1 \leq l \leq m, 1 \leq j \leq n. \end{aligned}$$

- **Case 2:** Let X' have m values $x'_1, \dots, x'_m$. W.L.O.G., let $k \leq m < n$.

Then the bounds of the probability, $P(y_{1x_1}, \dots, y_{kx_k})$,

is exactly the same as the bounds of the probability, $P(y'_{1x_1}, \dots, y'_{kx_k})$, where

$$\begin{aligned} P(y_{jx_l}) &= 0, P(y_{mx_l}) = 1, P(x_l) = 0, \\ \text{for } m+1 \leq l \leq n, 1 \leq j \leq m-1, \end{aligned}$$

and,

$$\begin{aligned} P(y_{jx_l}) &= P(y_{jx'_l}), P(x_l, y_j) = P(x'_l, y_j) \\ \text{for } 1 \leq l \leq m, 1 \leq j \leq m. \end{aligned}$$

3.6 Replaceability in Probability of Causation

Theorem 7 (Replaceability in probabilities of causation). Suppose variable X has n values $x_1, \dots, x_n$ and Y has n values $y_1, \dots, y_n$, then the bounds of the probability, $P(y_{1x_1}, \dots, y_{i-1x_{i-1}}, y_{ix_i}, y_{i+1x_{i+1}}, \dots, y_{kx_k})$, can be obtained by replacing y_{ix_i} with y_{ix_i} for any i , such that $1 \leq i, \hat{i} \leq n$, in the bounds of the probability, $P(y_{1x_1}, \dots, y_{kx_k})$.

Note that both theorems apply to the cases of PNS(k), Psub(k, p), PRep(k, q), and PN(k, p, q). For instance, to derive the bounds for $P(y_x, y_{x'}, y''_{x''} \mid x''', y''')$, we can first apply PN(3, 4, 4) to obtain $P(y_x, y'_{x'}, y''_{x''}, x''', y''')$. If $|X| \neq |Y|$, we should first apply Theorem 6. Then, using the replaceability property of probabilities of causation (Theorem 7), we can convert $P(y_x, y'_{x'}, y''_{x''}, x''', y''')$ into $P(y_x, y_{x'}, y''_{x''}, x''', y''')$. Finally, dividing by $P(x''', y''')$ yields the desired conditional probability: $P(y_x, y_{x'}, y''_{x''} \mid x''', y''')$.

4 Examples

In this section, we will discuss several examples to demonstrate the effectiveness and usefulness of our results.

4.1 Cross-Treatment Heterogeneity in Medicine

Consider patients with hypertension who are eligible for three alternative treatment options. Standard clinical summaries report average improvement rates for each treatment, yet they fail to identify patients who would benefit from one treatment but be harmed by another. Since the same patient cannot receive all treatments simultaneously, and responses observed at different times may be confounded by disease progression, such cross-treatment responses are inherently counterfactual, motivating the use of multi-valued probabilities of causation.

Let X denote the treatment assigned to a patient, where x_1, x_2 , and x_3 correspond to Treatment 1, Treatment 2, and Treatment 3, respectively. Let Y denote the clinical outcome category, with y_1, y_2 , and y_3 indicating poor outcome, moderate outcome, and good outcome.

We focus on the probability of causation $P(y_{3x_1}, y_{1x_2}, y_{2x_3})$, which quantifies the fraction of patients who would achieve a good outcome under Treatment 1, a poor outcome under Treatment 2, and only a moderate outcome under Treatment 3. Such cross-treatment response patterns cannot be inferred from marginal outcome rates alone.

To obtain the required estimates, clinicians conduct a randomized clinical trial with 900 patients, where each patient

	Treatment 1	Treatment 2	Treatment 3
poor	46	270	40
moderate	23	8	223
good	231	22	37
Overall	300	300	300

Table 1: Experimental data collected by clinicians.

is assigned to one of the three treatments. The experimental data are summarized in Table 1.

Notably, each marginal experimental probability appearing in the target conjunction, namely $P(y_{3x_1})$, $P(y_{1x_2})$, and $P(y_{2x_3})$, exceeds 0.5. This might suggest that the joint probability $P(y_{3x_1}, y_{1x_2}, y_{2x_3})$ is also large. However, this inference is not justified without causal bounds that account for cross-world dependencies.

In addition, clinicians collected observational data from 900 patients whose treatment choices were self-selected, as summarized in Table 2.

By plugging both the experimental and observational data into Theorem 2 together with the replaceability property (Theorem 7), we obtain the following bounds:

$$0.509 \leq P(y_{3x_1}, y_{1x_2}, y_{2x_3}) \leq 0.588.$$

In contrast, the recursive bounds obtained using the method of Li and Pearl [2024] yield $[0.428, 0.588]$, which leaves substantial ambiguity as to whether this response type constitutes a majority of patients.

Our tighter lower bound resolves this ambiguity by showing that more than half of patients exhibit this cross-treatment response pattern, indicating a large subgroup for which Treatment 1 is beneficial, Treatment 2 is harmful, and Treatment 3 is intermediate. Such risk-relevant heterogeneity is invisible to marginal outcome comparisons and highlights how tighter multi-valued PoC bounds can guide heterogeneity-aware treatment selection.

4.2 Personalized Decision-Making in Educational Interventions

An education platform provides academic support at multiple intensity levels for learners preparing for certification exams. While the platform advertises that increased instructional intensity leads to better learning outcomes on average, advisors are concerned that overly aggressive intervention may harm some learners.

Consider a learner with stable but unexceptional performance. The advisor seeks to assess whether moderate support

	Treatment 1	Treatment 2	Treatment 3
poor	131	45	38
moderate	68	22	483
good	1	51	61
Overall	200	118	582

Table 2: Observational data collected by clinicians.

	Intensive	Moderate	Light	None
Fail	195	11	80	100
Pass	51	266	198	147
Excellent	54	23	22	53
Overall	300	300	300	300

Table 3: Experimental data summarized by the education platform.

would be beneficial, while intensive support could instead be detrimental. Such questions are central to personalized education policy, where one-size-fits-all strategies may be sub-optimal.

Let X denote the level of academic support, ranging from intensive (x_1), moderate (x_2), light (x_3) to none (x_4). Let Y denote the learner’s grade at the end of the term, where y_1 indicates fail the exam (e.g., grade F), y_2 indicates pass the exam, and y_3 indicates excellent performance (e.g., grade A).

The advisor’s query is captured by the probability of causation $P(y_{1x_1}, y_{2x_2}, y_{2x_3} \mid x_4, y_2)$, which asks: among learners who currently pass without support, what is the probability that they would still pass under light or moderate support, but fail under intensive intervention?

To answer this question, the advisor obtains historical data collected by the platform over several academic terms, during which learners participated in different intensity levels while following comparable curricula and assessment standards. The data include both experimental studies (where 1200 learners were assigned to different intervention levels) and observational records (where 1200 learners self-selected their level of support). These results are summarized in Tables 3 and 4.

Based on an observational study, the platform reports that higher levels of intervention are associated with improved performance on average: 16.1%, 7.3%, 5.8%, and 3.2% of learners achieve excellent performance under intensive instruction, moderate instruction, light instruction, and no additional support, respectively, suggesting that intensive training may be beneficial. However, such averages do not address individual-level risks.

We therefore consider:

$$P(y_{1x_1}, y_{2x_2}, y_{2x_3} \mid x_4, y_2) = \frac{P(y_{1x_1}, y_{2x_2}, y_{2x_3}, x_4, y_2)}{P(x_4, y_2)},$$

which quantifies the probability that increasing intervention intensity would negatively affect learners who are currently stable without support.

Applying the experimental and observational estimates to

	Intensive	Moderate	Light	None
Fail	67	11	53	46
Pass	129	17	53	436
Excellent	193	87	70	38
Overall	389	115	176	520

Table 4: Observational data summarized by the education platform.

Theorem 5 and Theorem 7, we obtain:

$$0.0125 \leq P(y_{1x_1}, y_{2x_2}, y_{3x_3}, x_4, y_2) \leq 0.363,$$

which is strictly tighter than the bounds obtained using the method of Li and Pearl [2024]: $[0, 0.363]$.

Since $P(x_4, y_2) = \frac{436}{1200}$, this yields:

$$0.344 \leq P(y_{1x_1}, y_{2x_2}, y_{3x_3} \mid x_4, y_2) \leq 1,$$

which is substantially more informative than the vacuous bound $[0, 1]$ obtained by Li and Pearl [2024] method.

These results allow academic advisors to conclude that, for a non-negligible subset of learners, intensive intervention poses a significant risk of adverse outcomes, while moderate support remains safe.

5 Simulated Results

In this section, we empirically demonstrate that the conjectured closed-form bounds yield tighter estimates than the recursive bounds of Li and Pearl [2024]. Here we focus on the probability of necessity and sufficiency (PNS) to provide a simple and visually direct illustration of how much improvement is obtained in practice. Our generated data satisfy the consistency relation between experimental and observational distributions. We uniformly generate each interventional distribution $P(Y_{x_j} = y_i)$ first, and then uniformly generate observational distributions subject to compatibility constraints with the experimental data. In particular, following Tian-Pearl’s bounds, for any $i, j \in \{1, \dots, n\}$, we require $P(x_j, y_i) \leq P(y_{ix_i}) \leq 1 - P(x_j, y_{i'})$, for all $i' \neq i$.

Dimension sweep ($n \in [3, 20]$). We fix $k = 3$ (i.e., X and Y each take three values) and vary the dimensionality $n \in [3, 20]$, focusing on the probability of necessity and sufficiency $PNS(k)$. For each n , we randomly generate 10,000 ground-truth instances of $PNS(3)$, i.e., $P(y_{1x_1}, y_{2x_2}, y_{3x_3})$. For every instance, we construct compatible observational and experimental sample distributions. Our data-generation procedure enforces the standard feasibility constraint $P(x_i, y_i) \leq P(y_{ix_i}) \leq 1 - P(x_i) + P(x_i, y_i)$. For each n , we compute the average improvement of our bounds over the 10,000 instances and plot the resulting differences in Figure 1. Across all dimensions, the improvements

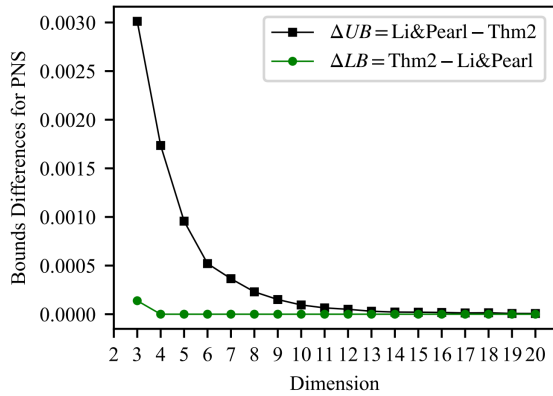


Figure 1: Bounds differences between Li–Pearl and Theorem 2.

n, k	Avg. inc. LB	Avg. dec. UB	Li–Pearl PNS gap	Thm. 2 PNS gap	Bounds narrower
3, 3	0.0002	0.0029	0.1691	0.1661	1737
4, 4	0.0000	0.0024	0.1006	0.0983	1652
5, 5	0.0000	0.0018	0.0669	0.0651	1628
5, 4	0.0000	0.0015	0.0789	0.0775	1186
4, 3	0.0000	0.0018	0.1257	0.1239	1088
5, 3	0.0000	0.0009	0.0952	0.0943	746

Table 5: Comparison of bound tightening and PNS-gap reduction across dimensions over 10,000 sample distributions.

are never negative. Moreover, the average improvement decreases as n increases, which is expected because probability mass is spread over more categories, making typical values of $P(x_i, y_i)$ and $P(y_{ix_i})$ smaller.

Per-instance comparisons. To further illustrate the improvement at the level of individual instances, we consider several representative settings, $n \in \{3, 4, 5\}$, $k \in \{3, 4, 5\}$ (chosen for convenience of presentation). For each setting, we generate 10,000 compatible sample distributions and compute bounds by both methods. For sample i , let $[a_i, b_i]$ denote the bounds from our theorem and $[c_i, d_i]$ denote the bounds from Li and Pearl [2024]. We summarize the following quantities:

- Average increase in the lower bound : $\frac{\sum (a_i - c_i)}{10000}$
- Average decrease in the upper bound : $\frac{\sum (d_i - b_i)}{10000}$
- Average gap from Li and Pearl : $\frac{\sum (d_i - c_i)}{10000}$
- Average gap from theorem 2 : $\frac{\sum (b_i - a_i)}{10000}$
- Count of sample distributions where Theorem 2 yields narrower bounds : $\sum f_i$ where, $f_i = 1$ if $(a_i > c_i)$ or $(b_i < d_i)$ and $f_i = 0$ otherwise.

Table 5 reports these summary statistics across different dimensions. Notably, the bounds are narrowed in approximately 10%–20% of the samples.

Visualizing per-instance improvements. To complement the aggregate trends in Fig. 1 and the summary statistics in

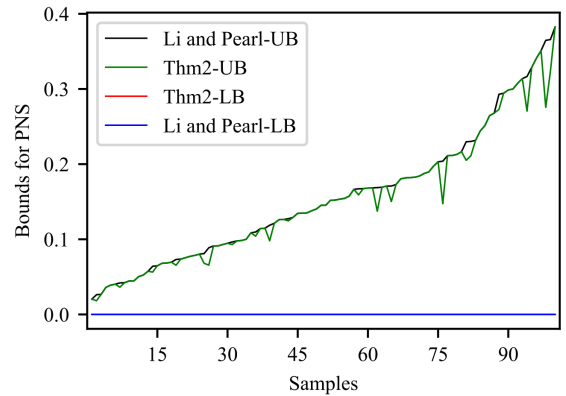


Figure 2: PNS bounds over 100 instances for $n = k = 3$.

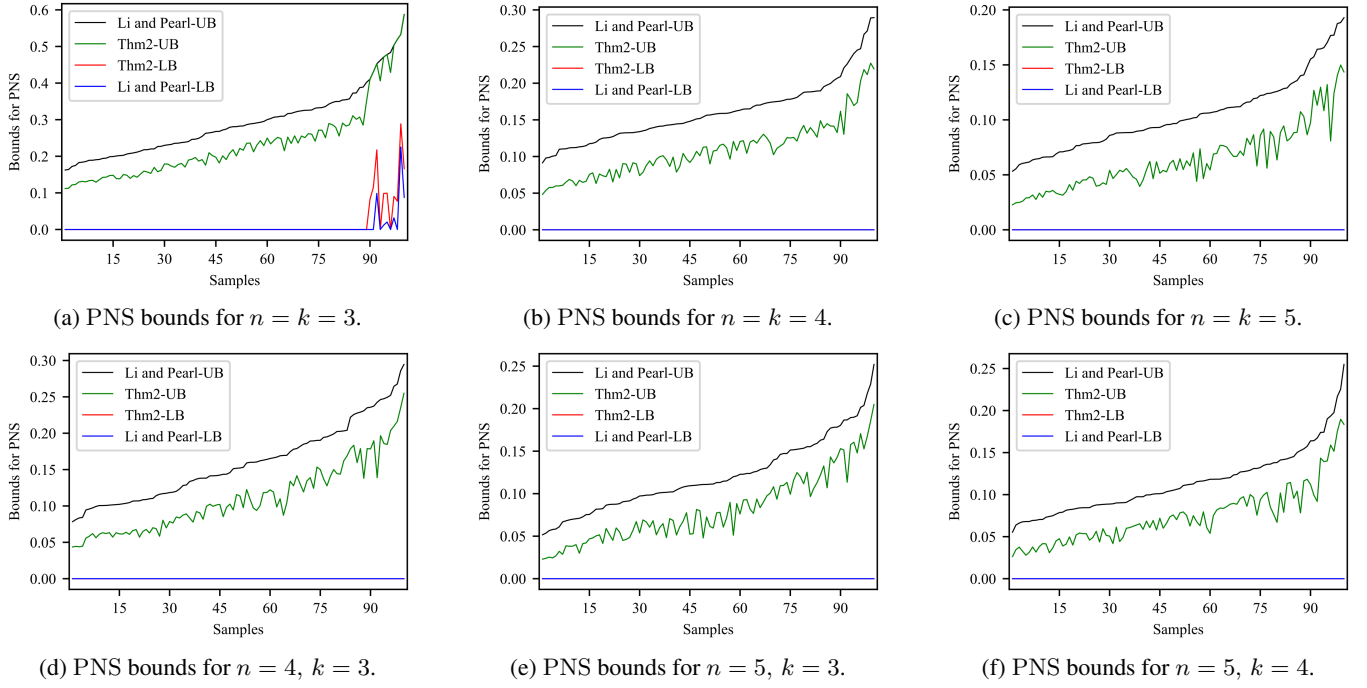

 Figure 3: Per-instance PNS bounds across different (n, k) settings.

Table 5, we further examine per-instance behavior to illustrate how bound tightening manifests at the level of individual samples across dimensions.

Figure 2 shows the lower and upper bounds for 100 representative instances with $n = k = 3$, comparing the bounds from Theorem 2 and from Li and Pearl [2024]. Extending this analysis, Fig. 3 reports results across multiple (n, k) settings. For each configuration, we generate 10,000 compatible sample distributions and select the 100 instances with the largest discrepancies between the two methods, ordered by their PNS bounds to facilitate comparison.

As dimensionality increases, the lower bounds increasingly concentrate near zero. This behavior is expected, since probability mass is spread over a larger number of categories, reducing typical values of terms such as $P(x_i, y_i)$ and $P(y_{i|x_i})$, and thereby limiting potential improvements in the lower bounds.

Importantly, this does not diminish the advantage of the proposed approach. Instead, with increasing n , the *relative* separation between the bounds produced by Theorem 2 and those of Li and Pearl becomes more pronounced, driven mainly by systematic reductions in the upper bounds. This trend is particularly evident in Fig. 3a–3c, where $n = k$ and the relative gap widens as dimensionality grows.

Moreover, Fig. 3d–3f show that this separation persists when $n \neq k$, indicating that the closed-form bounds remain accurate beyond the diagonal ($n = k$) regime. Overall, these results demonstrate that the proposed closed-form expressions are not only computationally simpler, but also consistently tight across a wide range of dimensional configurations.

6 Conclusion

We studied multi-valued probabilities of causation (PoCs) in general Structural Causal Models using experimental and observational distributions, without imposing structural restrictions. Our main contribution is a closed-form framework that reduces arbitrary discrete PoCs to a representative family with analytically derived bounds. Equivalence classes and replaceability transfer these bounds across value permutations and mismatched treatment–outcome cardinalities, yielding a compact toolkit for non-binary PoCs.

We established soundness in all dimensions and empirically verified tightness in low-dimensional settings via Balke’s linear programming when feasible. Simulations show that our bounds consistently tighten the recursive bounds of Li and Pearl [2024] while being simpler to compute and suitable for repeated queries. These results suggest that the proposed bounds may be complete in general, although a formal tightness proof for all dimensions remains open.

Our examples show that multi-valued PoCs capture cross-world response patterns that marginal comparisons cannot reveal. In both medical and educational settings, they quantify latent response types—such as individuals who benefit under one intervention but may be harmed under another—thereby supporting heterogeneity-aware and personalized decisions.

Several open directions remain. A general completeness proof is still open beyond the low-dimensional cases verified here. Future work may incorporate covariates and causal graphs to sharpen identification, and seek interpretable non-binary structural assumptions for stronger bounds. In particular, extending monotonicity-type ideas such as Monotonic Incremental Treatment Effect (MITE) [Mueller, 2025] under partial-order constraints is a promising direction.

References

- [Balke, 1995] Alexander Abraham Balke. *Probabilistic counterfactuals: semantics, computation, and applications*. University of California, Los Angeles, 1995.
- [Dawid *et al.*, 2017] A Philip Dawid, Monica Musio, and Rossella Murtas. The probability of causation. *Law, Probability and Risk*, 16(4):163–179, 2017.
- [de Aguas *et al.*, 2025] Johan de Aguas, Sebastian Krum-scheid, Johan Pensar, and Guido Biele. The probability of tiered benefit: Partial identification with robust and stable inference. *arXiv preprint arXiv:2502.10049*, 2025.
- [Galles and Pearl, 1998] David Galles and Judea Pearl. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3:151–182, 1998.
- [Halpern, 2000] Joseph Y Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.
- [Li and Pearl, 2019] Ang Li and Judea Pearl. Unit selection based on counterfactual logic. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019.
- [Li and Pearl, 2022a] Ang Li and Judea Pearl. Bounds on causal effects and application to high dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5773–5780, 2022.
- [Li and Pearl, 2022b] Ang Li and Judea Pearl. Unit selection with causal diagram. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 5765–5772, 2022.
- [Li and Pearl, 2024] 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.
- [Mueller and Pearl, 2022] Scott Mueller and Judea Pearl. Personalized decision making – a conceptual introduction. Technical Report R-513, Department of Computer Science, University of California, Los Angeles, CA, 2022.
- [Mueller *et al.*, 2021] Scott Mueller, Ang Li, and Judea Pearl. Causes of effects: Learning individual responses from population data. *arXiv preprint arXiv:2104.13730*, 2021.
- [Mueller, 2025] Scott Mueller. *Personalized and Situation-Specific Decision Making*. University of California, Los Angeles, 2025.
- [Pearl, 1999] Judea Pearl. Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese*, pages 93–149, 1999.
- [Tian and Pearl, 2000] Jin Tian and Judea Pearl. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28(1-4):287–313, 2000.
- [Zhang *et al.*, 2022] Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In *International Conference on Machine Learning*, pages 26548–26558. PMLR, 2022.