Causes of effects

Identifying individual effects from population-level data is a common but challenging scientific goal. Three variations of causes of effects are employed in practice: probability of necessity (PN), probability of sufficiency (PS), and probability of necessity and sufficiency (PNS). The latter, PNS, is the focus of this note.

Probability of necessity (PN) is the probability of no effect when no treatment is applied given a treated individual is affected.

$$PN = \Pr\left(y'_{x'} \mid y, x\right)$$

where y denotes a detected effect, y' denotes no detected effect, x denoted exposure, and x' denotes no exposure.

Probability of sufficienty (PS) is the probability that treatment causes the effect given an individual is unaffected when untreated.

$$PS = \Pr\left(y_x \mid y', x'\right)$$

Probability of necessity and sufficiency (PNS) is the probability of both effects[1]

$$\begin{gathered} PNS = \Pr\left(y_x, y'_{x'}\right) \\ PNS = PN \Pr(y, x) + PS \Pr\left(y', x'\right) \end{gathered}$$

That is, PNS is the probability that outcome y responds to exposure x both ways. Clearly, all three probabilities are counterfactual as we do not simultaneously observe an individual's response to both exposure conditions.

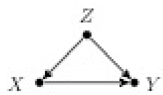
These probabilities are more demanding than other point-identified counterfactuals such as average treatment effects and accordingly are typically not point-identified. Tian and Pearl [2000] provide tight bounds for all three probabilities without using a causal graph.Tian and Pearl bounds for PNS are

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

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

[1] The second expression for PNS is lemma 1 in Pearl [1999]. Consistency dictates $\left(y_x \mid x\right) = y$ and $\left(y_{x'} \mid x'\right) = y$. Combining consistency with $\wedge$ (and) and $\vee$ (or) logic gives $\left(y_x \wedge y'_{x'}\right) = \left(y_x \wedge y'_{x'}\right) \wedge \left(x \vee x'\right) = \left(y \wedge x \wedge y'_{x'}\right) \vee \left(y_x \wedge y' \wedge x'\right)$. Applying probabilities to both sides completes the lemma. $\Pr\left(y_x, y'_{x'}\right) = \Pr\left(y_x, y', x'\right) + \Pr\left(y'_{x'}, y, x\right)$

Mueller, Li, and Pearl [2021] employ a causal graph typically producing tighter bounds. Given there exists a back-door variable Z,



the bounds for PNS are

$$\sum_z \max\{0, \Pr(y \mid x, z) - \Pr(y \mid x', z)\} \times \Pr(z) \leq PNS$$

–

$$PNS \leq \sum_z \min\{\Pr(y \mid x, z), \Pr(y' \mid x', z)\} \times \Pr(z)$$

Example

Suppose a struggling organization is approached by a professional manager claiming to be the better option for turning around the organization's prospects (better off with and worse off without the manager's leadership). The professional manager backs up the claim with data indicating a 68% turn around success rate for professional managers versus a 54% turn around success rate for employee owner-managers.

However, further scrutiny of the data reveals the following tabulated results regarding successful turn around where the state of the industry is deemed causal to (precedes) manager selection as well as impacting turn around success (in other words, state of the industry is a back-door).[2]

	professional manager	employee-owner manager
declining industry	1 of 110 (1%)	13 of 120 (11%)
rising industry	313 of 354 (88%)	114 of 116 (98%)
total	314 of 464 (68%)	127 of 236 (54%)

Let z indicate rising industry and z' indicate declining industry. Also, let x denote professional manager and x' denote employee-owner manager. Of course, y denotes successful turn around and y' denotes otherwise. Tian and Pearl PNS bounds without the use of the causal graph are fairly wide, $0 \leq PNS \leq 0.297$, where professional manager is the exposure variable of interest.[3]

Tian and Pearl solve a linear program to find the bounds. Define the joint distribution for Y, X, Y_x, and $Y_{x'}$ by letting (where the right hand expression

[2]Data is simulated, not drawn from actual sources.

[3]If the exposure variable is employee-owner manager, Tian and Pearl bounds are $0.099 \leq PNS \leq 0.396$.

follows from consistency)

$$\begin{aligned}
p_{111} &= \Pr(y_x, y_{x'}, x) = \Pr(x, y, y_{x'}) \\
p_{110} &= \Pr(y_x, y_{x'}, x') = \Pr(x', y, y_x) \\
p_{101} &= \Pr(y_x, y'_{x'}, x) = \Pr(x, y, y'_{x'}) \\
p_{100} &= \Pr(y_x, y'_{x'}, x') = \Pr(x', y', y_x) \\
p_{011} &= \Pr(y'_x, y_{x'}, x) = \Pr(x, y', y_{x'}) \\
p_{010} &= \Pr(y'_x, y_{x'}, x') = \Pr(x', y, y'_x) \\
p_{001} &= \Pr(y'_x, y'_{x'}, x) = \Pr(x, y', y'_{x'}) \\
p_{000} &= \Pr(y'_x, y'_{x'}, x') = \Pr(x', y', y'_x)
\end{aligned}$$

Now, the minimization (maximization) linear program is

$$\min_{p_{ijk} \geq 0} (\max) PNS = p_{101} + p_{100}$$

$$s.t.$$

$$\begin{aligned}
\Pr(y_x) = p_{111} + p_{110} + p_{101} + p_{100} &= \frac{1}{110} * \frac{230}{700} + \frac{313}{354} * \frac{470}{700} \\
\Pr(y_{x'}) = p_{111} + p_{110} + p_{011} + p_{010} &= \frac{13}{120} * \frac{230}{700} + \frac{114}{116} * \frac{470}{700} \\
\Pr(x) = p_{111} + p_{101} + p_{011} + p_{001} &= \frac{464}{700} \\
\Pr(y, x) = p_{111} + p_{101} &= \frac{341}{464} * \frac{464}{700} \\
\Pr(y', x) = p_{011} + p_{001} &= \frac{150}{464} * \frac{464}{700} \\
\Pr(y, x') = p_{110} + p_{010} &= \frac{127}{236} * \frac{236}{700} \\
\Pr(y', x') = p_{100} + p_{000} &= \frac{109}{236} * \frac{236}{700}
\end{aligned}$$

However, employing the causal structure of the graph and the Mueller et al bounds allows considerable tightening of the bounds, $0 \leq PNS \leq 0.015$. Also, solving the CSM linear program for this example produces even further tightening of the bounds, $0.002 \leq PNS \leq 0.013$.[4] This provides little support (around one percent) to the professional manager's claims that a professional manager produces successful turn around and an employee-owner management team does not.

On the other hand, focus on the employee-owner manager provides a little more support, $0.099 \leq PNS \leq 0.113$ for the Mueller et al bounds, and $0.100 \leq PNS \leq 0.112$ for the linear program applied to this example. The probability that an employee-owner management team produces successful and a professional manager does not produce a turn around is about ten percent.

[4]See the appendix for details of the CSM (causal structural model) linear program.

It's instructive to compare PNS with population-level causal effects (based on the back-door adjustment). $\Pr(y_x) = 0.597$ while $\Pr(y_{x'}) = 0.695$ where x refers to the professional manager and x' refers to employee-owner manager. The difference is approximately ten percent in favor employee-owner management.

Since the state of the industry is frequently known before managerial selection, it may be more instructive to check the bounds on conditional PNS. Conditional PNS is the joint probability that management team x produces a turn around and management team x' fails to turn around the organization's prospects in industry state z', or alternatively, in industry state z. When x refers to a professional manager the conditional bounds are $0 \leq PNS(z') \leq 0.009$ and $0 \leq PNS(z) \leq 0.017$.

On the other hand, when x refers to a employee-owner manager the conditional bounds are $0.099 \leq PNS(z') \leq 0.108$ and $0.099 \leq PNS(z) \leq 0.116$. The PNS evidence conditional on state of the industry again favors employee-owner management to professional management (around ten percent to one percent).

Again, we can compare conditional PNS with conditional causal effects. $\Pr(y_x \mid z') = 0.091$, $\Pr(y_x \mid z) = 0.884$, $\Pr(y_{x'} \mid z') = 0.108$, and $\Pr y_{x'} \mid z) = 0.983$ where x refers to the professional manager, x' refers to the employee-owner manger, z refers to rising industry conditions, and z' refers to declining industry conditions. The difference ranges from approximately one percent for declining industry conditions to ten percent for rising industry conditions in support of employee-owner management.

Appendix

The SCM linear program expands Tian and Pearl by adding the back-door Z to the joint distribution to form p_{ijkl} where l refers to the value of Z. With this addition, the objective function is $PNS = p_{1011} + p_{1001} + p_{1010} + p_{1000}$. The added constraints include

$$\Pr(z) = \sum_{i,j,k} p_{ijk1}$$
$$\Pr(y, x, z) = p_{1111} + p_{1011}$$
$$\Pr(y, x, z') = p_{1110} + p_{1010}$$
$$\Pr(y, x', z) = p_{1101} + p_{0101}$$
$$\Pr(y, x', z') = p_{1100} + p_{0100}$$
$$\Pr(y \mid x, z) = (p_{1111} + p_{1011}) / \Pr(x, z)$$
$$\Pr(y \mid x, z') = (p_{1110} + p_{1010}) / \Pr(x, z')$$
$$\Pr(y \mid x', z) = (p_{1101} + p_{0101}) / \Pr(x', z)$$
$$\Pr(y \mid x', z') = (p_{1100} + p_{0100}) / \Pr(x', z')$$

As Z is a back-door, the latter four constrains are equivalent to $\Pr(y \mid do(X), Z)$ by rule 2 of do-calculus and employed in the back-door adjustment.