TECHNICAL REPORT
R-439
SEPTEMBER 2014

A Note on Causes of Effects

Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
(310) 825-3243 / judea@cs.ucla.edu

September 29, 2014

1 Background

Interest in applying counterfactual logic to legal settings has resulted in disagreements regarding the proper interpretation of the legal term "but for," as in "It is more probable than not that the injury would not have occurred *but for* the defendant action" (Robertson, 1997). Let $A = 1$ stands for the defendant's action, $R = 1$ for the observed response (e.g., injury or damage), and R_0 (respectively R_1) for the value that R would have had the action not taken ($A = 0$). The standard interpretation of the "but for" criterion is captured by the inequality $PN \geq \frac{1}{2}$ where PN stands for counterfactual probably

$$PN = P(R_0 = 0 | A = 1, R = 1) \tag{1}$$

termed "probability of necessity" in Pearl (2000a). The same interpretation was used by Greenland and Robins (1988); Balke and Pearl (1994a,b); Pearl (1999, 2009a); and Tian and Pearl (2000). Equation (1) is a direct translation of the "but for" test into counterfactual language, saying that R would not have occurred in the absence of A, given that R and A did in fact occur. Implicit in PN is the understanding that the probability P is defined relative to a reference class of individuals who are exchangeable with the defendant. In other words, P embeds all other information we have about the incident, for example, that the defendant is a red hair lawyers who owns a black Mercedes, and that the claimant was a reckless driver.

Ironically, Eq. (1) was also used in Pearl (2000b)to demonstrate that counterfactuals can handle CoE-type questions, contrasting Dawid's dismissal of counterfactuals as "metaphysical" concepts that "can lead to distorted understandings and undesirable practical consequences" (Dawid, 2000, p. 408). "I challenge Dawid to express Query II ["My headache has gone. Is it because I took aspirin?"], let alone formulate conditions for its estimation in a counterfactual-free language (Pearl, 2000b, p. 429).

2 The Mystification of "CoE"

In a recent article, Dawid et al. (2014a) (henceforth DFF) urge statisticians to pay attention to "Causes of Effects" (PN), in contrast to "Effects of Causes" (CoE), especially to the unique and challenging problems that CoE presents in legal settings, where individual cases are the focus of deliberation, and population data rarely provide sufficient evidence. Oddly, instead of PN, DFF proposed another counterfactual expression for CoE, termed "probability of causation":

$$PC = P(R_0 = 0|R_1 = 1) \tag{2}$$

which in our context reads: The probability that an injury would not occur had an action (like $A = 1$) not been taken, given that an injury would occur had that action been taken.

Clearly, PC is the wrong measure to use in CoE problems, as can be seen from fact that the reference class $R_1 = 1$ does not entail that the injury $R = 1$ actually occurred, or that the action $A = 1$ actually took place. As a result of this mistaken reference class, PC fails to properly represent the evidence typically available in litigation cases (Pearl, 2014, footnote 5). It represents the probability that a response obtained under experimental regime $A = 1$ would not occur had the regime been changed to $A = 0$. It does not take into account the fact that, in litigation cases, actions are typically executed by choice, by actors who have the capabilities to implement those actions and reasons to expect their consequences. Such actors constitute a select subpopulation that are not distinguished by PC.

DFF go to a great length describing the difficulty of comprehending the meaning of PC, and of estimating, or bounding it from empirical data, unless one is wiling to make strong, untestable assumptions. Perhaps it was this sort of difficulties that led DFF to conclude that the problem requires "an alternative framing of the 'CoE' that differs substantially from that found in the bulk of the scientific literature" (Dawid et al., 2014a, p. 359).

Undeterred by the semantic inadequacy of PC, DFF showed that a lower bound to PC can be estimated if one is willing to assume "exogeneity," or "non-confoundedness," namely, that the defendant chose his action "as if at random," independent of any factor that may affect the response R. Under such assumption DFF show that PC can be lower bounded by the *Excess Risk Ratio* (ERR):

$$ERR = 1 - P(R = 1|A = 0)/P(R = 1|A = 1) \leq PC$$

Readers will recognize the inequality $ERR > \frac{1}{2}$ as the standard criterion used by epidemiologists to meet the "more probable than not" test in court cases (Schlesselman, 1982; Greenland and Robins, 1988 Pearl, 2000a, p. 292). ERR is also known to be a lower bound to PN when exogeneity holds (Greenland and Robins, 1988) and it was refined in (Tian and Pearl, 2000) to allow for confounding.

In a discussion following DFF's paper, Nicholas Jewell alerted the authors to the "more relevant" interpretation of "but for" in terms of PN, to the extensive work done on CoE under this and other interpretations and, in particular, to the tight bounds derived by Tian and Pearl (2000) under a variety of assumptions, using both observational and experimental data (Jewell, 2014). In their rejoinder (Dawid et al., 2014b), DFF explained their choice of the PC measure in these words:

> Jewell notes the close connection with earlier work of Robins and Greenland (1989; Greenland and Robins, 2000), and of Pearl and his collaborator Tian (Pearl, 2009b; Tian and Pearl, 2000). We were aware of this work, having referenced it in earlier articles, and were remiss in not including discussion of it here. Robins and Greenland, using different notation and statistical formalisms, focus on what we and they call the PC although without the potential outcome labels, and they present the same lower bound, which come from the standard Fréchet bounds for 2×2 tables. They also address the assigned shares approach to interpreting the role of the relative risk used by the courts to address the CoE.
>
> Jewell suggests that we should have focused on $P(R_0 = 1|R_1 = 1 \text{ and } A = 1)$ where A denotes the observed exposure condition–which is Pearl's Probability of Necessity (PN).This was in fact the way in which the CoE problem was initially formulated by Dawid (2011), the simplification to $Pr(R_0 = 1|R_1 = 1)$ being based on the "(questionable) assumption that the decision to take aspirin was unrelated to the (then hidden) values of the potential responses." Now this additional assumption is unreasonable unless the joint probability distribution being manipulated can be regarded as that fully specific to the given individual; and, to the extent that knowledge of this individual distribution is informed by EoC-type data, it will be essential that probabilities estimated from these data are computed relative to a suitably refined reference class. Without this requirement, focusing on bounds for $P(R_0 = 1|R_1 = 1 \text{ and } A = 1)$ will not be the right thing to do.
>
> We also note that the difference in the condition for our PC and Pearl's is what led to the upper bound in Pearl's work with Tian, which is not necessarily 1 for PN. Moreover, the work of Pearl and others to sharpen these bounds and to identify PN rests on heroic assumptions that we deem inappropriate for the present discussion, especially when they ignore the distinctions between populations and samples, and observational and experimental data. Dawid et al. (2014b) do provide a more general treatment than the one we do in our article, which does allow for an upper bound that can differ from 1, but again it differs from that of Tian and Pearl for the reasons given previously.

These paragraphs are laden with inexplicable oversights. I will first list these oversights, and then trace their origin to the paper by Dawid et al. (2014b) (henceforth DMF), which DFF cite as a "more general treatment" of the problem. Finally, I will summarize the features of the CoE problem that were missed by DFF and DMF and show how restoring these features can improve our ability to discern the "more probable than not" criterion. The latter is based on (Pearl, 2014).

3 List of Puzzles and Oversights in DFF

1. DFF attempt to repair the inadequacy of PC by strong assumptions fails to distinguish "definition" from "identification." Definitions should capture the intent of the research question universally, over all models; they should not change with assumptions about

one scenario or another. In our context, the defining expression should faithfully represent the "but for" criterion in all models, regardless of whether confounding is present in the model or not. The "simplification" that DFF made in using PC (instead of PN) does not represent the "but for" criterion and, therefore, cannot be justified by any model-specific assumption, including the one made my DFF, that A and (R_1, R_0) are independent.

2. If PC was an unfortunate "simplification" based on "questionable assumptions," then

$$P(R_0 = 1 | R_1 = 1 \text{ and } A = 1)$$

should be free of those assumptions and deemed the proper parameter to focus on. Why then would "focusing on bounds for $P(R_0 = 1|R_1 = 1$ and $A = 1)$ not be 'the right thing to do'?" Once we decide on the right parameter to focus on, we should derive all the information we can get from it.

However, while DFF deem many parameters "improper," they do not tell us what the proper parameter is that we should focus on. Once we commit to the proper parameter, we must also commit to the proposition that, if the lower bound for that parameter exceeds 50% the "more probable than not" criterion would be satisfied. This, of course would mean that CoE problems can be solved by standard counterfactual logic, and do not require "an alternative framework of the 'CoE' that differs substantially from that found in the bulk of the scientific literature" as DFF state in their abstract.

3. DFF's assertion that "the work of Pearl and others rests on heroic assumptions" does not sit well with the facts. Their basis for the assertion reads: "they [Pearl and others] ignore the distinction between population and samples, and observational and experimental data." The facts tell a different story.

   3a. Pearl consistently separates population aspects of the CoE problem from its sample aspects. DFF are using the same separation, which is a wise move; the two subproblems deserve separate treatments.

   3b. Tian and Pearl's work rests on combining observational and experimental data. They distinguish between them chapter and verse; in mathematical notation, in verbal description, in examples, and in logic. It is hard to imagine a more incisive and colorful distinction anywhere in the statistical literature (see also Pearl, 2000a, Ch. 9).

   3c. The assumptions we make for bounding PN are in fact milder than those made by DFF (2014a), as well as by DMF (2014b). For example, we do not assume "no-confounding," which DFF assume.

   3d. To say that Tian and Pearl analysis rests on "heroic assumptions [that are] inappropriate for the present discussion" is like saying that CoE analysis is inappropriate for CoE analysis. Indeed, DMF have embraced these same assumptions by adopting PN instead of PC.

4. DFF confound "generality" with "appropriateness." The analysis of DMF (2014b) is not "more general" than the one done by DFF; it merely corrects the research question,

and brings it to the fold of standard CoE analysis. In other words, DMF discard the parameter

$$PC = P(R_0 = 0|R_1 = 1) \tag{3}$$

and replaces it with the appropriate parameter

$$PC_A = P(R_0 = 0|H, A = 1, R_1 = 1) \tag{4}$$

which is equivalent to

$$PN = P(R_0 = 0|H, A = 1, R = 1) \tag{5}$$

(H stands for "any other information we have about the episode, and is implicit in PN). This follows from the consistency rules

$$(A = 1) \text{ and } (R_1 = 1) \implies (R = 1) \tag{6}$$

and

$$(A = 1) \text{ and } (R_1 = 0) \implies (R = 0) \tag{7}$$

By restoring the analysis to the PN fold, DMF recaptured the "but for" criterion and should have been able to obtain the bounds of Tian and Pearl (2000). Unfortunately, the syntactic transformation from PN to PC_A led DMF to make unnecessarily strong assumptions and to miss the more informative bounds that were derived in Tian and Pearl (2000).

4 How Opportunities Were Missed?

DMF formally define the PROBABILITY OF CAUSATION as the conditional probability:

$$PC_A = P_A(R_0 = 0|H, A = 1, R_1 = 1)$$

where P_A denotes the subjective probability distribution of attributes of the actor or decision maker. As we discussed earlier, this expression is identical to PN, but differs from it in syntactic form; the conditioning event contains $R_1 = 1$, instead of $R = 1$. This led DMF to conclude that PC_A "involves a joint distribution of (R_0, R_1)," which is non-estimable from either observational or experimental data. Accordingly, DMF derived upper and lower bounds in terms of the counterfactual parameter $P_A(R_0|H, A = 1)$ which is also non-estimable from observational or experimental data without further assumptions.[1] DMF therefore made the unnecessary assumption of "strong ignorability" (also called "sufficiency").

$$(R_0, R_1) \perp\!\!\!\perp E|H$$

[1]This parameter belongs to the ETT variety (the effect of treatment on the treatment), which cannot be identified from experimental data alone (See (Pearl, 2009b, pp. 396–397; Shpitser and Pearl, 2009) for complete identification conditions.)

which permitted them to finally express the lower bound (of PN) in terms of an estimable parameter, the *causal risk ratio*

$$RR_A = P_A(R_1|H)/P_A(R_0|H)$$

In comparison, the bounds obtained by Tian and Pearl enjoy the following properties:

1. Ignorability (or sufficiency) is not assumed.
2. PN is lower bounded by an observational parameter, ERR, and an experimental parameter
$$CF = [P(R = 1|A = 0) - P(R_0 = 1)]/P(R = 1, A = 1),$$
which accounts for possible confounding.
3. The only parameter that comes from experimental data is $P(R_0 = 1)$; $P(R_1 = 1)$ need not be estimated.
4. The lower bound may be improved by confounding, whenever $CF > 0$.
5. The upper bound can be reduced by confounding, whenever $CF < 0$.
6. Regardless of confounding, the gap between the upper and lower bounds is given by one observational parameter, $P(A = 0)/P(A = 1)$.
7. When R is monotonic with A, PN is identifiable from observational data.
8. These bounds are tight, i.e., they cannot be improved without strengthening the assumptions.
9. Contrary to prevailing lore, these bounds do not require knowledge of the data-generating model; population data from observational and experimental studies are all that is needed.

Vivid illustratioins of how the PN bounds vary with observational and experimental parameter are given in Pearl (2014).

Conclusions

I fail to understand why Dawid, Faigman and Fienberg would not embrace a mathematical analysis of Causes of Effects that is based on weaker assumptions and yields more meaningful and informative conclusions than any of those reported in the literature.

Acknowledgment

This research was supported in parts by grants from NIH #1R01 LM009961-01, NSF #IIS-0914211 and #IIS-1018922, and ONR #N000-14-09-1-0665 and #N00014-10-1-0933.

References

Balke, A. and Pearl, J. (1994a). Counterfactual probabilities: Computational methods, bounds, and applications. In *Uncertainty in Artificial Intelligence 10* (R. L. de Mantaras and D. Poole, eds.). Morgan Kaufmann, San Mateo, CA, 46–54.

Balke, A. and Pearl, J. (1994b). Probabilistic evaluation of counterfactual queries. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, vol. I. MIT Press, Menlo Park, CA, 230–237.

Dawid, A. (2000). Causal inference without counterfactuals (with comments and rejoinder). *Journal of the American Statistical Association* **95** 407–448.

Dawid, A. (2011). The role of scientific and statistical evidence in assessing causality. In *Perspectives on Causation* (R. Goldberg, ed.). Hart Publishing, Oxford, England, 133–147.

Dawid, A., Fienberg, S. and Faigman, D. (2014a). Fitting science into legal contexts: Assessing effects of causes or causes of effects? *Sociological Methods and Research* **43** 359–390.

Dawid, A., Musio, M. and Fienberg, S. (2014b). From statistical evidence to evidence of causality. Tech. rep., Statistical Laboratory, University of Cambridge, UK. Submitted to *Bayesian Analysis*. ArXiv: 1311.7513.

Greenland, S. and Robins, J. (1988). Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology* **128** 1185–1197.

Greenland, S. and Robins, J. (2000). Epidemiology, justice, and the probability of causation. *Jurimetrics* **40** 321–340.

Jewell, N. P. (2014). Assessing causes for individuals: Comments on Dawid, Faigman, and Fienberg. *Sociological Methods and Research* **54** 391–395.

Pearl, J. (1999). Probabilities of causation: Three counterfactual interpretations and their identification. *Synthese* **121** 93–149.

Pearl, J. (2000a). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. 2nd edition, 2009.

Pearl, J. (2000b). Comment on A.P. Dawid's, Causal inference without counterfactuals. *Journal of the American Statistical Association* **95** 428–431.

Pearl, J. (2009a). Causal inference in statistics: An overview. *Statistics Surveys* **3** 96–146.

Pearl, J. (2009b). *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge University Press, New York.

Pearl, J. (2014). Causes of effects and effects of causes. Tech. Rep. R-431-L, <http://ftp.cs.ucla.edu/pub/stat_ser/r431-L.pdf>, Department of Computer Science, University of California, Los Angeles, CA. Short version forthcoming, *Journal of Sociological Methods and Research*.

Robertson, D. (1997). The common sense of cause in fact. *Texas Law Review* **75** 1765–1800.

Robins, J. and Greenland, S. (1989). The probability of causation under a stochastic model for individual risk. *Biometrics* **45** 1125–1138.

Schlesselman, J. (1982). *Case-Control Studies: Design Conduct Analysis*. Oxford University Press, New York.

Shpitser, I. and Pearl, J. (2009). Effects of treatment on the treated: Identification and generalization. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, Montreal, Quebec, 514–521.

Tian, J. and Pearl, J. (2000). Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence* **28** 287–313.