

1-1-2015

# On the Causes of Effects Response to Pearl

David Faigman

UC Hastings College of the Law, faigmand@uchastings.edu

Philip A. Dawid

Steven E. Feinberg

Follow this and additional works at: [http://repository.uchastings.edu/faculty\\_scholarship](http://repository.uchastings.edu/faculty_scholarship)

---

## Recommended Citation

David Faigman, Philip A. Dawid, and Steven E. Feinberg, *On the Causes of Effects Response to Pearl*, 44 *Soc. Methods & Res.* 165 (2015).  
Available at: [http://repository.uchastings.edu/faculty\\_scholarship/1304](http://repository.uchastings.edu/faculty_scholarship/1304)

This Article is brought to you for free and open access by UC Hastings Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of UC Hastings Scholarship Repository. For more information, please contact [marcusc@uchastings.edu](mailto:marcusc@uchastings.edu).

# On the Causes of Effects: Response to Pearl

Sociological Methods & Research

2015, Vol. 44(1) 165-174

© The Author(s) 2014

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0049124114562613

smr.sagepub.com



**A. Philip Dawid<sup>1</sup>, David L. Faigman<sup>2</sup>,  
and Stephen E. Fienberg<sup>3</sup>**

## Abstract

We welcome Professor Pearl's comment on our original article, Dawid et al. Our focus there on the distinction between the "Effects of Causes" (EoC) and the "Causes of Effects" (CoE) concerned two fundamental problems, one a theoretical challenge in statistics and the other a practical challenge for trial courts. In this response, we seek to accomplish several things. First, using Pearl's own notation, we attempt to clarify the similarities and differences between his technical approach and that in Dawid et al. Second, we consider the more practical challenges for CoE in the trial court setting, and explain why we believe Pearl's analyses, as described via his example, fail to address these. Finally, we offer some concluding remarks.

## Keywords

effects of causes, causes of effects, law, evidence, causation

---

<sup>1</sup> Centre for Mathematical Sciences, Cambridge University, Cambridge, UK

<sup>2</sup> Hastings College of the Law, University of California, San Francisco, CA, USA

<sup>3</sup> Department of Statistics, Heinz College, Machine Learning Department, and Cylab, Carnegie Mellon University, Pittsburgh, PA, USA

## Corresponding Author:

David L. Faigman, Hastings College of the Law, University of California; School of Medicine, University of California San Francisco; UCSF/UC Hastings Consortium on Law, Science & Health Policy, San Francisco, CA, USA.

Email: faigmand@uchastings.edu

## Introduction

Our focus in Dawid, Faigman, and Fienberg (2014) on the distinction between the “Effects of Causes” (EoC) and the “Causes of Effects” (CoE) concerned two fundamental problems, one a theoretical challenge in statistics and the other a practical challenge for trial courts. We welcome Judea Pearl’s valuable reminder (Pearl 2015) of the analysis by Tian and Pearl (2000) of the problem of making inferences about CoE.

Pearl believes he has solved the technical problem we posed, and expects his solution to be adopted by trial courts. He assures us that analysis of CoE “has not lagged behind that of EoC.” He is confident that, using his approach, “[b]oth modes of reasoning enjoy a solid mathematical basis, endowed with powerful tools of analysis, and researchers on both fronts now possess solid understanding of applications, identification conditions, and estimation techniques.”

In the second section, using Pearl’s own notation, we attempt to clarify the similarities and differences between his technical approach and that in Dawid et al. (2013). In the third section, we turn to the more practical challenges for CoE in the trial court setting, and explain why we believe Pearl’s analyses, as described via his example, fail to address these. The fourth section offers some concluding remarks.

## Technical Details of CoE

### Comparison

For expository purposes, the analysis in our article was based on some simplifying, albeit implicit, assumptions:

1. Exchangeability: The pair of potential responses  $Y = (Y_0, Y_1)$  for the new individual ( $i$ ) can be regarded as exchangeable with such pairs in the group ( $G_0$ ) on which we have experimental data.
2. No confounding: For  $i$ , and for any member of  $G_0$ , the pair  $Y = (Y_0, Y_1)$  of potential responses to treatment is independent of the treatment actually applied.

These conditions were made explicit, and carefully discussed and criticized, in the articles Dawid (2011) and Dawid, Musio, and Fienberg (2014) referenced in our article.

Pearl, drawing on Tian and Pearl (2000), retains (again implicitly) our Assumption 1, but drops Assumption 2. This vitiates our own simple

analysis, since it is now necessary to take further account of the fact that  $i$  chose to take the treatment. Our approach in such a case (see Dawid et al. 2014) was to assume that there is an observable covariate which is “sufficient,” in the sense that, conditional on it, the assumptions become valid—then our analysis can be localized accordingly. Pearl, on the other hand, supposes we have further observational data (or equivalent information) on a group  $G_1$  of individuals who can be regarded as fully exchangeable with  $i$ —including in the way their treatment was assigned. His new bounds apply to this situation. This is certainly a valuable extension of our analysis when its conditions are satisfied. We caution, however, that it may not be easy to obtain experimental and observational data that can *both* be regarded as exchangeable with the individual  $i$  under current consideration.

To link these two approaches, we introduce a variable  $D$  to represent the treatment that an individual desires. We extend the exchangeability Assumption 1 to require that the triples  $(D, Y_0, Y_1)$ , for all individuals, can be regarded as arising from the same joint distribution, and further that there is no additional relevant information available on these individuals.

In purely observational circumstances, there is no constraint preventing the desire for treatment from being realized, and so the actual treatment  $X$  will be identical with the desired treatment  $D$  (in particular,  $D$  is observable in these circumstances). In an experimental setting, however, the desire  $D$  will be overridden by the externally imposed treatment, and we will typically not be in a position to observe  $D$ . We can expect  $D$  to be associated with  $Y$ ; but  $D$  should behave as a sufficient covariate, since the (probabilistic) response behavior of an individual in the experiment who was assigned active treatment  $X = 1$  and also had  $D = 1$  (and thus would have taken the treatment anyway) should be the same as that of an individual in the observational setting having  $X = 1$  (and thus also  $D = 1$ ).

What is somewhat surprising about the Tian and Pearl (2000) analysis of this situation is that we can make use of the special sufficient covariate  $D$  even though we can observe it only in  $i$  and the observational group  $G_0$ , and not in the experimental group  $G_1$ .

We compare and contrast this consequence of Pearl’s approach with the analysis in Dawid (2011, appendix)—see also Dawid (2000, section 11.1)—which shows how (in a setting where both Assumptions 1 and 2 do hold) it is possible to get additional information from a covariate that is observed in the experimental group  $G_1$ , but not observed for the current individual  $i$ . For example, with  $X$ ,  $Y$ , and covariate  $S$  all binary, we might learn, from the experimental data:

$$\begin{aligned}
 P(S = 1) &= 0.50 \\
 P(Y = 1 | X = 1, S = 1) &= 0.60 \\
 P(Y = 1 | X = 0, S = 1) &= 0 \\
 P(Y = 1 | X = 1, S = 0) &= 0 \\
 P(Y = 1 | X = 0, S = 0) &= 0.24
 \end{aligned}$$

Since, for  $i$ , we know  $X_i = 1$  and  $Y_i = 1$ , we can deduce that we must have  $S_i = 1$ , and so  $Y_i(0) = 0$ . Hence,  $PN = 1$ . If we had not measured  $S$  in the data, we would only have discovered  $PN \geq 1 - 1/((.30)/.12) = .60$ .

### Frameworks

In his second section, Pearl reminds us of his favored framework for modeling statistical causality, which is based on fully deterministic functional relationships between variables, stochasticity being introduced only through (typically unobserved) “background” variables  $U$ —implicitly supposed unaffected by interventions (factual or counterfactual) in the system. It is therefore interesting to note that none of this machinery is used in the analysis of his third section, which requires only the assumption of the simultaneous existence (and imperviousness to intervention) of the potential responses ( $Y_0, Y_1$ ). In a straightforward model such as Pearl has used here, there is indeed—as he points out—a formal isomorphism between these two different frameworks: Pearl’s equation (1) translates from a functional model to a potential response model, while the reverse translation is effected by defining  $U = (Y_0, Y_1)$ , and  $\mathbf{Y} = (X, U)$ , where  $f(x, (y_0, y_1))$  is taken to be  $y_x$ . However, such a definition of the “background variable”  $U$  itself begs many interpretive and philosophical questions.

Our own analysis was likewise based on the assumption of the simultaneous existence of  $Y_0$  and  $Y_1$ —though deliberately making no assumptions (such as monotonicity) about their joint behavior that are untestable in the factual world. Nonetheless, we confess to being less than fully comfortable with even this minimalist framework. An alternative framework for conducting counterfactual reasoning, avoiding direct or indirect assumptions of the simultaneous existence of incompatible potential responses, was presented in Part III of Dawid (2000)—this has some similarities with Pearl’s three-step computation in his second section, but replaces the assumption of functional determinism, required for step 3, with an assumption of conditional independence across parallel worlds. We hope to revisit the problem of “causes of effects” from this standpoint in due course (though we conjecture that this will not yield any new bounds for *Probability of Necessity* [PN]).

## *Effect of Treatment on the Treated*

In his footnote 4, Pearl mentions the problem of estimating the “effect of treatment on the treated (ETT)”. There are both similarities and differences with the estimation of PN. Comparing Pearl’s definition of ETT with that of PN in his equation (2), the only difference is the additional conditioning, in the latter, on the observed outcome,  $Y = y$ . This difference is, however, utterly crucial—in its absence, we can define and interpret ETT without any recourse to counterfactual reasoning by introducing again the “desire” variable  $D$ , so that  $X = D$  in an observational, but not in an experimental, setting.<sup>1</sup> Then we can define ETT as:

$$\text{ETT} = P(Y = 1|X = 1, D = 1) - P(Y = 1|X = 0, D = 1), \quad (1)$$

where  $X$  is supposed set by external intervention. Since the desired treatment  $D$  is a pretreatment variable, now uncoupled from the actually applied treatment  $X$ , this makes perfect sense in the actual world, with no need for any counterfactual considerations: It is the effect of switching from Treatment 0 to Treatment 1, for an individual who would, if unconstrained, choose to take the treatment. It can be shown that this definition agrees with that of Pearl and that it can be identified from suitable experimental and observational studies. See Geneletti and Dawid (2011) for a fuller analysis.

## **Fitting CoE Into the Legal Context**

Our article focused on the distinction between the “Effects of Causes” (EoC) and the “Causes of Effects” (CoE), with special emphasis on the latter as a practical challenge for trial courts. Pearl believes he has solved these problems and that his solution will be adopted by the courts—which, he concludes, can simply get on with tackling the business of CoE using his “solid mathematical [and logical] basis,” since “the logical gulf between [EoC and CoE] is no longer a hindrance to systematic analysis.” However, even if we accept Pearl’s assurances that the gulf between EoC and CoE “has been bridged by the structural semantics of counterfactuals,” his statistical analysis is far removed from the practicalities of the courtroom, leaving a yawning gap between his theoretical solution and the Law’s practical needs.

The EoC to CoE issue involves what is increasingly referred to in the courts as the “G2i problem”: that is, reasoning from group data to individual cases (Faigman, Monahan, and Slobogin 2014). Research data are introduced in court in a wide variety of legal contexts (e.g., civil litigation, criminal prosecutions, sentencing hearings, and civil commitment proceedings),

involving a score of scientific disciplines, ranging from aerodynamics to zoology. This research typically arrives initially as group data involving general phenomena, such as the factors associated with inaccurate eyewitness identifications, or whether benzene exposure is associated with leukemia. In the courtroom, the usual focus is on the individual case. Courts need to determine whether a particular case is an instance of some general phenomenon. Was the bank teller's identification of the gun-wielding perpetrator accurate? Did benzene cause the plaintiff's leukemia?

Pearl's treatment of CoE provides little guidance on the real issues courts face with G2i. He (1) misunderstands how the burden of proof is operationalized in legal settings, (2) fails to appreciate the complexity of the legal questions presented by CoE issues, and (3) is inattentive to legal decision makers' unfamiliarity with basic mathematical reasoning. Owing to space limitations, we can only consider these three problems briefly; but in terms of developing any practical solutions to reasoning from EoC to CoE in court, these and others are likely to pose profound, if not insurmountable, challenges to both statisticians and legal scholars.

First, Pearl's use of the legal burden of proof in civil cases is misleading and simplistic. In particular, he aligns the CoE statistical issue with the trier of fact's (jury or judge) ultimate determination regarding causation. He notes that the court's task in his example is to determine "whether it is 'more probable than not' that drug x was in fact the cause of Mr. A's death." Yet in the sixth section, he makes clear that his statistical analysis does not, indeed cannot, take into account "all the anatomical and psychological variables that determine an individual's behavior, and, even if we knew, we would not be able to represent them in the crude categories provided by the distribution at hand." But it is exactly these variables that jurors will be considering alongside any statistical proof of CoE. Pearl is wrong in believing that "[b]y using the wording 'more probable than not,' lawmakers have instructed us to ignore specific features for which data is not available, and to base our determination on the most specific features for which reliable data is available."

It may be that experts should be limited to "the most specific features for which reliable data is available," but it rarely operates that way in practice. In any case, this proscription does not apply to legal decision makers who are charged with applying the "more probable than not" standard. Indeed, this simplification to the point of caricature is reminiscent of the famous "Blue Bus Company" hypothetical (see, e.g., Fienberg 1986; Nesson 1985; Tribe 1971), which asks whether a plaintiff can recover for damages in an accident with a bus when the Blue Bus Company operates 80 percent of the buses on

the street where the accident occurred, but no other evidence is proffered by either side. Legal scholars generally agree that the proportion of blue buses is relevant evidence (and likely admissible), but it is not sufficient to sustain a verdict for the plaintiff (see Cheng 2013). The courts do expect to hear about “all the anatomical and psychological variables that determine an individual’s (or bus’s) behavior.” In the absence of such proof—much of it likely to be descriptive and anecdotal—the plaintiff’s claim will be dismissed.

Pearl’s second problem is his failure to appreciate the complexity of the legal questions presented by CoE issues. He applies his “solution” to “the most specific features for which reliable data is available.” He concedes, however, that more might be needed, though he ultimately dismisses this concern:

If additional properties of Mr. A became known, and deemed relevant (*e.g.*, that he had red hair, or was left-handed), these too could, in principle, be accounted for by restricting the analysis to data representing the appropriate subpopulations. However, in the absence of such data, and knowing in advance that we will never be able to match all the idiosyncratic properties of Mr. A, the lawmakers’ specification must be interpreted relative to the properties at hand.

But in court the ultimate issue, to which the burden of proof applies, will be informed by a myriad of idiosyncratic properties for which little or no data will be available. Such idiosyncrasies in the Blue Bus example might include the plaintiff’s blood alcohol level, the witness’s biases against the Blue Bus Company, memory issues, the validity of the police officer’s accident report, and so forth.

Finally, there are the problems associated with the ways in which these types of statistical evidence will inevitably be introduced in court by expert witnesses. The actual provenance of the data matters, as do the details of the experimental protocols, the scientific controls, and the biases, as well as the qualifications, of the experts, who may well favor the evidence provided by experimental data published in a peer-reviewed high-quality medical journal over the unpublished and uncontrolled observational data gathered by the witness for trial purposes.

Pearl envisions his solution to CoE as responsive to the ultimate issue when, in fact, it is merely one piece of evidence in the fabric of proof. This implicates our third problem: That the end users, judges, or jurors, need to be able to understand it. It is not simply the mathematics that participants in the courtroom will need to grasp, but all of the underlying or limiting assumptions. Virtually, none of Pearl’s “mathematization of counterfactuals” would



be understood by judges, lawyers, or jurors, and they would likely find the spate of assumptions he makes troubling, even if they could figure them out. Contrary to Pearl's stated hope, courts remain a considerable distance from employing "the structural semantics of counterfactuals . . . to decide individual cases of all kinds."

## **Conclusion**

Pearl concludes by inviting us to "reap the benefits of the counterfactual theory of CoE." Since such reasoning was already conducted in part III of Dawid (2000), and the analysis in our article on which he is commenting is firmly based on this approach, we are somewhat bemused by this invitation. But (for all its admitted technical successes) we cannot share Pearl's uncritical belief in the all-conquering power of his own framework, since it begs too many deep philosophical questions and offers little guidance as to how any given real problem should be modeled and how the empirical evidence should be used in the complex real-world settings of legal disputes and courtroom expert testimony; compare with the earlier efforts of Robins and Greenland (1989); Greenland and Robins (2000).

Lest we give the wrong impression, we humbly admit that our own efforts to advance the technical framing of CoE remain in an oversimplistic state. We continue to work on the development of these ideas (Dawid et al. 2014), and we gladly invite Pearl and our other readers to contribute to the task that remains.

## **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of this article was supported, in part, by a grant from the John D. and Catherine T. MacArthur Foundation. Its contents reflect the views of the authors, and do not necessarily represent the official views of either the John D. and Catherine T. MacArthur Foundation or the MacArthur Foundation Research Network on Law and Neuroscience ([www.lawneuro.org](http://www.lawneuro.org)).

## **Note**

1. This node-splitting ploy is very similar to the "SWIG" construction of Richardson and Robins (2013).

## References

- Cheng, E. J. 2013. "Reconceptualizing the Burden of Proof." *Yale Law Journal* 122: 1254-79.
- Dawid, A. P. 2000. "Causal Inference without Counterfactuals (with Discussion)." *Journal of the American Statistical Association* 95:407-48.
- Dawid, A. P. 2011. "The Role of Scientific and Statistical Evidence in Assessing Causality." Pp. 133-47 in *Perspectives on Causation*, edited by R. Goldberg. Oxford, UK: Hart Publishing.
- Dawid, A. P., D. L. Faigman, and S. E. Fienberg. 2014. "Fitting Science into Legal Contexts: Assessing Effects of Causes or Causes of Effects? (with Discussion)." *Sociological Methods and Research* 43(3):359-421.
- Dawid, A. P., M. Musio, and S. E. Fienberg. 2014. "From Statistical Evidence to Evidence of Causality." Submitted to *Bayesian Analysis*. <http://arxiv.org/abs/1311.7513>.
- Faigman, D. L., J. Monahan, and C. Slobogin. 2014. "Group to Individual (G2i) Inference in Scientific Expert Testimony." *University of Chicago Law Review* 81:417-80.
- Fienberg, S. E. 1986. "Comment: Gatecrashers, Blue Buses, and the Bayesian Representation of Legal Evidence." *Boston University Law Review* 66: 693-99.
- Geneletti, S. and A. P. Dawid. 2011. "Defining and Identifying the Effect of Treatment on the Treated." Pp. 728-49 in *Causality in the Sciences*, edited by P. M. Illari, F. Russo, and J. Williamson. Oxford: Oxford University Press.
- Greenland, S. and J. M. Robins. 2000. "Epidemiology, Justice, and the Probability of Causation." *Jurimetrics* 40:321-40.
- Nesson, C. 1985. "The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts." *Harvard Law Review* 98:1357-92.
- Pearl, J. 2015. "Causes of Effects and Effects of Causes." *Sociological Methods and Research* 44:150-65.
- Richardson, T. S. and J. M. Robins. 2013. "Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality." Technical Report 128, Center for Statistics and the Social Sciences, University of Washington, Seattle, WA, USA.
- Robins, J. M. and S. Greenland. 1989. "The Probability of Causation under a Stochastic Model for Individual Risk." *Biometrics* 45:1125-38.
- Tian, J. and J. Pearl. 2000. "Probabilities of Causation: Bounds and Identification." *Annals of Mathematics and Artificial Intelligence* 28:287-313.
- Tribe, L. H. 1971. "Trial by Mathematics: Precision and Ritual in the Legal Process." *Harvard Law Review* 84:1329-93.

## **Author Biographies**

**A. Philip Dawid** is Emeritus Professor of Statistics at the University of Cambridge. He has been awarded the Royal Statistical Society's Guy Medal in Silver for his wide-ranging research, which includes a long-standing interest in the interpretation of evidence and causal reasoning. His recently edited volumes include "Simplicity, Complexity and Modelling", "Evidence, Inference and Enquiry", and "Causality: Statistical Perspectives and Applications", as well as the book of the Darwin College Cambridge Lecture Series on "Beauty".

**David L. Faigman** is John F. Digardi Distinguished Professor of Law at the University of California, Hastings College of the Law. He also holds an appointment in the School of Medicine, Department of Psychiatry, at the University of California San Francisco and is Co-Director of the UCSF-UC Hastings Consortium on Law, Science & Health Policy. His scholarship primarily concerns the integration of scientific research into legal decision making. His most recent book is "Constitutional Fictions: A Unified Theory of Constitutional Facts," and he is a coauthor/coeditor of the five volume treatise "Modern Scientific Evidence: The Law and Science of Expert Testimony."

**Stephen E. Fienberg** is Maurice Falk University Professor of Statistics and Social Science at Carnegie Mellon University and Director of the Living Analytics Research Centre, a joint activity with Singapore Management University. Best known for his work on categorical data analysis, he has longstanding interests in the use of statistics in legal contexts and causal inference more broadly. He is a member of the National Academy of Sciences and a fellow of the American Academy of Arts and Sciences, and the Royal Statistical Society.