Comments on "The Tale Wagged by the DAG"

Judea Pearl
University of California, Los Angeles
Computer Science Department
Los Angeles, CA, 90095-1596, USA
judea@cs.ucla.edu

Abstract
This comment clarifies the role of structural models in epidemiological research and answers questions raised in a previous discussion on this topic.

Keywords: structural framework, triangulation, explanation, graphical models, transportability

I am grateful to the Editors for the opportunity to comment on Nancy Krieger and George Davey Smith's (KDS) article, "The Tale Wagged by the DAG" [1].

KDS raise several objections to the direction taken by modern epidemiology, which they deem to be too narrowly wedded to a DAG-counterfactual framework. In this framework, graphical models (DAGs) are used to express scientific knowledge, and counterfactuals (or potential outcomes) are used to express queries of interest.

Since the article does not demonstrate concrete alternatives to current methodologies, I speculate that it is the dazzling speed with which epidemiology has modernized its tools that lies behind the author's discomfort, and that it will subside as soon as researchers gain greater familiarity with the capabilities and flexibility of these new tools.

Epidemiology, as I have written on several occasions, has been a pioneer in accepting the DAG-counterfactuals symbiosis as a ruling paradigm – way ahead of mainstream statistics and its other satellites [2].

In examining the specific limitations that KDS perceive in DAGs, I must note that these limitations coincide precisely with the strengths for which DAGs are praised.

For example, the article complains that DAGs provide no information about variables that investigators failed to include in the model. In their words, "the DAG does not provide a comprehensive picture. For example, it does not include paternal factors, ethnicity, respiratory infections or socioeconomic position..." (from the Editorial introduction). This should not be taken as a limitation of DAGs or of any other scientific modeling. Quite the contrary. It would be a disaster if models were allowed to produce information unintended by the modeller. Instead, I have come to admire the ease with which DAGs enable researchers to incorporate new knowledge about new variables, or new mechanisms, when the need arises.

Model misspecification is a problem that plagues every exercise in causal inference, no matter what framework one chooses to adapt. It can only be cured by careful model-building strategies, and by enhancing the modeller's knowledge. Yet, when it comes to preventing misspecification errors, DAGS have no match. The transparency with which DAGs display the causal assumptions in the model, and the ease with which the DAG identifies the testable implications of those assumptions are incomparable; these facilitate speedy model diagnosis and repair and, in this way, brings misspecification errors to a minimum. This is the most one can ask from a judgmental-based model.

Or, to take another example, the authors call for an ostensibly unavailable methodology, which they label \causal triangulation" (it appears 19 times in the article). In their words: "In our field, involving dynamic populations of people in dynamic societies and ecosystems, methodical triangulation of diverse types of evidence from diverse types of study settings and involving diverse populations is essential." Ironically, however, the task of treating "diverse type of evidence from diverse type of study settings and involving diverse populations" has been shown to require the DAG-counterfactual framework even for the most elementary examples analyzed by [3]. In these examples, one takes diverse type of study settings, or diverse populations, and infer causal parameters in new settings, not yet studied. It is hard to imagine how more elaborate triangulations, say those integrating physiological studies on rats, evolutionary theory and knowledge of intracellular chemical reaction, could be accomplished without the machinery provided by the DAG-counterfactual framework. In other words, tools that are indispensable in solving simple problems are unlikely to become dispensable when problems become more complex.

Another conceptual paradigm which the authors hope would liberate us from the tyranny of DAGs and counterfactuals is Lipton's [4] romantic aspiration for "Inference to the Best Explanation (IBE)." IBE is a compelling mantra, going back at least to Charles Peirce theory of abduction [5] which, unfortunately, has never operationalized its key terms: "inference-to," "Best," and "explanation." Again, I know of only one framework in which this aspiration has been explicated with sufficient precision to produce tangible results – the structural framework of DAGs and counterfactuals. See for example, the works on causal explanations and causes of effects in [6], [7], and [8].

In summary, what Krieger and Davey Smith aspire to achieve by abandoning the structural framework has already been accomplished with the help and bliss of that very framework. More generally, what we can learn from these examples is that the DAG-counterfactual symbiosis is far from being a narrow "ONE approach to causal inference?" which "may potentially lead to spurious causal inference" [1] (p. 1787). It is in fact a general and flexible framework within which a plurality of tasks and aspirations can be formulated, analyzed and implemented. It is based on a rigorous logic of causation and, therefore, it is as protected from "leading to spurious causal inferences" as mathematics itself.

I was pleased to note that, by and large, other commentators on Krieger and Davey Smith paper seemed to be aware of the powers and generality of the DAG counterfactual framework, albeit not exactly for the reasons that I have described here. (I have many disagreements with other commentators as well, but I wish to focus here on the TALE WAGGED DAG, where the problems appear more glaring.) In a recent paper [9], I provide a concise summary of the achievements of the structural framework, and explained why one is justified in calling those achievements a "Causal Revolution."

All in all, epidemiology has made an incredible progress in the past two decades by adapting the DAG-counterfactual symbiosis as a unifying language for communicating research problems, analytical tools, and research results [8] [10]. To appreciate this progress, one needs only compare to the status of causal inference in fields such as economics [11] [12], where selecting covariates for confounding control is still a black magic, and where testable implications of one's assumptions are a wishful dream. The flexibility of the DAG-counterfactual language is likely to position epidemiology among the first beneficiaries of the big data and machine learning generation. Here, inferences across populations and across experimental settings become feasible but their management requires the machinery of structural modeling [3]. At the same time, I hope that the discomfort that Krieger and Davey Smith's have expressed will be short lived and that it will inspire a greater understanding of the modern tools of causal inference.

Bibliography

[1] N. Krieger and G. D. Smith, "The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology.," *International Journal of Epidemiology,* vol. 45, no. 6, pp. 1787-1808, 2016.

[2] J. Pearl, Causality: Models, Reasoning, and Inference, 2009, 2nd ed., New York, NY: Cambridge University Press, 2000.

[3] E. Bareinboim and J. Pearl, "Causal inference and the data-fusion problem," in *Proceedings of the National Academy of Sciences*, USA, 2016.

[4] P. Lipton, Inference to the Best Explanation., London: Routledge, 2004.

[5] C. Peirce, "What is Pragmatism," *The Monist,* vol. 15, no. 2, pp. 161-181, 1905.

[6] J. Pearl, "Causes of Effects and Effects of Causes," *Journal of Sociological Methods and Research,* vol. 44, pp. 149-164, 2015.

[7] J. Halpern and J. Pearl, "Causes and explanations: A structural-model approach–Part I: Causes," *British Journal of Philosophy of Science,* vol. 56, pp. 843-887, 2005.

[8] T. VanderWeele, Explanation in causal inference: Methods for mediation and interaction,

New York, NY: Oxford University Press, 2015.

[9]   J. Pearl, "The Eight Pillars of Causal Wisdom," http://ftp.cs.ucla.edu/pub/stat ser/r470.pdf, Los Angeles, 2017.

[10] M. A. Hernán and J. M. Robins, Causal Inference, Boca Raton: Chapman & Hall/CRC, 2018, forthcoming.

[11] G. W. Imbens and J. M. Rubin, Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, Cambridge, MA: Cambridge University Press, 2015.

[12] J. Pearl, "Trygve Haavelmo and the emergence of causal calculus," *Econometric Theory,* vol. 31, pp. 152-179, 2015.