Commentary: On Causes, Causal Inference, and Potential Outcomes

International Journal of Epidemiology, 2016, 1809–1816
doi: 10.1093/ije/dyw230
Advance Access Publication Date: 27 January 2017

Tyler J VanderWeele

Harvard T.H. Chan School of Public Health

Accepted 20 July 2016

I would like to structure my remarks on the papers by Vandenbroucke et al.[1] and by Krieger and Davey Smith[2] around a series of propositions that articulate my views concerning causes, causation, and potential outcomes. I will then relate each of the propositions to the material discussed by Vandenbroucke et al. and by Krieger and Davey Smith and discuss where I agree and disagree with their comments. Certain claims about the potential outcomes framework made by Vandenbroucke et al. have also been clarified in a recent letter.[3] See also related discussion[4] on another recent similar article by Schwartz et al.[5]

(1) A distinction ought to be drawn between conditions under which we can reasonably describe something as a cause and under which we can reasonably define a quantitative causal effect estimand. The potential outcomes framework provides an approach to defining casual effect estimands

Krieger and Davey Smith speak of 'what the current counterfactual framework. . . is precluding from being deemed a "cause".' A distinction should be drawn between under what circumstances it is reasonable to refer to something as a cause and under what circumstances it is reasonable to speak of an estimate of a causal effect i.e. a particular number (possibly with accompanying confidence interval) that we attempt to give a precise causal interpretation. Said another way, the circumstances may differ with regard to when we are willing to talk about a cause qualitatively versus when we think it is reasonable to quantify. The potential outcomes framework provides a way to quantify causal effects. For a hypothetical intervention, it defines the causal effect for an individual as the difference between the outcomes that would be observed for that individual with versus without the exposure or intervention under consideration. To make this definition precise, the potential outcomes framework imposes certain assumptions, such as that the hypothetical intervention is sufficiently well defined for the outcome under the intervention to be, for each individual, unique or to have a unique distribution.[6,7] The potential outcomes framework provides a set of sufficient conditions for defining a quantitative causal effect, i.e. a causal estimand. The term "causal effect" is itself somewhat ambiguous insofar as it might simply be used to refer to an effect of a cause; however, it is often used in a more precise sense within the potential outcomes framework as a numeric quantity that one seeks to estimate from the data, also referred to as a "causal estimand." From this point onward I will refer to this simply as a "causal estimand."

The potential outcomes framework then formally articulates the assumptions under which such causal estimands can be estimated on average (or in distribution) for a population. When there is an intervention in view, it generally makes sense to speak about causal effects quantitatively e.g. if we are considering the intervention taking pill X, we can consider how much blood pressure might differ for an individual with or without the pill. Generally, whenever it is reasonable to talk about a specific causal estimand then, provided it is non-zero for someone, it is also reasonable to talk of the hypothetical intervention or the exposure as a cause. However, the reverse is not true: there are many contexts in which it is reasonable to speak of something as a cause but more difficult to articulate what one means by a causal effect estimand. As will be described below, I think it is quite reasonable to say that obesity is a cause of longevity, but it is much more difficult to define what precisely one means by the causal estimand for obesity.

Thus, contrary to Vandenbroucke et al. and Krieger and Davey Smith, the potential outcome framework thus does not delimit what is "a cause". It provides sufficient conditions to conceptualize, quantitatively, causal effect estimands, and I argue below, there are not many satisfactory alternatives for this at present. Quantitative definitions, assumption articulation, interpretation, and effect estimation are, at present at least, largely restricted to the potential outcomes framework, or those from which potential

outcomes can be derived or have analogues such as causal Bayes' nets or intervention diagrams.[8–10]

(2) All counterfactuals are vague to a certain extent but can be made more precise by specifying further the contrary-to-fact scenario in view. When such contrary-to-fact scenarios are sufficiently precise, they admit potential outcomes. The potential outcomes framework then provides the conceptual and mathematical link between data and causal effect estimands

A position that is often articulated is that all counterfactuals are vague to a certain extent, but that they are made more precise by specifying further the relevant contrary-to-fact scenario or by specifying further what is meant by the counterfactual.[11–14] I believe this pertains both to discussions of causal estimands and to discussions of causes.

This position has also has effectively been adopted by those employing the potential outcomes framework. Potential outcomes are always ill-defined to a certain extent but can be made more precise by further specifying the hypothetical intervention or contrary-to-fact scenario under consideration. The contrary-to-fact scenario under consideration need not correspond to a human action, but if it does, this helps to specify further what is meant. However, specifying a hypothetical intervention may not eliminate all ambiguity. Thus, with regard to the example of taking pill X, there may be still questions as to the timing of taking the pill and whether the pill is taken with or without food. Specifying some of these may make the counterfactual under consideration more precise. But some of these aspects may not be relevant for the outcome and can thus be potentially ignored.[15] Conceptualizations that allow the potential outcomes for each individual to follow a distribution yet further relax the requirements.[7,15] The general principle, however, is that the further we specify the hypothetical intervention, the more precise is our counterfactual. If the hypothetical intervention is sufficiently precise, then we might entertain using numeric potential outcomes, with and without the intervention, and can consider causal estimands.

However, even if our exposure or hypothetical interventions are not precise we can still pose counterfactual queries or ask if particular exposures are causes. With obesity, for example, the statement that "obesity is a cause of longevity" is vague. One could, however, specify further what is meant. If someone is queried further as to the meaning of the statement, it might be understood along the lines of that, if one could reduce the body mass index (BMI) for those with BMI of 30 to bring this down to a BMI of 25, then this intervention would result in a longevity benefit for some of those individuals. But the counterfactual is still vague: any practical way to implement the reduction in BMI, whether by exercise, diet, stomach surgery, or liposuction may well have different effects on longevity.[16] However, if the truth value of every possible interpretation or more precise specification of the counterfactual were the same, then it might be reasonable to assign that truth value to the ill-defined imprecise counterfactual as well. In such a case we might be willing to say that "obesity is a cause of longevity," even though formal potential outcomes, and thus causal estimands, are not well-defined.

There is a continuum of how well-defined counterfactuals are. As is made clear in the example above, the requirements in terms of removing ambiguity for numeric potential outcomes is more stringent than that for general counterfactual inquiries or discussion of causes. Of course, simply writing down potential outcomes notation does not mean that the potential outcomes are well-defined. They are well-defined to the extent that the hypothetical intervention or contrary-to-fact scenario is specified.

When it is reasonable to use potential outcomes, the framework provides the conceptual and mathematical link between the data and the causal effect estimands. In actual epidemiologic research, investigators collect data – encoded as strings of numbers. Then, a data analyst performs various computations on these strings. Based on these computations, the investigators may produce written or verbal sentences concerning causal relationships.[17] The relationship between these strings of numbers and the resulting sentences needs a formal framework for translation. The potential outcomes and related framework provides the needed formalization. There now exists an entire mathematical calculus for reasoning about potential outcomes framework based on causal diagrams or directed acyclic graphs (DAGs).[8,18] In some settings, such as is the case with the effects of time-varying exposures, adequate methodology required to assess such effects arguably could not be derived without the sort of formality that potential outcomes and related frameworks provide.[14,19–21] Likewise, certain questions about mediation and direct and indirect effects for which it was long thought, in practice at least, that a simple regression approach was sufficient, turned out to be subject to far more pitfalls than previously realized, insights again made possible through potential outcomes and causal diagrams.[13,22,23] The above truly are major advances over what was available previously, and I believe lies at the heart of the current interest in these methods. Thus while I agree with Krieger and Davey Smith that "DAGs and counterfactual approaches are but one set of conceptual tools that epidemiologists can employ, and

should not occupy a privileged place in delimiting the kinds of questions we ask or causes we theorize, I also think that potential outcomes and related approaches do have a privileged position in the estimation of causal effect estimands.

(3) Some progress has been made in defining causal effect estimands for settings which lie outside of the traditional potential outcomes framework but this remains challenging

When we move on to try to talk of causal effect estimands for more composite exposures (i.e. those which do not correspond to a hypothetical intervention) such as obesity or cholesterol, precisely articulating what is meant becomes more difficult. The potential outcomes framework is not as obviously applicable to such composite exposures. If one tries to frame the effect of obesity on mortality within the potential outcomes framework it becomes clear that different interventions to reduce obesity – diet, exercise, surgery – may well have different effects on mortality.[16] It thus becomes difficult to speak of *the* causal effect estimand comparing a BMI of 30 versus 25. That is not to say that obesity is not a cause, but rather that trying to give an estimate a precise interpretation is challenging.

Some research has addressed such questions by extending the potential outcomes framework so as to include multiple versions of the exposure or intervention.[24,25] Under various assumptions, the estimate of a composite exposure such as obesity can be conceived of as a weighted average of the effects of the different, more precise, "versions of the exposure", weighted by the likelihood of each version naturally arising within the subpopulation that was actually exposed.[24,25] However, the meaning of such an interpretation is difficult when the underlying distribution of the "versions of exposure" are unknown, as is often the case. Moreover the no unmeasured confounding assumptions required for this interpretation are difficult to assess when the different underlying versions of treatment or exposure are unknown. [25] Some progress has thus been made in providing a more precise interpretation of the causal effect estimands of such composite exposures, but whether and how such interpretations are helpful in practice remains to be seen.

Pearl[8] has proposed that causal effect estimand of a composite exposure be viewed, not as the effect of a hypothetical intervention, but as an effect arising from an underlying structural equation model. An analogy is made with the laws of nature whereby the exposure, the state of the universe and the underlying laws suffice to set the outcome or its distribution. If the exposure were different, the outcome would be as well. Within the framework of DAGs or causal diagrams, the structural equation model takes the place of the laws of nature. While such an interpretation is promising in principle, its use in practice is problematic. First, unlike in the physical sciences, in the social and biomedical sciences, the complexity of the systems studied are such that we are almost never in a position to be able to articulate a causal diagram that would adequately capture the natural laws. Second, even if we could, when we have a composite exposure in view, many of the constructs we use in the social and biomedical sciences are not sufficiently precise so as to be a part of the underlying laws of nature. If we are considering the effect on mortality of a change in BMI from 30 to 25, the fundamental underlying laws that result in death are such that there is not a one-to-one mapping from BMI to the underlying variables relevant to the natural laws.[26] Implicit within the proposed structural equation model interpretation is that it is possible to conceive of the exposure being different but all else within the system being the same, and, with many of our composite exposures, this is not possible. This does not mean that causal diagrams are irrelevant here – often a simplified diagram can help diagnose bias in such a way that one can see it would also be present on any more adequate representation of reality. It also does not mean that we cannot talk about causation – if every possible mapping from BMI to the underlying relevant variables implied that a change in BMI from 30 to 25 resulted in a change in mortality for some persons, we might well speak of obesity as a cause.[13] However, it does make it more difficult to talk of a causal effect estimate. Perhaps again the best we can do in such settings is to appeal to what is still a somewhat vague interpretation under theory concerning "multiple versions of treatment," as described above. Thus, although the approach to conceptualizing causal effect estimands of composite exposures through the underlying structural equation model is theoretically appealing, this interpretation is, in fact, still difficult in practice. I have not yet seen any compelling empirical examples.

The potential outcomes framework begins to reach its limits with such composite exposures. It instead points the researcher to precise hypothetical interventions, which do not correspond to the complexity of some of the exposures that are studied. Responses to these difficulties may vary. If policy intervention or action is of interest, then the potential outcomes framework can be helpful in thinking through the range of interventions to alter a composite exposure and which of these might best improve outcomes. One can alternatively attempt to pursue a more nuanced interpretation as in the proposals above. Or one can abandon any attempt at a precise interpretation of a quantitative estimand and simply speak of evidence for general causation, for something being a cause. One can, of course, criticize or complain, as do Vandendroucke et al., that the

potential outcomes framework does not adequately addresses questions of such composite exposures, but this amounts to simply a complaint that a tool does not address questions for which it was not intended or that are even well-defined. Nor do such complaints help in any way to provide a precise interpretation of the effects of such composite exposures. What little progress has been made on these questions has in fact come out of extensions of the potential outcomes or causal diagram frameworks.

(4) Although policy should be informed by many considerations, estimands from causal inference using potential outcomes have a central role in deciding between policies and interventions based on the quantification of causal effects. There are, however, important questions that are not amenable to the potential outcomes framework

Perhaps a more reasonable and profound complaint, suggested by both Vandenbroucke et al. and Krieger and Davey Smith, and also in a recent article by Schwartz et al.[26], is that the potential outcomes framework itself is restricting the types of questions that are being posed. As alluded to above, some of the issue is simply the complexity of the systems or exposures being studied. Krieger and Davey Smith make reference, for example, to social movements, societal trends such as more married women going to work, and even war. With these questions the complexity of the "exposures" or "systems" under study compounds even further the issues described in the previous section. The potential outcomes framework simply is no longer the right tool. There is no single intervention that corresponds to "the civil rights movement." Other modes of inquiry and evaluation are needed. There are subfields of history, for example, that do attempt to address these questions. In complex settings concerning wars or social movements, even if data are aggregated over many different wars, say, the precise meaning of the counterfactual inquiries is very ambiguous as discussed in the second and third propositions above. If epidemiologists want to and are successfully able to address questions concerning individual historical events that is fine and good, but these do require very different disciplinary tools and different data than epidemiologists typically have had at their disposal. These are still counterfactual queries, but of a very different nature, and ones which are often more difficult to answer.

The potential outcomes framework can be useful in narrower policy evaluations and decisions. When various actions are being considered it can help with articulating what precisely is under study, the data and the assumptions needed to estimate the effects of specific actions, and thus to provide guidance on the relative magnitudes of those effects and to choose between actions.

Arguably a great deal of the popularity of the potential outcomes approach within epidemiology has come about because of the precision the framework offers and its success in evaluating numerous types of causal questions concerning effect estimates. However, as suggested by both Vandenbroucke et al. and Krieger and Davey Smith, that very success may well have led to a shift in focus in the types of questions being asked. Those questions which are not amenable to a potential outcomes analysis are perhaps receiving less attention. And again, as noted by Krieger and Davey Smith, some of these questions are of tremendous societal importance. The solution, it seems to me, is not an attack on, or caricature of, the potential outcomes framework, but rather ensuring that, within teaching, and in the published literature, more examples of broader questions concerning systems and movements are discussed, along with examples of where other types of reasoning and evidence have proved to be useful. Causal inference using potential outcomes are but a small subset of courses offered in most departments of epidemiology or schools of public health. This other material can and should be covered in other courses and this is of course already happening at many institutions.

(5) Although no complete characterization of causation is yet available, there are well established sufficient conditions for attributing causation and the potential outcomes framework provides one such set of sufficient conditions. While the open questions on causation are important, this should not be used to obscure the progress that has already been made

As noted by Vandenbroucke et al. and by Krieger and Davey Smith, there are ongoing discussions in the philosophical literature concerning the nature of causation, with different conceptions put forward. No complete characterization of what is meant by causation is yet available. In particular, there is not yet any complete characterization (i.e. a set of necessary and sufficient conditions) for the statement "X caused Y".[28,29] Although we can often come to consensus in any given instance, and perhaps seemingly do so according to some unspecified set of rules, no one to date has yet been able to precisely articulate those rules. This does not, however, mean we cannot sometimes, even often, reason about causation. Although there is not yet a complete characterization, there are sufficient conditions under which we can attribute causation. In particular, I believe there would be near universal consensus that if it were the case that a particular event X, along with the

current state of the universe and the laws of nature entailed some subsequent event Y, and if the absence of the event X, along with the current state of the universe, and the laws of nature entailed the absence of Y, and if X and Y were in fact present, then almost all would say "X caused Y." These are not the only circumstances under which we would say "X caused Y" but they are sufficient conditions for our saying so. This is a type of relationship of counterfactual dependence. That such counterfactual dependence is not necessary for attributing causation can be made clear by so-called cases of overdetermination[30,31]: if an explorer's water supply is poisoned by enemy 1 and then subsequently emptied by enemy 2 and the explorer goes off to the desert and dies from dehydration, we would generally say the emptying of the water caused the explorer's death, even though the explorer would still have died (by poison) had the water not been emptied.

The potential outcomes framework effectively takes the event X in sufficient conditions above as some sort of hypothetical intervention; if the outcome differs depending on whether the intervention is present or absent, then the intervention is said to have affected the outcome for that individual. The potential outcomes framework thus provides a sufficient condition for attributing causation. However, like all approaches to date, it does not provide a complete characterization of causation. It does not, contrary to Vandenbroucke et al., delimit the bounds of causality; it is concerned with a subset of causal questions that can be defined as a contrast of hypothetical interventions. So while I strongly disagree with Vandebroucke et al. on this point about delimitation, I agree with both them and with Krieger and Davey Smith that multiple perspectives are currently needed in our thinking about causal reasoning. The potential outcomes framework provides one important approach, and again, as above, essentially the only one for thinking quantitatively about causal effect estimands.

(6) Explanation is a much broader concept than causal explanation; scientific reasoning is a much broader concept than causal inference. Inference to be best explanation is important in causal inference and diverse types of evidence can and should be used. Often, in the end, however, the most compelling evidence for causal reasoning comes from counterfactual comparisons

I agree with both Vandenbroucke et al. and with Krieger and Davey Smith that diverse sorts of evidence ought to be considered and I agree with them on the importance of temporal trends, of negative controls, and of using different types of study designs. I agree further with Krieger and Davey Smith that inference to the best explanation is an important approach to scientific reasoning, though, often, before consensus is reached, there is a period during which scientists disagree about what that best explanation is. I also agree that scientific discovery involves numerous steps that are not deductive, that hypotheses are often generated from creative conjecturing or moments of sudden insight, and that thinking through multiple levels of explanation can be helpful in gaining a fuller understanding. None of this seems open to much dispute.

However, I believe that when it comes to reasoning about causes, it is often counterfactual contrasts that provide the most compelling form of evidence. In Textbox 3 of Krieger and Davey Smith, every method and study design put forward as of use in "triangulation and inference to the best explanation" has been formalized within the potential outcomes framework[32–37] and effectively involves counterfactual contrasts. In Textbox 2 of Krieger and Davey Smith, in their example of inference to the best explanation drawn from Peter Lipton's exposition of Semmelweis's research programme, the definitive final step confirming Seemelweis's cadaveric hypothesis involved the counterfactual contrast of what occurred before versus after medical students were required to wash their hands.[38] The generation of the hypothesis itself, of course, involves insight arising from multiple individual pieces of evidence, but the confirmation, and ultimately the strongest evidence came from a counterfactual comparison, and it is my view that in most cases the most compelling evidence in causal reasoning will come from counterfactual comparisons.

It is such comparisons that the potential outcomes framework, and that causal diagrams, attempt to formalize. Within the broader context of explanation and of scientific inference, potential outcomes and causal diagrams are just tools. They do not address all questions. As per Krieger and Davey Smith they certainly do not address which measures to use or how to conceptualize these measures. Nor, as in Krieger and Davey Smith's description, do the diagrams supply the biological explanation, but they are not intended to. They do illustrate biological explanations, however, and when used in the context of the birthweight paradox, explanations have accompanied them. Causal diagrams are important tools for causal reasoning but even there they have their limits. There is no dispute that DAGs, for example, can only go so far in representing interactions or quantitative relations. As Krieger and Davey Smith note, they do not quantify the magnitude of biases, but there are other tools for that, within the potential outcomes framework, such as sensitivity analysis.[39–41] There are still disputes perhaps about the precise nature of the limits of various approaches (contrary to what seems to

be suggested by Krieger and Davey Smith, for example, DAGs can and have been used to represent feedback) but they do not address all questions or all forms of causal reasoning.

(7) While it may not be unreasonable to claim that race is a cause of health and of other outcomes, it still remains difficult to precisely define what one might mean by a quantitative causal effect estimand for race. Counterfactual-based methodology concerning how much of a racial inequality could be reduced by potentially implementable interventions constitutes an important approach to understanding how best to reduce disparities

Vandenbroucke et al. and also Krieger and Davey Smith bring up causal questions concerning race. In their Textbox 4, Krieger and Davey Smith cite some of my work as examples of 'counterfactual reasoning that "race" cannot be a cause.' Similar claims are made by one of the authors of the Vandenbroucke et al. commentary.[42] Contrary to what they suggest, I have made no such claims.

The mistake of both Krieger and Davey Smith[2] and Broadbent[42] is to confuse conditions under which it is reasonable to speak of something as a cause with conditions under which it is reasonable to give a causal interpretation to a numeric effect estimand, as per the first proposition above. Questions of race and racism are undoubtedly of importance and ought to be addressed. Diverse approaches to thinking about these questions and policy measures to address them ought to be pursued. However, when it comes to talking about a numerical causal effect estimand for race, this becomes more difficult. It is this task of considering under what conditions one can provide a precise causal interpretation of race, or of regression coefficients for race, or of aspects of race and racism, that has occupied some of the potential outcomes literature on race.[43–45] Much of this work concludes that such interpretations with respect to race, while perhaps not impossible, are difficult, as per the second and third propositions above. However, to claim that such research is therefore asserting that race is not a cause is again to confuse when it is reasonable to speak of something as a cause and when it is possible to interpret a numeric causal effect estimand. One might say that "race is a cause" in the sense that, whatever might be meant by race, or however one might define or conceptualize race, that conception would entail also certain features such as skin color, and that these themselves are perceived by others and, as a result of discrimination, affect health, income, and other outcomes. One might thus maintain, that under any reasonable conceptualization of race, and any reasonable conception of causation, race satisfies sufficient conditions for attributing causation – essentially that the truth value of "race is a cause of health" is the same irrespective of how the causal claim is interpreted or further specified. Making such statements is thus perhaps not unreasonable. I think a more helpful approach, however, is trying to more precisely specify the claim or claims that may actually be in view. It is interesting to note that Krieger and Davey Smith themselves do not explicitly assert that race is a cause, nor deny it, but rather state that "the relevant counferfactual pertains to racism, not 'race.'" I think it is indeed easier to conceptualize causal effect estimands for various aspects of racism and in fact some of the potential outcomes work on this topic has done precisely that.[43]

In some of my work[45,46], I have also written about methodology whereby race coefficients in regression models with covariates (such as childhood conditions) or potentially mediating variables (such as high school educational achievement) can be interpreted as by how much a racial inequality could be reduced and how much of it would remain under hypothetical interventions to equalize the distributions of the covariates (e.g. childhood socioeconomic conditions) or potentially mediating variables (e.g. educational achievement) across racial groups. I have argued that such analyses and interpretations can perhaps be useful in identifying what sorts of the interventions would be most effective in reducing disparities.[45,46] The methodology does *not* somehow implicitly assume that race is not a cause; the use of the methodology does *not* imply that we should ignore broader societal factors, social movements, and racial discrimination. Rather, in keeping with potential outcomes thinking, the methodology focuses on a set of questions concerning interventions that might reduce racial inequalities. I would argue that for purposes of, not simply documenting and understanding, but addressing and reducing racial disparities, such methodology is of use. The methodology was apparently considered sufficiently useful by Krieger at least so as to be willing to participate in a study that employed it in examining racial inequalities in colorectal cancer survival.[47] Questions concerning race and racism certainly should be addressed by tools other than the potential outcomes framework, but even here, I would argue the framework is of value for a number of policy-relevant questions.

Concluding Remarks

I have, in this commentary, attempted to state more precisely my own views on issues of causation, causal effects and potential outcomes and to relate those to what has

been expressed in the articles of Vandenbroucke et al. and Krieger and Davey Smith. I have drawn a number of distinctions and have put forward a series of propositions that I believe are helpful in reasoning about causation and causal effects. The drawing of distinctions and precision in terminology is essential to science. However, also important to the progress of science is correct attribution of views. Krieger and Davey Smith cite some of my work as examples in their Textbox 4 of 'counterfactual reasoning that "race" cannot be a cause.' I have made no such claim. This is not a view that I hold. Vandenbroucke et al. claim that the potential outcomes approach "equates causal claims with precise predictions about contrary-to-fact statements" and that it "denies the meaningfulness... of causal claims that do not readily yield predictions" and cite my work as representative of that view. These again simply are *not* views that I hold. The citations chosen are cherry-picked, and taken out of context, and still they do not yield the claims Vandenbroucke et al., and elsewhere Broadbent[42], make. More careful distinctions and closer readings would have avoided these errors. Careful causal reasoning and careful distinctions are important for science, and so also with correct attribution as well.

Funding

This research was funded by United States NIH grant R01 ES017876.

Acknowledgements

I would like to thank Jamie Robins for helpful discussions. We were originally writing together but, due to the preponderance of points that we wanted to make and the desired differences in emphasis, we have written two distinct commentaries; but we are each indebted to the other.

References

1. Vandenbroucke JP, Broadbent A, Pearce N. Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol* 2016;**45**:1776–86.
2. Krieger N, Davey Smith G. The tale wagged by the DAG: broadening the scope of causal inference and explanation for epidemiology. *Int J Epidemiol* 2016;**45**:1787–808.
3. VanderWeele, TJ Hernán, MA, Tchetgen Tchetgen, EJ, and Robins, JM. Re: Causality and causal inference in epidemiology: the need for a pluralistic approach. *Int J Epidemiol* 2016;**45**: 2199–200.
4. Hernán, MA. Does water kill? Causal inferences anchored to target trials or how to make less casual causal inferences. Annals of Epidemiology, in press.
5. Schwartz S, Campbell UB, Gatto BM. Causal identification: reclaiming an abandoned charge of epidemiology. Annals of Epidemiology, in press.
6. Rubin DB. Comment on: "Randomization analysis of experimental data in the fisher randomization test" by D. Basu. *J Am Stat Assoc* 1980; **75**: 591–93.
7. Robins, JM and Greenland, S. The probability of causation under a sto-chastic model for individual risk. *Biometrics* 1989;**45**, 1125–38.
8. Pearl J. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2009.
9. Dawid, AP. Causal inference using influence diagrams: The problem of partial compliance (with Discussion). In *Highly Structured Stochastic Systems*, edited by Peter J. Green, Nils L. Hjort and Sylvia Richardson. Oxford University Press, 45–83, 2003.
10. Spirtes, P, Glymour, C and Scheines, R. *Causation, Prediction and Search*. MIT Press, Cambridge, MA, 2000.
11. Lewis, D *Counterfactuals*. Harvard University Press, Cambridge, 1973.
12. Robins JM, Greenland S. Comment on "Causal inference without counterfactuals." *Journal of the American Statistical Association*, 2000;**95**:477–482.
13. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press, 2015.
14. Hernán MA, Robins JM. Causal Inference. Chapman Hall, forthcoming.
15. VanderWeele, TJ Concerning the consistency assumption in causal inference. *Epidemiology*, 2009;**20**:880–883.
16. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well defined interventions to answer causal questions. *International Journal of Obesity* 2008; **32**:S8–S14.
17. Robins JM. Addendum to A new approach to causal inference in mortality studies with sustained exposure periods – Application to control of the healthy worker survivor effect. *Comput math appl* 1987; **14**:923–45.
18. Richardson, TS. and Robins, JM. Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality. Working Paper Number 128, Center for Statistics and the Social Sciences, University of Washington, 2013.
19. Robins JM. A new approach to causal inference in mortality studies with sustained exposure period – application to control of the healthy worker survivor effect. *Math Model* 1986; **7**: 1393–512.
20. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 2000;**11**:550–560.
21. Robins JM, Hernán MA. In: *Estimation of the causal effects of time-varying exposures*. Fitzmaurice G, M Davidian, Verbeke G Molenberghs G, eds. New York: Chapman and Hall/CRC Press, 2009.
22. Robins, JM and Greenland, S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 1992;**3**:143–155.
23. Pearl J. Direct and indirect effects. In: *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*. San Francisco: Morgan Kaufmann, 2001: 411–420.
24. Hernán MA, VanderWeele, TJ. Compound treatments and transportability of causal inference. *Epidemiology*, 2011;**22**: 368–377.

25. VanderWeele, TJ. and Hernán, MA. Causal inference under multiple versions of treatment. *Journal of Causal Inference*, 2013;**1**:1–20.
26. VanderWeele, TJ and Hernán MA. Casual effects and natural laws: towards a conceptualization of causal counterfactuals for non-manipulable exposures with application to the effects of race and sex. In: C. Berzuini, P. Dawid and L. Bernardinelli (eds.): *Causality: Statistical Perspectives and Applications*. Wiley. Chapter 9, p. 101–113, 2012.
27. Schwartz S, Prins SJ, Campbell UB, Gatto NM. Is the "well-defined intervention assumption" politically conservative? Social Science in Medicine, 2015, http://dx.doi.org/10.1016/j.socscimed.2015.10.054.
28. Vander Weele TJ. Criteria for the characterization of token causation. *Logic and Philosophy of Science*, 2009;**7**:115–127.
29. Glymour C, Danks D, Glymour B, Eberhardt F, Ramsey J, Scheines R, Spirtes P, Man Teng P, Zhang J. Actual causation: a stone soup essay. *Syntheses* 2010;**175**:169–192.
30. Collins J, Hall N, Paul LA. *Causation and Counterfactuals*. Cambridge, MA: MIT Press.
31. Hall, N. and Paul, LA. Causation and preemption. In P. Clark and K. Hawley (eds.), *Philosophy of Science Today*, Oxford, Oxford University Press, 2003, pp. 100–129.
32. Abadie A. Semiparametric difference-in-differences estimators *The Review of Economic Studies* 2005;**72**:1–19.
33. Athey S, Imbens GW. Identification and inference in nonlinear difference-in-]differences models. *Econometrica* 2006;**74**: 431–497.
34. Didelez V, Sheehan NA. Mendelian randomisation as an instrumental variable approach to causal inference, *Statistical Methods in Medical Research*, 2007;**16**, 309–330.
35. Sjölander A, Frisell T, Öberg S. Causal Interpretation of Between-Within Models for Twin Research. *Epidemiologic Methods*; 2012;**1**:10, p. 216–237.
36. Tchetgen Tchetgen E. The control outcome calibration approach for causal inference with unobserved confounding. *Am J Epidemiol*. 2014 Mar 1;**179**(*5*):633–40.
37. Sofer T, Richardson DB, Colincino E, Schwart J, and Tchetgen Tchetgen EJ. On Simple Relations Between Difference-in-differences and Negative Outcome Control of Unobserved Confounding (August 2015). Harvard University Biostatistics Working Paper Series. Working Paper 194. http://biostats.bepress.com/harvardbiostat/paper194
38. Lipton P. *Inference to the Best Explanation*. 2nd ed. London: Routledge, 2004.
39. Robins, JM, Scharfstein, D, and Rotnitzky, A. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: *Statistical Models for Epidemiology, the Environment, and Clinical Trials*. Halloran, E., and Berry, D. (eds), 1–95. New York: Springer-Verlag, 2000.
40. Rothman, KJ., Greenland, S., Lash, TL. *Modern Epidemiology*, 3rd Edition. Lippincott Williams & Wilkins, 2008.
41. Ding, P, and VanderWeele, TJ. Sensitivity analysis without assumptions. *Epidemiology*, 2016; **27**:368–377.
42. Broadbent A. Causation and prediction in epidemiology: a guide to the" Methodological Revolution." *Studies in the History and Philosophy of Biological and Biomedical Sciences* 2015;**54**: 72–80.
43. Greiner J, Rubin D. Causal effects of perceived immutable characteristics. *Rev Ec Stat* 2011; **93**:775–785.
44. Sen M, Wasow O. Race as a "bundle of sticks": designs that estimate effects of seeming immutable characteristics. *Annual Rev Polit Sci* 2016 (in press).
45. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiol* 2014; **25**:473-454.
46. VanderWeele, TJ and Robinson, WR. (2014). How to reduce racial disparities? Upon what to intervene? *Epidemiology*, **25**:491–493.
47. Valeri, L, Chen, JT, Garcia-Albeniz, X, Krieger, N., VanderWeele, TJ, and Coull, BA. The role of stage at diagnosis in colorectal cancer racial/ethnic survival disparities: a counterfactual causal inference approach. *Cancer Epidemiology, Biomarkers and Prevention*, 2016;**25**:83–89.