PROJECT MUSE®

Hill's Causal Considerations and the Potential Outcomes Framework

Tyler J. VanderWeele

Observational Studies, Volume 6, Issue 2, 2020, pp. 47-54 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2020.0008>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/793349/summary>

Hill's Causal Considerations and the Potential Outcomes Framework

Tyler J. VanderWeele

tvanderw@hsph.harvard.edu

Department of Epidemiology and Department of Biostatistics

Harvard University

Boston, MA, USA 02115

Hill's considerations for assessing causality (Hill, 1965) are arguably still important today, as indeed they were at the time of writing. The formalization of causal inference has, however, advanced considerably since Hill's paper. The potential outcomes or quantitative counterfactual framework (Rubin, 1974; Rosenbaum and Rubin, 1983a; Robins, 1986; Imbens and Rubin, 2015; Hernán and Robins, 2019) has clarified and made more precise our reasoning about causation and the underlying assumptions for much of causal inference. It has also allowed for dramatic expansion of the types of causal questions that can be addressed (Angist et al., 1995; Robins et al., 2000; Pearl, 2009; Hudgens and Halloran, 2008; Imai et al., 2010; Tchetgen Tchetgen and VanderWeele, 2012; VanderWeele, 2015). And it has moreover, I will argue below, altered how evidence concerning causation is often assessed in practice.

Within the potential outcomes framework, in its most basic form, we can let $Y(a)$ denote the potential outcome or counterfactual outcome that would have been observed for an individual if the exposure A had, possibly contrary to fact, been set to level a . We let $Y(1) - Y(0)$ denote the causal effect for an individual i.e. the difference in potential outcomes for that individual. We say that the covariates $\mathbf{C}$ suffice to control for confounding (or that “exchangeability holds conditional on $\mathbf{C}$ ” or “ignorability conditional on $\mathbf{C}$ ”) if the counterfactuals $Y(a)$ are independent of A conditional on $\mathbf{C}$ i.e. if within strata of $\mathbf{C}$, the group that actually had exposure status $A = a$ is representative of what would have occurred had the entire group with $\mathbf{C} = \mathbf{c}$ been given exposure $A = a$. We say that the assumption of consistency holds if whenever the actual exposure A equal level a , then $Y(a) = Y$. Under these assumptions of exchangeability and consistency we have that the average causal effect conditional on $\mathbf{C} = \mathbf{c}$ is given by $E[Y(1)|\mathbf{c}] - E[Y(0)|\mathbf{c}] = E[Y(1)|A = 1, \mathbf{c}] - E[Y(0)|A = 0, \mathbf{c}] = E[Y|A = 1, \mathbf{c}] - E[Y|A = 0, \mathbf{c}]$, where the first equality holds by exchangeability and the second by consistency. This final expression is just the association between exposure A and outcome Y on a difference scale. Under the assumptions of exchangeability and consistency, this association is equal to the conditional causal effect; under these assumptions, association implies causation.

The potential outcomes framework thus effectively employs deductive logic in its conclusion about causation. In contrast, as will be discussed further below, Hill's considerations or viewpoints, approach causation from an inductive approach. In this commentary, I will offer some discussion on how the potential outcomes framework has contributed to, or clarified, each of Hill's considerations, how it contrasts with those considerations by using deduc-

tive rather than inductive inference, and how it has thus also altered approaches to causal reasoning in practice.

1. Strength

Hill noted that, other things being equal, the greater the strength of association between exposure and outcome, the more plausible that the relationship is causal. The underlying reasoning is that a stronger association is more difficult to entirely attribute to explanations other than causality. This intuitive notion has been formalized through various sensitivity analysis approaches (Rosenbaum and Rubin, 1983b; Imbens, 2003; Hsu and Small, 2013). Indeed sensitivity analysis to assess sensitivity or robustness to unmeasured confounding has been around since before Hill’s paper (Cornfield et al., 1959). However, these approaches to sensitivity analysis have expanded dramatically since the introduction of the potential outcomes framework to observational studies, and have been extended to address other possible biases such as measurement error and selection bias (Greenland, 2005; Rothman et al., 2008; Lash et al., 2009). For some of these techniques, the relevant sensitivity analysis parameters that are required to explain away an association can be expressed as simple monotonic functions of the magnitude of the association between the exposure and the outcome (VanderWeele and Ding, 2017; Smith and VanderWeele, 2019; VanderWeele and Li, 2019). Such techniques reinforce the contention that strength of association is a critical consideration in assessing evidence for causation.

2. Consistency

Hill devoted considerable discussion to consistency, to arriving at the same conclusion repeatedly by different persons, in different places, circumstances, and times, and using different designs. Meta-analysis and systematic review is of course often useful in accumulating evidence over different persons, places, times and circumstances and much of the statistical development of meta-analysis proceeded independent of the potential outcomes framework. More recently, however, methods have been developed that employ the potential outcomes framework to combine meta-analytic approaches assessing consistency over studies, with the possibility of unmeasured confounding in those studies (Mathur and VanderWeele, 2019). Such approaches effectively tackle both the first and the second of Hill’s considerations simultaneously. In a different context, recent work on evidence factors provides another quantitative approach to assess robustness of evidence from different sources of data (Rosenbaum, 2010, 2017). Perhaps an even more profound contribution from the potential outcomes framework to Hill’s consideration of consistency arises from his point about consistency of conclusions across different designs. Hill explicitly references prospective and retrospective studies; however he also makes mention, in his discussion of nickel refiners, of what might be referred to today as an interrupted time series design. Indeed, the number of designs that employ assumptions other than “control for confounding by covariates” to assess causation has grown dramatically, often motivated by, grounded in, or made precise through the potential outcomes framework. These include instrumental variable approaches including Mendelian randomization, regression discontinuity designs, difference-in-difference methods, and interrupted time series designs (Angist et al.,

1995; Didelez and Sheehan, 2007; Angrist and Pischke, 2009; Morgan and Winship, 2015). Each of these employ different assumptions and consistency of evidence across these different designs will render more plausible the conclusion of causality. There has been recent discussion of this as a process of “triangulation” (Lawlor et al., 2016). However, it should also be noted that each of these designs also often identifies a different causal effect, for different subpopulations, and thus discrepant results across these different study designs do not necessarily imply a contradiction, or that one of them is wrong, and thus, in words of Hill “different results of a different inquiry certainly cannot be held to refute the original evidence.” Nevertheless, the potential outcomes framework has arguably dramatically expanded the different ways to look for consistency of evidence, and has also clarified when discrepant results are or are not problematic.

3. **Specificity**

Hill indicated that if the association with an exposure is specific to a particular outcome (or is considerably stronger for that outcome) then this may render a causal conclusion more plausible. Recent work on negative controls has formalized this notion within a potential outcomes framework (Tchetgen Tchetgen, 2013) and such more formal quantitative approaches to leveraging specificity, while perhaps not yet widely used, could potentially contribute a great deal to our causal reasoning. However, as with consistency, so also with specificity, Hill notes, “if it is not apparent, we are not thereby necessarily left sitting irresolutely on the fence.”

4. **Temporality**

Causes of course must precede effects. However, Hill’s discussion of the temporality consideration in fact concerns ruling out the possibility of reverse causation i.e. the possibility that the supposed outcome might in fact precede and subsequently alter the variable taken as the exposure. This temporality or “reverse-causation” consideration is mirrored in calls within the potential outcomes literature to use longitudinal data to assess evidence for causation in observational studies and to control for baseline outcome to rule out that possibility of reverse causation (VanderWeele et al., 2016). Such control will often effectively be needed to ensure the “no-confounding” or “ignorability” assumption within the potential outcome approach does indeed hold.

5. **Biological Gradient**

Hill argues that in some cases a causal relationship may be more plausible if there is a monotonic dose-response relationship. Some discussion of when such dose-response relationships may render evidence for causality more plausible has been formalized using potential outcomes (Rosenbaum, 2003). However, here too it must be acknowledged that the absence of a monotonic dose-response relationship is not necessarily evidence against causality. Causal relationships can plausibly be U-shaped, for example.

6. **Plausibility**

Hill writes “It will be helpful if the causation we suspect is biologically plausible.” However, Hill does not place a great deal of weight on this consideration in his discussion because it is entirely possible that the relevant biological knowledge may be

lacking. Certainly one consideration with regard to biological plausibility concerns knowledge of the mechanisms. The potential outcomes framework has supplied a set of concepts and methods, and has clarified the underlying assumptions, to quantitatively assess the relative importance of various proposed mechanisms (Robins and Greenland, 1992; Pearl, 2009; Imai et al., 2010; VanderWeele, 2015), or, said another way, to assess the extent to which each potential mechanism might mediate the effect. Unfortunately, the assumptions required by these approaches are in fact generally considerably stronger than those required to establish causation between the exposure and the outcome itself. These strong assumptions are needed for quantification. However, from a more qualitative perspective the logic of mediation may still be of help for contributing biological plausibility: if we have evidence that the exposure alters particular mechanisms, and also separate evidence that the mechanisms alter the outcome, then this renders more plausible that the exposure itself alters the outcome (VanderWeele, 2015).

7. Coherence

Hill places some emphasis on the coherence with existing knowledge, that “the cause-and-effect interpretation of our data should not seriously conflict with the generally known facts.” This is of course an important principle. It is perhaps one that has not as often been explicitly and quantitatively exploited within the counterfactual framework. However, consideration of “triangulation” may come into play here as well (Lawlor et al., 2016).

8. Experiment

Of course experimental evidence from a trial randomizing the exposure can contribute substantially to the case for causality. The potential outcomes framework helps make clear the logic, but more profoundly, the theory and methods that have developed from the potential outcomes framework has been useful in evaluating the strength of evidence from randomized trials when something goes wrong e.g. when there is non-compliance, or attrition, or censoring due to death (Robins, 1986; Angrist et al., 1995; Scharfstein et al., 1999; Frangakis and Rubin, 2002).

9. Analogy

Hill’s consideration of analogy (e.g. that when we have evidence for a causal effect of one exposure of some class on an outcome of a particular class, it may render causality more plausible between other members of the same exposure class and outcome class) does not have a direct counterpart within the potential outcomes framework of which I am aware. There are perhaps some indirect resonances with certain questions concerning surrogacy. For example, when it is known from randomized trials that a particular drug has beneficial effects on a health outcome, then when is it reasonable to draw conclusions using a randomized trial for the effect of a different drug, but in the same drug class, if data is only available on a surrogate, rather than the primary outcome of interest? Such questions have been formalized using counterfactual approaches (Frangakis and Rubin, 2002; Gilbert and Hudgens, 2008; Joffe and Green, 2009; VanderWeele, 2013), but there is almost always some reasoning by analogy in

such cases when these approaches are employed in practice. Such reasoning by analogy can of course go wrong, in general, and of course with surrogacy as well, and there have been a number of disastrous failures concerning such reasoning by analogy with surrogates (Moore, 1995; Fleming and DeMets, 1996). Caution must be exercised.

Inductive vs. Deductive Reasoning

Hill’s conclusion is that these “are nine different viewpoints from all of which we should study association before we cry causation.” He believes that “None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*.” Hill’s considerations for assessing evidence for causation essentially relies on inductive reasoning. Hill states that the “fundamental question” is whether there is “any other way of explaining the set of facts before us, is there any other answer equally, or more likely than cause and effect?” and he believes the nine viewpoints are helpful in this type of inductive reasoning. As I have argued above, I believe these viewpoints are indeed still helpful from an inductive standpoint.

The potential outcomes framework, in contrast, follows a deductive form of logic to reach the conclusion of causation. Under the potential outcomes framework, under a certain set of assumptions (e.g. sometimes stated as “exchangeability, consistency, and positivity” (Hernán and Robins, 2019) or as “stable unit treatment value assumption and ignorable treatment assignment” (Imbens and Rubin, 2015)), then any actual association implies causation. While the potential outcomes framework in principle employs deductive logic to reach the conclusion of causation, in practice, there are still inductive elements. We are never sure that the premises hold. In fact, we often are sure that they do not exactly hold and must use sensitivity analysis to assess whether our conclusions are robust to violations. We also often cannot definitively conclude that an association is present; rather, we use statistical inference to assess the evidence for an association. We are still left only with evidence, and must evaluate whether, and to what extent, that evidence justifies a conclusion about causation. There is thus an inductive element to the application of the potential outcomes approach in practice. Nevertheless, the deductive form of logic underlying the potential outcomes approach itself simplifies the inductive process. Because we understand the logic relating premises to conclusions, we can carefully evaluate each premise. Rather than having to rule out all possible alternative explanations to a causal relation that either we, or someone else, may come up with, we instead can focus more narrowly on each premise that is used by the potential outcomes framework to deduce causation and can carefully evaluate these premises. We can often do so quantitatively by means of sensitivity analysis (Rosenbaum and Rubin, 1983b; Imbens, 2003; Hsu and Small, 2013). This can sometimes lend considerable confidence to the conclusions that are drawn, and can also indicate when caution may be needed. However, when such sensitivity analyses for unmeasured confounding, and measurement error, and selection bias (Greenland, 2005; Rothman et al., 2008; Lash et al., 2009) do indicate substantial robustness, the confidence with which we can draw causal conclusions may often be considerably greater than if we had had, in a more purely inductive approach, to consider every other non-causal alternative explanation for an association.

It is my view that it is this move of only needing to evaluate the premises imbedded in the potential outcomes framework’s logic to draw causal conclusions that has altered, for many, the form of reasoning and the types of evidence considered when assessing causal claims in practice. The focus has shifted from Hill’s considerations to the premises characteristic of potential outcomes logic. It has simplified our reasoning. This has been, in my view, an advance, and has resulted in real changes in practice with regard to how causal evidence is assessed. Nevertheless, with simplified reasoning can come also the danger of complacency. Because the deductive logic is relatively straightforward, we may be tempted to no longer even carefully evaluate the premises. Hill’s considerations can still be helpful today in countering that complacency, by looking at the association through a different, more inductive, lens and thinking creatively about what alternative explanations might be. However, if we return to the potential outcomes framework itself, then I think the central tool in confronting complacency over the assumptions is sensitivity analysis. As described above, sensitivity analysis is the principal means by which the premises fundamental to the potential outcomes approach can be evaluated. As I have written elsewhere (VanderWeele and Ding, 2017; VanderWeele et al., 2019), I believe these should be routine in analysis of observational data and should *always* take place when such data is being used to argue for a conclusion of causality. In attempt to help facilitate this, I have elsewhere proposed a set of extremely simple easy-to-use sensitivity techniques that do not require sophisticated software resources or a more technical background in statistics and that are applicable to unmeasured confounding, selection bias, and measurement error (VanderWeele and Ding, 2017; Smith and VanderWeele, 2019; VanderWeele and Li, 2019); they can effectively be implemented by hand, and are relatively easy to interpret. There are many other good sensitivity analysis techniques available, but these sometimes require considerably more effort to implement. My hope is that these simple techniques leave empirical researchers without any excuse for not carrying out what is at least a very simple sensitivity analysis. Such sensitivity analysis, in conjunction with the simple deductive logic of the potential outcomes framework, can, at least sometimes, establish very strong evidence indeed for a causal conclusion.

References

- Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, 91:444–472.
- Angrist JD, and Pischke J-S. (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press.
- Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., and Wynder, L. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, 22:173–203.
- Didelez, V., and Sheehan, N. (2007). Mendelian randomization as an instrumental variable approach to causal inference. *Statistical Methods in Medical Research*, 16(4), 309–330.
- Fleming, T. R. and DeMets, D. L. (1996). Surrogate end points in clinical trials: Are we being misled? *Annals of Internal Medicine*, 125, 606–613.

- Frangakis, C. E., and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1), 21-29.
- Gilbert, P. B., and Hudgens, M. G. (2008). Evaluating candidate principal surrogate end-points. *Biometrics*, 64(4), 1146-1154.
- Greenland, S. (2005). Multiple-bias modeling for analysis of observational data (with discussion). *Journal of the Royal Statistical Society Series A*, 168:267–308.
- Hernán MA, Robins JM (2019). *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming.
- Hill, A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 295-300.
- Hsu, J. Y., and Small, D. S. (2013). Calibrating sensitivity analyses to observed covariates in observational studies. *Biometrics*, 69(4), 803-811.
- Hudgens, M. G., and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482), 832-842.
- Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4):309–334.
- Imbens, G., and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.
- Joffe, M. M., and Greene, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics*, 65(2), 530-538.
- Lash, T.L., Fox, M.P., and Fink, A.K. (2009). *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York: Spring.
- Lawlor, D. A., Tilling, K., and Davey Smith, G. (2016). Triangulation in aetiological epidemiology. *International Journal of Epidemiology*, 45(6), 1866-1886.
- Mathur, M. B., and VanderWeele, T. J. (2019). Sensitivity analysis for unmeasured confounding in meta-analyses. *Journal of the American Statistical Association*, in press, <https://doi.org/10.1080/01621459.2018.1529598>
- Moore, T. (1995). *Deadly Medicine: Why Tens of Thousands of Patients Died in America's Worst Drug Disaster*. New York:Simon and Schuster.
- Morgan, S. L., and Winship, C. (2015). *Counterfactuals and Causal Inference*. Cambridge, UK: Cambridge University Press, 2nd Edition.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd Edition. Cambridge: Cambridge University Press.
- Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure period – Application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393–1512.
- Robins, J. M., and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 143-155.
- Robins J.M., Hernán M.A. and Brumback B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550-560.
- Rosenbaum, P. R. (2003). Does a dose-response relationship reduce sensitivity to hidden bias?. *Biostatistics*, 4(1):1-10.
- Rosenbaum, P. R. (2010). Evidence factors in observational studies. *Biometrika*, 97(2), 333-345.

- Rosenbaum, P. R. (2017). The general structure of evidence factors in observational studies. *Statistical Science*, 32(4), 514-530.
- Rosenbaum, P.R. and Rubin, D.B. (1983a). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983; 70:41-55.
- Rosenbaum, P. R., and Rubin, D. B. (1983b). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2), 212-218.
- Rothman KJ, Greenland S, Lash TL. (2008) *Modern Epidemiology*. 3rd Edition. Lippincott.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66 :688–701.
- Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94(448), 1096-1120.
- Smith, L. and VanderWeele, T.J. (2019). Bounding bias due to selection. *Epidemiology*, 30(4), 509-516.
- Tchetgen Tchetgen, E. (2013). The control outcome calibration approach for causal inference with unobserved confounding. *American Journal of Epidemiology*, 179(5), 633-640.
- Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research – Special Issue on Causal Inference*, 21:55–75.
- VanderWeele, T. J. (2013). Surrogate measures and consistent surrogates (with Discussion). *Biometrics*, 69(3), 561-581.
- VanderWeele, T.J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.
- VanderWeele, T.J., Jackson, J.W., and Li, S. (2016). Causal inference and longitudinal data: a case study of religion and mental health. *Social Psychiatry and Psychiatric Epidemiology*, 51:1457-1466.
- VanderWeele, T.J. and Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of Internal Medicine*, 167:268-274.
- VanderWeele, T.J. and Li, Y. (2019). Simple sensitivity analysis for differential measurement error. *American Journal of Epidemiology*, in press. <https://doi.org/10.1093/aje/kwz133>
- VanderWeele, T. J., Mathur, M. B., and Chen, Y. (2019). Outcome-wide longitudinal designs for causal inference: A new template for empirical studies. *Statistical Science*, in press.