PROJECT MUSE®

Plausibility and the Benefit of Theoretical Reasoning:
Comment on Austin Bradford Hill (1965)

A. James O'Malley

Observational Studies, Volume 6, Issue 2, 2020, pp. 20-23 (Article)

Published by University of Pennsylvania Press

DOI: <https://doi.org/10.1353/obs.2020.0004>

➔ *For additional information about this article*

<https://muse.jhu.edu/article/793345/summary>

Plausibility and the Benefit of Theoretical Reasoning: Comment on Austin Bradford Hill (1965)

A. James O'Malley

James.OMalley@dartmouth.edu

Department of Biomedical Data Science and
The Dartmouth Institute of Health Policy and Clinical Practice
Geisel School of Medicine, Dartmouth University
Lebanon, NH 03756, USA

I greatly enjoyed reading the paper “The Environment and Disease: Association or Causation” by Professor Austin Bradford Hill (ABH). The paper continues to be held in high regard and is massively cited (9,359 citations as of October 20, 2019). The paper is centered around nine aspects of statistical association that ABH suggests be considered before deciding if causation should be claimed. In some fields, these points have erroneously been viewed and taught as causal criteria (Phillips and Goodman 2004). Given that “Association or Causation” discernment is receiving increasing attention in the statistical literature, and that we’re in the midst of the emergence of data science and ever-growing volumes of observational data, reviewing ABH (1965) and assessing whether there are insights that can inform contemporary statistical analysis is timely and valuable.

ABH (1965) has been the subject of much review and discussion over many years (e.g., Philips and Goodman 2004; Thygesen, Andersen and Andersen 2005; Fedak et al 2015). Often reviews have systematically considered each of the nine points and offered discussion and critique about each. Because the appropriateness or not of these points has already received an extensive amount of attention, I have focused my comment on a specific topic that while being inherent to ABH (1965) has not been directly addressed. I focus on the role of theoretical models – often used in economics and sociology to generate causal stories and allied hypotheses prior to analyzing data or even designing a study – in relation to the task of distinguishing cause from association. Is the ability to construct and exploit theoretical models in order to devise ways of testing their legitimacy an under-appreciated skillset that ought to be part of statistical (and data science) practice and education? In the following, I first provide some general comments and then give an illustrative example.

The point in ABH (1965) that is most directly relevant to my comment is Plausibility, one of the least considered of the nine points. ABH states “It will be helpful if the causation we suspect is biologically plausible.” I agree with this statement and think that it can be extended to “It will be helpful if the causation we suspect is plausible in the scientific or other setting being analyzed.” In other words, does the causal story make sense given current knowledge. In order to answer this question, it is imperative that one understand the underlying science or subject area knowledge. If the causal story is plausible then the skill of being able to think theoretically or conceptually about it may allow tests of hypotheses and ways of identifying parameters in the theoretical model to be constructed that data can subsequently adjudicate.

Since 1965, the presence and importance of statistics in science and medicine has grown. However, in many settings, roles of statisticians have been restricted to data analysis as opposed to the art of linking the broader scientific perspective underlying research or other investigation to data. To prepare students for such data-focused careers, the budding statistician is typically supplied with practice problems to solve. This tailor-made approach bypasses a lot of the artfulness of being able to reason theoretically in the application-side of a problem to identify the key aspects of it that can be tested by data. Thus, even if their empirical work is performed with much creativity and skill, in terms of overall impact the role of a traditionally-trained statistician may not be as great as it could be.

The statistician is often a “Devil’s advocate”. While being a Devil’s advocate is certainly needed, if the ultimate goal is to get the best answer possible using current resources, the skill of being able to discern what should be tested in the first place is vital. I think of the construction of a theoretical model based on current knowledge as constructive causal thinking. This process can enlighten one of the data that will be most useful to acquire in order to strengthen a claim of causality and of what identification strategies can be used. I advocate for the involvement of statisticians in the role of constructive causal thinking as well as a Devil’s advocate.

In methodological statistics it is common to assume certain complications don’t exist in order to focus on the conditions for which a particular result applies. Yet, from a purist standpoint, observational data analysis is inherently doomed. By its nature, there is always an assumption which, if violated, leads to bias as there is no way of ensuring that all assumptions, particularly those involving unobserved variables, hold. Therefore, it seems important to pair methodological work on a ready-made or stylized problem with the ability to identify the key problem to work on in the first place. This encompasses being able to justify why a question is important, why assumptions are plausible, recognizing and constructing sanity checks to inform the validity of a model, and so on. Working through these steps may help enable the methodological problem that has the biggest impact to be realized.

Technical approximation is another critical but often untaught skill. It is a means of assessing what is a reasonable “ball-park” answer to a numerical problem. The concept can be generalized to the ability to construct a reasonable ball-park theoretical model that can in turn be used to assess whether the stated question of interest is the right one and whether a proposed analysis answers the question of interest. In practice, the best research questions and empirical strategies can be ingenious for their ease of understanding. Exceedingly complicated solutions are seldom universally popular. Explanations involving elegant sanity checks based on a plausible theoretical model can be far more convincing than the output of uninformed but complex empirical analyses.

To conclude the first part of my comment, many social science fields (e.g., economics, sociology) place substantial value on theoretical models that develop hypotheses and allied identification strategies. I believe that the skill of being able to theoretically construct relationships between variables and testable constructs is underutilized in statistics. A general example of this skill is seen in the identification of an instrumental variable (IV); because IVs cannot be fully tested by the data, their construction is a skill that relies on broad and artful thinking.

To illustrate theoretical (non-empirical) reasoning and its benefits, consider the tie-directionality identification strategy developed in Christakis and Fowler (2007) to account for unmeasured common causes in the estimation of peer-effects (the effect of one individual on another). This example features the points in ABH (1965) concerning Plausibility, Analogy and Dose-response.

Christakis and Fowler (2007) hypothesized that a health behavior or characteristic (e.g., obesity or body-mass-index) of an individual is affected by the same trait in their peers. One form of bias that may afflict the estimator of a peer-effect is from unmeasured common causes – uncontrolled factors that impart similar effects on a group of individuals. Unmeasured common causes may give the appearance that the change in the outcome of one individual tracks that of other individuals, even if there is no actual peer effect.

If there exist dyads for which there should be no inter-individual influence, then such dyads can be used as a control group whose estimated peer effect captures any unmeasured common causes (Christakis and Fowler 2007). Under the homogeneity assumption that all types of dyads are equally affected by exposure to an unmeasured common-cause, differencing a peer-effect against that for the control dyad group is able to recover the true peer effect. This is an example of argument by Analogy. Christakis and Fowler (2007) also theorized that if peer effects exist they should increase with the strength of relationship between individuals. This led them to theorize that in the case of mutable relationships such as friendships, the peer association will be greatest in dyads for which the friendship is mutual, second-greatest along network ties when friendship is only from the focal actor to the other actor, third-greatest along network ties when friendship is only from the other actor to the focal actor, and least in dyads with no relationship (null dyads). (The “focal actor” is the individual providing the dependent variable.) A requirement that peer-effect estimates follow this theoretically-implied ordering of effect-sizes is a more rigorous test of the existence of peer-effects than a single comparison of the estimated peer-effects for non-null and null dyads as the former encapsulates all hypotheses simultaneously.

Although all sources of unmeasured confounding are not guarded against, which is impossible in an observational study, the theoretical derivation and implied empirical strategy for testing the peer-effects hypothesis is appealing for featuring Plausibility, Analogy and Dose-response. The underlying theoretical model makes sense, is easily understood and can be constructively debated, which is what has transpired. For example, in a subsequent paper, Shalizi and Thomas (2011) introduced a form of confounding, known as latent ignorability, which is not accounted for by the Christakis and Fowler approach. They argued that individuals may form relationships because they have similar characteristics that continue to affect their outcomes post tie-formation and thus that latent ignorability is indistinguishable from the true peer-effects in the Christakis and Fowler study. They embellished their theory with a compelling example of how latent homophily could occur. By articulating the theoretical model or causal mechanism clearly, both the original paper and subsequent critiques written about it can triggered valuable discussion and insights resulting in further advances in understanding. Plausibility assessment and theoretical reasoning are important skills!

Another important point implied by ABH (1965) is that one should not hold back from publishing results in order to wait for every contingency to be accounted for. It is better to advance knowledge with results once available as long as reports and publications are

accompanied with clear descriptions of the theoretical model, assumptions and identification strategy. The final section of ABH (1965), “The Case for Action,” argues that evidence is there to be judged on its merits and should be independent of what hangs on it. This is consistent with the notion that while a great idea isn’t foolproof, it might still be worthwhile to pursue. This progressive approach allied with full acknowledgement and explanation of assumptions and an adept sensitivity analysis that evaluates the impact of plausible violations of the assumptions (another invocation of the wisdom of ABH (1965)) supports fully-informed and timely decision making.

While the technical correctness of the points made in ABH (1965) have been extensively debated, I believe that ABH (1965) remains a valuable paper to read and that statisticians can benefit from doing so, including by embracing the importance of plausibility and related points. Social scientists have practiced the art of developing and applying theoretical models for many years. I conjecture that being able to evaluate theoretical models for their plausibility and apply them to construct identification strategies will help to increase the prominence of statisticians in science, medicine and policy-formation (and other areas of society).

References

- Christakis, N.A., Fowler, J.H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357:370–379.
- Fedak, K.M., Bernal, A., Capshaw, Z.A., Gross, S. (2015). Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerging Themes in Epidemiology*, 12:14.
- Hill, A.B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58:295–300.
- Phillips, C.V and Goodman, K.J. (2004). The missed lessons of Sir Austin Bradford Hill. *Epidemiological Perspectives and Innovations*, 1(1):3.
- Shalizi, C.R., Thomas, A.C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods and Research*, 40(2):211–239.
- Thygesen, L.C., Andersen, G.S., Andersen, H. (2005). A philosophical analysis of the Hill criteria. *Journal of Epidemiology and Community Health*, 59(6):512-6.