The Messed Lessons of Sir Austin Bradford Hill

Carl V. Phillips, MPP PhD
Epiphi Consulting Group, www.epiphi.com
Minnesota Center for Philosophy of Science, University of Minnesota
phillips@ephphi.com

Karen J. Goodman, PhD
University of Texas-Houston School of Public Health
Epiphi Consulting Group, www.epiphi.com
kgoodman@sph.uth.tmc.edu

Poster #765, Congress of Epidemiology, Toronto, 15 June 2001

Classic paper:

Austin Bradford Hill. The Environment and Disease: Association or Causation? Proceedings of the Royal Society of Medicine 58:295-300, 1965.

Frequency of citation:

Very high *(440 citations between 1996-2001 in Web of Science citation databases, plus frequently invoked without citation).*

Frequency of actually reading:

Apparently fairly low.

Reason it is one of the most cited papers in health research:

Source of what has become known as the "Bradford-Hill criteria" for inferring causation from epidemiologic associations.

What Hill called his list of nine considerations:

"...viewpoints from...which we should study association before we cry causation."

*What Hill **did not** call his list of nine considerations:*

Criteria; hard and fast rules.

Our guess about how Hill would feel about his "criteria" remaining popular in the 21st century:

Appalled.

What this classic paper should be remembered for:

One of the most succinct lessons in the literature, presented in a few paragraphs of concluding remarks, that effectively summarizes how we should think about scientific uncertainty, interpretation of study results, tradeoffs, and policy decisions.

Opportunity to completely reinterpret the lessons of a classic paper:

Priceless.

The Wrong Lesson Learned: The "Bradford-Hill Criteria"

- The "criteria" have been effectively criticized as being of limited value in causal inference.[1]
- No set of criteria can identify a cause-effect relation with certainty.
- Empirical analysis has shown that reviewers of epidemiologic evidence interpret and apply lists of causal criteria in inconsistent ways.[2]
- Serious attention is being devoted to the notion of causality in epidemiology, offering approaches that supplant appeals to highly imperfect rules of thumb.[3-6]

The Right Lessons Missed

Amidst the misguided attention to Hill's list of considerations is a lack of attention to the last page and a half of the article and its insufficiently heeded lessons:

- Warning against the over-emphasis on statistical significance testing and claims of statistical precision in the face of systematic uncertainty.

- Guiding principles for formulating policy decisions on the basis of epidemiologic observations.

Over-emphasis on statistical significance testing

> "*the glitter of the t table diverts attention from the inadequacies of the fare*" (Hill, p. 299)

Hill lamented that the justifiable criticisms of drawing conclusions from inadequate study samples had led to a problematic backlash:

- Overemphasis on tests of statistical significance as a basis for causal inference,
- Sources of systematic error (bias) were frequently overlooked in drawing causal conclusions.

This warning is at least as relevant today as it was in 1965.

- The outpouring of thousands of "significant" but often contradictory results confuses the public and decision makers.
- Despite the education of several new generations of researchers, and the advances in thinking listed below, there has been little improvement in practice.

Modern advances in thinking about statistical inference

Improved representation of random sampling error

- An emphasis on the p-value alone, has been soundly denounced, though the practice has hardly been eliminated.
- Current textbooks recommend against statistical hypothesis testing that dichotomizes the p-value into significant or not.
- Modern textbooks and practice recognize that the confidence interval offers the advantage of distinguishing imprecise estimates (poor statistical power) from apparent null associations, and associations with large magnitudes from those that are small but statistically significant by virtue of a large sample.

Quantification of systematic error

- No statistical test based solely on random sampling error informs us about uncertainty due to sources of bias.
- Hill hints at this when he notes that one of his own studies had great potential for a particular systematic error. In effect, he asks *"why would I bother to do an exaggeratedly precise statistical test when I know that the other sources of error are likely so large?"*
- Hill points out the folly of mistaking statistical precision for validity; recent work improves upon this, by introducing methods to quantify systematic errors.[7-12]

Formulating Policy Decisions from Epidemiologic Observations

"Our object is usually to take action" (Hill, p. 300)

Largely overlooked is Hill's analysis of what should flow from causal inferences.

Epidemiology has its roots in specific policy questions (e.g., "can we do something about cholera outbreaks?").

But modern epidemiologists have ambivalent attitudes towards the policy decisions associated with their research. Researchers often justify costly research based on its practical benefits, but deny the need to assess the policy contributions, defending the value of the science for its own sake.

When policy implications are presented, they are seldom carefully analyzed. (One major journal forbids policy recommendations tacked on to original research reports, suggesting that policy analysis is too complex and serious to be an afterthought by researchers whose expertise lies elsewhere.)

Judging from Hill's comments, he might call for more careful policy analysis attached to epidemiology research, rather than none at all (though it is not clear he would have any solution to the challenge of fitting such analyses into the standard 3000-word, single-finding health research paper).

Lessons in tradeoffs

We need *"differential standards before we convict"* (Hill, p. 300)

Hill argues (as any policy analyst or economists would) that we ought to pay attention to the costs and benefits associated with an exposure.

The level of certainty regarding the cause-effect relation that is sufficient for policy action should depend on the cost-benefit tradeoff associated with the action, rather than merely the degree of certainty that there would be *some* benefit.

(Those who might dismiss this as socio-political concerns interfering with proper science should note that Hill clearly states that the science per se, including the data analysis, should not be influenced by what is at stake. Rather, public health researchers should recognize that the stakes matter, rather than ignoring such concerns and leaving them to a rather unreliable post-science political process.)

Hill's examples of tradeoff-based decisions

> *"on relatively slight evidence we might decide to restrict the use of a drug for early-morning sickness in pregnant women. If we are wrong in deducing causation from association no great harm will be done. The good lady and the pharmaceutical industry will doubtless survive."* (Hill, p. 300)

(Bonus lesson: A bad choice of example can make a good point look bad.)

Setting aside the impolitic claim and the unsupported assertion that there is no great harm at stake (as well as the irony of the rise, fall, and current possible reemergence of the morning sickness drug, Bendectin) the underlying point is a good one.

The subsequent examples are better:

> *"On fair evidence we might take action on what appears to be an occupational hazard, e.g. we might change from a probably carcinogenic oil to a noncarcinogenic oil in a limited environment and without too much injustice if we are wrong. But we should need very strong evidence before we made people burn a fuel in their homes that they do not like or stop smoking the cigarettes and eating the fats and sugar that they do like."* (Hill, p. 300)

The Most Important Lesson: Improving Epidemiology-Based Decision Making

> *"All scientific work is incomplete -- whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time."* (Hill, p. 300)

Making a good decision does not depend on having studies with confidence intervals that exclude the null.

A best decision can be made based on whatever information we have now.[11-12]

Pursuit of the low p-value,

- leaves our society postponing apparently useful policy choices while we do more research to try to show what we already believe to be true,
- creates the incentive to use dubious methods to squeeze out significant results,
- creates an opportunity for those who would prefer to find no association to make real causal relationships disappear below the 95% horizon.

If epidemiologists helped empower policy makers to, for example, ban an easily-replaced chemical with about a 50-50 chance it is a health hazard (based on an honest assessment of all uncertainty), then the payoff for fiddling with the data to show the certainty is a bit higher or a bit lower would be eliminated.

We can escape from the trap of letting ignorance trump knowledge

> *"In asking for very strong evidence I would, however, repeat emphatically that this does not imply crossing every 't', and swords with every critic, before we act."* (Hill, p. 300)

Regulators often fail to act because we have not yet statistically "proven" an association between an exposure and a disease, even when there is enough evidence to strongly suspect it.

There is a growing movement to make a similar mistake by adopting precautionary principles, calling for regulation until we have "proven" a lack of causal association.

But if we can escape from the false "proven vs. not proven" dichotomy that statistical tests create, and from the notion that causality can be definitively inferred from a list of criteria, we can make decisions based on what we do know rather than what we don't.

Conclusions

Appealing to the untested "criteria" of a great luminary from days past are reminiscent of the "scientific" methods of the Dark Ages.

Judging from his own caveats, we believe Hill would likely agree.

He deserves credit for roughing out some rules of thumb for a very young science that surely needed them at the time.

But long after that list ought to have become no more than an historical note, it generates uncritical adherence while his recommendations about statistical tests and policy context remain key unlearned lessons.

References

1. Rothman KJ, S Greenland. Causation and causal inference. Chapter 2, KJ Rothman and S Greenland, eds., *Modern Epidemiology, 2nd Edition*, pp. 7-28, 1998.
2. Weed DL, LS Gorelic. The practice of causal inference in cancer epidemiology. *Cancer Epidemiology, Biomarkers & Prevention* 1996;5:303-311.
3. Maldonado G, S Greenland. Estimating causal effects. *Int J Epidemiol*, in press.
4. Greenland S, J Pearl, JM Robins. Causal diagrams for epidemiologic research. *Epidemiology* 1999;10(1):37-48.
5. Greenland S. Causality theory for policy uses of epidemiologic results. In Murray CJL, Mathers C, Salomon J, Lopez AD, and Lozano R, *Summary Measures of Population Health*, Cambridge: Harvard University Press/World Health Organization, in press.
6. Pearl J. *Causality*. Cambridge: Cambridge University Press, 2000.
7. Phillips CV. Applying fully-articulated probability distributions. *Am J Epidemiol* 2000;151(11):S41.
8. Phillips CV, G Maldonado. Quantifying uncertainty in epidemiology. Manuscript under review.
9. Rothman KJ, S Greenland. Approaches to statistical analysis. In KJ Rothman and S Greenland, eds., *Modern Epidemiology, 2nd Edition*, 1998, pp. 183-199. Philadelphia: Lippincott-Raven.
10. Phillips CV, G Maldonado. Using Monte Carlo Methods to Quantify the Multiple Sources of Error in Studies. *Am J Epidemiol* 1999;149(11): S17.
11. Phillips CV. Our Estimates are Uncertain, But That is OK. *Am J Epidemiol* 2000;151(11): S41.
12. Phillips CV. The Economics of "More Research is Needed". *Int J Epidemiol* 2001, in press.