

Learning from and responding to statistical criticism*

Andrew Gelman[†]

15 Mar 2017

Irwin Bross's article, "Statistical Criticism," gives advice that is surprisingly current, given that it appeared in the journal *Cancer* nearly sixty years ago. Indeed, the only obviously dated aspects of this paper are the use of the generic male pronoun and the sense that it was still an open question whether cigarette smoking caused lung cancer.

In his article, Bross acts a *critic of criticism*, expressing support for the general form but recommending that critics go beyond hit-and-run, dogmatism, speculation, and tunnel vision. This all seems reasonable to me, but I think criticisms can also be taken at face value. If I publish a paper and someone replies with a flawed criticism, I still should be able to respond to its specifics. Indeed, there have been times when my own work has been much improved by criticism that was itself blinkered but which still revealed important and fixable flaws in my published work.

I would go further and argue that nearly all criticism has value. Again, I'll place myself in the position of the researcher whose work is being slammed. Consider the following sorts of statistical criticism, aligned in roughly decreasing order of quality:

- A thorough, comprehensive reassessment. Of course this is valuable: if an expert goes to the trouble of (a) finding problems in my work, (b) demonstrating that my errors were consequential, and (c) providing an alternative, then this is a clear step forward.
- A narrow but precise correction. If I have made a mistake in data processing or analysis, or if I have missed some alternative explanation for my findings, I would like to know. Even if it turns out that my error did not affect my main conclusions, it will be helpful to myself and to future researchers to fix the immediate problem.
- Identification of a potential problem. What if someone criticizes one of my published papers by suggesting a problem without demonstrating its relevance? This can be annoying, but I don't see the problem with the publication of such a criticism: Readers should be made aware of this potential problem, and future researchers can explore it.
- Confusion. All too often, criticism reveals a misunderstanding on the part of the critic. But this can have value too, in revealing that I have failed to communicate some point in my original article. We can't hope to anticipate all possible misreadings of our work, but it is good to take advantage of opportunities to clarify.
- Hack jobs. Bross was writing about cancer studies, an area where cigarette companies paid big money for several decades to highly-credentialed M.D.'s and Ph.D.'s to criticize epidemiological research using any and all arguments at hand. It can be hard to deal with criticism that is motivated by a desire to muddy the waters rather than to get

*Discussion of the paper "Statistical Criticism" by Irwin Bross, for *Observational Studies*. We thank Dylan Small for inviting this paper.

[†]Department of Statistics, Columbia University, New York.

at the truth. Hack criticism can indeed have negative value, and the problem here is not so much in the criticism itself—after all, even a hack can make a good point, and hacks will use legitimate arguments where available. Rather, the hack problem comes in the critical *process*: a critic who aims at truth should welcome a strong response, while a hack will be motivated to avoid any productive resolution.

Another way to see the value of post-publication criticism, even when it is imperfect, is to consider the role of *pre*-publication review. It is perfectly acceptable for a peer reviewer to raise a narrow point, to speculate, or to point out a potential data flaw without demonstrating that the problem in question is consequential. Referees are encouraged to point out potential concerns, and it is the duty of the author of the paper to either correct the problems or to demonstrate their unimportance. Somehow, though, the burden of proof shifts from the author (in the pre-publication stage) to the critic (after the paper has been published). It is not clear to me that either of these burdens is appropriate. I would prefer a smoother integration of scientific review at all stages, with pre-publication reports made public and post-publication reports being appended to published articles.

Overall, I am inclined to paraphrase Al Smith and reply to Bross that the ills of criticism can be cured by more criticism. That said, I recognize that any system based on open exchange can be hijacked by hacks, trolls, and other insincere actors. The key issues in dealing with such people are economic and political, not statistical, but we still need to be able to learn from and respond to statistical criticisms, whatever their source.