

Discussion of “Statistical Inference: The Big Picture” by R. E. Kass

Steven N. Goodman

In this piece, Rob Kass brings to bear his insights from a long career in both theoretical and applied statistics to reflect on the disconnect between what we teach and what we do. Not content to focus just on didactic and professional matters, the focus of his 2009 article (Brown and Kass, 2009), in this commentary he proposes a remake of the foundations of inference. He proposes to replace two fundamental “isms”—frequentism and Bayesianism—with a new “ism”—“pragmatism;” an approach that he puts forward as more ecumenical and practical, enshrining in foundations what good statisticians already do.

There is a lot to commend in this piece, particularly the emphasis on the subjunctive nature of all model-based inference, and I am sure the other commentators will do justice to its strengths. But in spite of its clarity and initial promise, I found Kass’s proposal ultimately unsatisfying. It seems less a new foundational philosophy than a call for a truce, one of many over the years. It is telling that all of the examples show practical equivalence between Bayesian and frequentist estimates, so the biggest stakes here seem to be what people think, not what they do. The difficulty with “big tent” foundations is that in circumstances where different philosophies within the tent dictate different actions, there is no guidance as to what route to take.

It is interesting to contrast this with the philosophic version of “pragmatism,” originally put forth by the polymath C. S. Peirce in the late 1800s [also credited with proposing the log-likelihood ratio as a measure of evidence (Hacking, 1965)], whose intellectual heirs included William James, Thomas Dewey, W. V. O. Quine and Richard Rorty. Pragmatism embraced three maxims, the most important of which was that the meaning of ideas was defined by their practical, observable effects. Ideas that made no material difference in the real world had no meaning.

Steven N. Goodman is Professor of Oncology, Epidemiology and Biostatistics, Sidney Kimmel Comprehensive Cancer Center, Department of Oncology, Division of Biostatistics/Bioinformatics, 550 N. Broadway, Suite 1103, Baltimore, Maryland 21205, USA (e-mail: sgoodman@jhmi.edu).

Kass alludes to a possible difference in real-world consequences just once, in his mention of the analysis of high-dimensional data sets. But he states his pragmatist philosophy is agnostic on how to approach these, and that the choice should be “according to their performance under theoretical conditions thought to capture relevant real-world variation in a particular applied setting.” It would have been extraordinarily useful to see such an example, and if indeed there could be a model-based resolution of what are often quite difficult conundrums.

In the domain with which I am most familiar, clinical trials, the traditional frequentist-Bayesian inferential dilemma arises most commonly in the interpretation of “early stopped” trials, that is to say, should the inference depend upon the stopping rule, and if so, how? This becomes particularly acute when the stopping is due to an unplanned analysis. This particular situation arose recently in the high-profile case of the diabetes drug Avandia. In 2007, a meta-analysis was published that raised concern about the cardiovascular risks of Avandia (Nissen and Wolski, 2007), leading to calls that the FDA should remove the drug from the market. The RECORD trial was being conducted in Europe to examine the efficacy and safety of Avandia, and its industry sponsor requested an unplanned analysis in response to the new data. This analysis (arguably) indicated no excess cardiac risk, and this interim result was then published, at the behest of the sponsor (Home et al., 2007; Nissen, 2010). Many doubted that an interim result that had demonstrated excess risk would have been published and discounted the result. How should this be sorted out? What are the dimensions of “real-world variation” here that we should include in the model, and on what grounds do we determine how to measure the evidence and how to act? This was a real decision, with real, big-time consequences. What guidance would “statistical pragmatism” give us in skirmishes like that?

Kass says he has been guided by “past and present sages,” but leaves the job of naming them to Barnett (1999). Kass is right that many have preceded him on this path, and it would have been quite illuminating