Statistical Science
2006, Vol. 21, No. 3, 310–312
DOI: 10.1214/088342306000000286
Main article DOI: 10.1214/088342306000000114

Comment: Causal Inference in the Medical Area

Edward L. Korn

It is an honor to be a discussant to the Morris Hansen Lecture, and a pleasure to be discussing Don Rubin's talk. Dr. Rubin has clarified over the years many of the deep issues relating to causal inference.

Let me start with a story. About 20 years ago when I was teaching at UCLA, I was eating breakfast one morning at my kitchen table, and my two-and-a-half-year-old daughter was in the next room, lying on her back and kicking the wall with her feet. I told her to stop, which she did for a few seconds, and then began again. I told her to stop again, and that I really meant it. The kicking stopped for a longer period this time, maybe 30 seconds, and then started up again. Just then the Whittier–Narrows earthquake hit, 5.9 on the Richter scale. Our 50-year-old house started shaking like crazy. As I was running into the next room to get my daughter, I ran into her running into the kitchen screaming "I'm sorry, Daddy, I'm sorry. I didn't mean to do it!" Which brings me to my first point: causal inference can be tricky.

Causal inference can be tricky not just for small children, but for epidemiologists and biostatisticians, too. As an example, consider hormone-replacement therapy for postmenopausal women. Dozens of observational studies (including case-control studies and cohort studies) had suggested a 40–50% reduction in coronary heart disease (Stampfer and Colditz, 1991). However, the recently reported results of the Women's Health Initiative trial demonstrated that the treatment had an elevated incidence of coronary heart disease (Manson et al., 2003). Now the statisticians who worked on these epidemiologic studies thought they were making a valid causal inference. In fact, many women took estrogen replacement therapy partly because they believed that it would offer cardiovascular benefits. However, as the large randomized trial demonstrated, this causal inference from the observational data was completely wrong.

Edward L. Korn is Mathematical Statistician, Biometric Research Branch, National Cancer Institute, Bethesda, Maryland 90824, USA (e-mail: korne@ctep.nci.nih.gov).

Because of the difficulty of doing randomized clinical trials of certain interventions, and the public health importance of whether these interventions work, I have put the ability to perform causal inference on epidemiologic data on the top of my personal list of "practical importance" of causal inference methods (Figure 1). The hormone-replacement therapy example is, of course, not the only example of medical studies where incorrect causal inferences were made. Let me just mention one other: There were many observational studies that suggested beta carotene would reduce lung cancer incidence; see International Agency for Research on Cancer (1998, pages 64–103) for a summary. However, randomized trials of beta carotene supplements showed that it actually increased the risk of lung cancer. In fact, the epidemiologic data were so strong that when the results of the first trial came out (Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group, 1994), an editorial suggested the possibility that trial results might be due to an "extreme play of chance" (Hennekens, Buring and Peto, 1994). However, after the results of the second trial also showed beta carotene was causing an increase in lung cancer (Omenn et al., 1996), it became clear that the epidemiologic studies had been wrong. To the extent that Dr. Rubin's work can lead to better causal inferences with epidemiologic data of these sorts, it would be of tremendous practical importance.

A cynical colleague of mine suggested that one should not give a discussion like this without mentioning some of your own work. So as an aside, I want to briefly mention a causal analysis I did a few years ago. We were interested in estimating the effect of an orthodontic treatment from observational data (Figure 2). These data were from the University of the Pacific orthodontic clinic, so which orthodontist saw which patients could be assumed to be random. What was definitely not random was which patients received the extraction treatment and which received the nonextraction treatment, because this decision depends on the patient characteristics. Because the treatment decision is not random, one cannot just compare the outcomes for patients who received extraction with those who