

Comment: Complex Causal Questions Require Careful Model Formulation: Discussion of Rubin on Experiments with “Censoring” Due to Death

Stephen E. Fienberg

1. INTRODUCTION

I am very pleased to be able to offer some reactions to yet another masterful paper by Rubin on the topic of causal inference. It is a special honor to do so because the paper was initially presented at Carnegie Mellon University as the 2005 Morris H. DeGroot Memorial Lecture and I was in the audience. “Morrie” DeGroot was my colleague, collaborator and close friend. He always raised questions about the naive use of randomization to answer causal questions from experiments. Thus, I believe that, had he been there to offer his own discussion at Rubin’s oral presentation, he might have opined on the two issues that I address below, although perhaps with more wit.

The present paper fits quite naturally with Rubin [12], where he discusses the problematic nature of intermediate outcomes and R. A. Fisher’s failure to recognize this problem. But this paper departs from that earlier one by avoiding the presentation of the key ideas using formal notation and equations. This makes for an interesting story but also for difficulties when one tries to follow the argument. The recipe for the resolution of most complex causal questions, we are told, is to frame them using potential outcomes and principal stratification. This is all well and good, but I still am not sure how to follow Rubin’s recipe, either for the stylized example he uses or in other settings in the future.

2. ALTERNATIVE REPRESENTATIONS FOR CAUSAL INFERENCE

I share Rubin’s enthusiasm for representing causal questions using the formal framework of counterfactuals

of which our philosophy colleagues are so fond. Rubin refers to these using the label “potential outcomes,” harking back to Neyman’s [7]. The reason I like this counterfactual representation is that it forces one to represent everything in terms of random variables, including randomization or any other allocation or missingness (censoring) mechanism; see [9–11]. Unfortunately, counterfactuals by their very nature lead us to condition on “unobservables” and thus they violate de Finetti’s [3] dictum that conditional probabilities only make sense when we condition on actual observables, not simply potential ones. This is at least in part why Dawid [1, 2] has attempted to present a framework for causal inference similar to Rubin’s but which avoids the counterfactual representation. Lauritzen [6] has a related graphical model approach to this which he links to Pearl’s [8] notion of “fixing” treatments or causes; see the similar ideas in [14].

My own preference is, as I suggested above, for representing every quantity under consideration using random variables, whether observed or unobserved, and then displaying these in graphical form using the standard methods for directed acyclic graphs. Thus, the act of randomization has a corresponding random variable and its introduction changes the graphical representation of the problem, often breaking the links between a treatment and an outcome variable; for example, see [4], as well as the more complete justification in [15]. This has the virtue of sidestepping Pearl’s “unnatural” embellishments to the notation and representation of causal effects.

This is a very long preamble to a plea: If the arguments in the current paper are truly to hold sway, then: (1) They must have formal representation so we can see precisely where the assumptions fit in, and (2) We need to formulate them using the different causal representations, not simply the potential outcome framework.

Stephen E. Fienberg is Maurice Falk University Professor of Statistics and Social Science, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA (e-mail: fienberg@stat.cmu.edu).