Rejoinder

Paul R. Rosenbaum

REPLY TO PAUL HOLLAND

It is always a great pleasure to receive comments from Paul Holland. The reader should understand that it takes a certain amount of effort if Paul and I are to find something to disagree about, but Paul has helpfully raised the issue of SUTVA, and so there is a small difference of language and emphasis worthy of discussion. I will not discuss Paul's intriguing CAI experiment in detail, primarily because I am not sure I absorbed enough about it from his brief description to offer useful comments. Of course, I cannot but agree with the conclusion he reaches at the end of this description in his paragraph 4: we should randomize whenever we can; we can far more often than we do; we should teach every beginning student about the importance of randomization in experiments; nonrandomized controls should be used only when the ethical or practical obstacles to experimentation are overwhelming. I suppose there has been little dispute among statisticians on these points for more than half a century.

Toward the end, Paul discusses what Don Rubin calls SUTVA, the stable unit-treatment value assumption. I would like to make a few general remarks about SUTVA, and then return to Paul's specific comments. This assumption concerns the notation that expresses treatment effects as comparisons of two potential responses for each subject; it says that this notation is adequate for the problem at hand. One might say it is the assumption, or perhaps the indefinite collection of assumptions, implicit in the notation. Don and Paul are certainly correct in emphasizing that virtually any notation carries assumptions buried within it, and they are especially correct in this case, for there are many ways in which the notation can be inadequate, and quite a few are of practical importance. Nonetheless, I do not love SUTVA as a generic label for all of these, for it seems to bear a distinct resemblance to an attic trunk; what does not fit is neatly folded and packed away. The more capacious the trunk, the more likely we are to have difficulty remembering precisely what is packed away. Periodically, we might open the lid and scan the top layer to illustrate what the trunk contains, but because it is so large, we are not inclined to take everything out, to sort the contents into piles: useful in season, useful if altered to fit, damaged beyond repair; still less are we inclined to begin the alterations, for the repair of each garment entails considerable effort. To press the metaphor, I would like to see the trunk opened, the contents sorted, alterations and repairs effected and to see what is beyond repair identified and clearly labeled. In other words, I would like to see SUTVA divided up into a series of more tangible assumptions with practical interpretations, so that violations could be quickly discerned and perhaps addressed. I know that Don and Paul do not want inadequacies of the notation to be forgotten, but I am concerned that the expansive concept SUTVA may tend to have this effect. Let me mention two parts to SUTVA, the second being relevant to Paul's comments.

One violation of SUTVA, "interference between units," is discussed by Cox (1958, Section 2.4) in the context of randomized experimentation. It is possible that the treatment assigned to one unit affects not just that unit, but other units as well. An example from observational studies is passive smoking: whether or not you get lung cancer from cigarette smoke depends not only on the treatment assigned to you—whether or not you smoke—but also on the treatments assigned to others—whether coworkers and family members smoke. In this case, the notation is not adequate: you do not have two potential responses depending solely on whether or not you smoke, but rather a multitude of potential responses depending on whether you smoke and whether those around you smoke. The usual solution to such interference in experiments is to amalgamate small units into larger units that do not interfere with one another; e.g., to replace students by classrooms in educational experiments. This solution works in some but not all observational studies. In studies of macroeconomics or of oligopolistic markets, it is usually impossible to identify two units that do not interfere with one another; in this case, the problem of interference between units must be tackled head on. I am not convinced that adequate methods for this task yet exist, although the interest economists have in simultaneous equation systems and cooperative games is clearly an attempt to grapple with this sort of problem.

A second violation of SUTVA is what Campbell and Stanley (1963) call "history" and I will call "intervening treatments." An intervening treatment is a second treatment, not the treatment of primary interest. The intervening treatment is applied after the primary treatment, and so the intervening treatment is not a covariate, but it is applied before the responses are observed, and so it may affect the responses. This is different from a factorial experiment in that the