Statistical Science
2009, Vol. 24, No. 1, 54–58
DOI: 10.1214/09-STS274A
Main article DOI: 10.1214/09-STS274

Comment: The Essential Role of Pair Matching

Jennifer Hill and Marc Scott

1. INTRODUCTION

We appreciate having the opportunity to comment on the well-motivated, highly informative and carefully constructed article by Imai, King and Nall (IKN). There has been a great deal of confusion over the years about the issue of pair-matching, often due to a conflation of the implications of design versus analysis choice. This article sheds light on the debate and offers a set of helpful alternative analysis choices.

Our discussion does not take issue with IKN's provocative assertion that one should pair-match in cluster randomized trials "whenever feasible." Instead we will explore the trade-offs between using the inferential framework advocated by IKN versus fitting fairly standard multilevel models (see, for instance, Gelman and Hill, 2007).

The IKN design-based treatment effect estimators have the advantage of being simple to calculate and having better statistical properties in general than the harmonic mean estimator that IKN view to be the most standard estimator in this setting. Variance estimators for SATE and CATE are not identified, but that is a function of not making the assumption of constant treatment effects, which we find realistic. IKN do provide upper bound variance estimators for these quantities of interest. Perhaps the biggest drawback to these methods is that they are not flexible if it is necessary or helpful to extend the framework to accommodate additional complications or information.

The strength of multilevel models in this estimation setting is the flexibility to build in complexity that could provide us with additional information, increase our precision, or sometimes even reduce bias (for instance, when correcting for "broken" randomization). As an example, while the IKN variance estimators accommodate varying treatment effects, the multilevel model provides a framework to actually examine these pair-to-pair differences. The model can also be extended to allow treatment effects to vary over covariate-defined subgroups which has the potential to substantially increase our understanding of effect transmission. Conditioning on pre-treatment covariates can also help to increase precision (and even reduce bias in situations where the randomization has been less pristine). Moreover, not only can multilevel models include covariates and random treatment effects quite readily, but the need for such terms can be evaluated statistically.

A further example is the ability of models to accommodate missing data at the individual level (rather than entire clusters being missing due to group-level noncompliance or attrition which IKN address). This can be naturally incorporated into a model-based framework as well; it's unclear how the IKN framework would handle this complication.

Of course, these advantages come at the cost of making some modeling assumptions. IKN go so far as to claim that these approaches "violate the very purpose of experimental work which goes to great lengths and expense to avoid these types of assumptions." However, the primary purpose of experimental work is to avoid the *untestable* assumption of ignorability (or strong ignorability) that is so difficult to avoid in observational work. While it is true that we do not need to build models post-randomization in order to estimate treatment effects, this can hardly be viewed as the goal of randomized experiments. In fact, randomization actually increases robustness to model-misspecification, creating a safer climate within which to build models than would otherwise exist. Moreover, the parametric assumptions we make with a multilevel model are testable, for instance, using graphical regression diagnostics.

It could be argued that multilevel models have the disadvantage of being more complicated to fit. However with the capabilities of current standard statistical software the level of technical expertise required to fit such models is well within the reach of most applied researchers today.

Jennifer Hill and Marc Scott are Associate Professors of Applied Statistics, Department of Humanities and Social Sciences, Steinhardt School of Culture, Education and Human Development, New York University, 246 Greene St., Room 300, New York, New York 10003, USA (e-mail: jennifer.hill@nyu.edu; marc.scott@nyu.edu).