NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Pair-Matching *vs* Stratification in Cluster-Randomized Trials

Background

There has been much scholarly debate regarding the respective advantages and disadvantages of complete randomization, stratification, and pair-matching in group- or cluster-randomized experiments. Matching can increase power by reducing study population heterogeneity and can guarantee balance on selected confounders by matching on them. However, matching decreases the degrees of freedom (df) and may lead to a loss of statistical power when matching is not effective (i.e., when the outcomes and the matching variables are poorly correlated).

Potential limitations of pair-matching

Diehr et al. [1] showed that matching may lead to some loss of power when the number of pairs is between three and nine, and recommended breaking the matches in statistical analyses to gain power. Donner et al. [2,3] cautioned that even though unmatched analysis for a pair-matched design is unbiased for the estimation of the intervention effect, it *is* biased for the estimation of the effects for other individual-level risk factors. They also pointed out that the pair-matching design makes the applications of generalized estimation equations or mixed-effects models no longer routine and further complicates statistical analyses. Moreover, with a pair-matching design, if one cluster is lost to follow-up, then the matched cluster is no longer useful and must be discarded. Due to these limitations, Donner et al. suggested that pair-matching designs should be used with caution.

Potential benefits of pair-matching

In direct contrast, Imai et al. [4] strongly advocated for the use of pair-matching in cluster-randomized experiments, suggesting that:

> *"...from the perspective of bias, efficiency, power, robustness or research costs, and in large or small samples, pairing should be used in cluster-randomized experiments whenever feasible; failing to do so is equivalent to discarding a considerable fraction of one's data."*

They proposed a simple design-based estimator and model-based approach which, according to the authors, offer unbiased, efficient, and robust inference from pair-matched cluster-randomized experiments.

Recommendations

Imbens [5] conducted a study for the International Initiative for Impact Evaluation (3ie) that examined the arguments from both sides. He concluded that in terms of precision of estimation, pair-matching is superior to stratification (more than two units/clusters in each stratum), which is then superior to complete randomization. At the same time, he acknowledged that the "statistical limitations" associated with pair-matching as pointed out by Donner [3] and Klar [6] are real. Thus, he concluded that

> *"The overall recommendation, irrespective of the sample size and correlation between covariates and outcomes, is to use stratified randomization with relatively many and relatively small strata (to capture as much as possible of the precision gain from stratification), rather than either complete randomization or paired randomization (which would capture all of the potential gain in precision, but at the price of the analytical complications)."*

Imbens also suggests that "It is particularly helpful to stratify or pair based on cluster size if (i) cluster size varies substantially, (ii) interest is in average effects for the entire population and (iii) cluster size is potentially correlated with the average effect of the intervention by cluster." [5]

Next steps

Imbens' recommendation introduces a number of possible ambiguities: What exactly does "relatively many and relatively small strata" mean? How many is "many" for a cluster-randomized experiment with 20 clusters and 100 subjects per cluster? Where stratum-specific effects are of interest, it may be necessary to have a larger number of clusters per stratum; where the stratification factors are essentially nuisance variables, it may be possible to create strata with four clusters each, thereby maximizing the number of strata and the gain in efficiency. At the same time, there is likely a tradeoff between having more strata to gain efficiency and having fewer strata to achieve a valid estimate of the within-stratum variation.

We propose that a comprehensive simulation study is needed to confirm the recommendations offered by Imbens and also to provide specific guidelines on design selection. A number of important design factors should be considered for the simulation study, including (but not limited to) number of clusters, cluster sizes, number of covariates that need to be matched or stratified on, and their correlations with the outcome. Such a comprehensive simulation study will help quantify the potential gain in precision with a pair-matched design or a stratified randomization design with varying numbers of strata to aid in design selection.

Resources

1. Diehr P, Martin DC, Koepsell T, Cheadle A. Breaking the matches in a paired t-test for community interventions when the number of pairs is small. Stat Med 1995;14(13):1491–504. PMID: 7481187.
2. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. Am J Public Health 2004;94(3):416–22. PMID: 14998805.
3. Donner A, Taljaard M, Klar N. The merits of breaking the matches: a cautionary tale. Stat Med 2007;26(9):2036–51. PMID: 16927437. doi: 10.1002/sim.2662.

4. Imai K, King G, Nall C. The essential role of pair matching in cluster-randomized experiments, with application to the Mexican Universal Health Insurance Evaluation. Stati Sci 2009;24(1):29–53. Available at: http://projecteuclid.org/DPubS?verb=Display&version=1.0&service=UI&handle=euclid.ss/1255009008&page=record. Accessed May 14, 2014.
5. Imbens GW. Experimental design for unit and cluster randomized trials. Prepared for the International Initiative for Impact Evaluation. 2011. Available at: http://cyrussamii.com/wp-content/uploads/2011/06/Imbens_June_8_paper.pdf. Accessed May 14, 2014.
6. Klar N, Donner A. The merits of matching in community intervention trials: a cautionary tale. Stat Med 1997;16(15):1753–64. PMID: 9265698. doi: 10.1002/(SICI)1097-0258(19970815)