Paper 2: CRT-Power

Abstract Body

Background / Context:

In a simple randomized trial, once the effect size and alpha are set, power is determined solely by the number of subjects. This is the only factor that the researcher can manipulate, and also the only factor that the researcher needs to manipulate. By contrast, in a cluster-randomized trial the researcher needs to be concerned with the number of units at each level – for example, the number of students, classes, schools, and districts – and each needs to be considered separately, since each has a specific impact on statistical power.

Purpose / Objective / Research Question / Focus of Study:

We will present a computer program, CRT-Power that researchers can use to design cluster-randomized (hierarchical) and randomized-block (matched, or multi-site) trials.

The program allows users to work with a relatively wide array of designs. It works with two-level, three-level, and four-level designs. Additionally, it allows randomization at any level.

The program will work with several indices of effect size. For continuous data it works with the standardized mean difference (d), which can be standardized by the standard deviation within-cells or the standard deviation across-cells. For binary data it works with the odds ratio or the risk difference.

The program is designed for clarity and transparency. For example, a key factor in computing power is the ICC, a number that reflects the degree of “clustering”, or the extent to which (for example) student scores within a school tend to form a cluster, distinct from scores at other schools. This value is important to the calculations, but may not be well understood. To help ensure that the value entered for the ICC is plausible, the program “translates” the ICC into more intuitive terms. If the researcher provides the ICC the program shows how this translates into a span of means (for example, an ICC of .20 for schools on the SAT might imply that 95% of the school means vary over a span of 400 to 600). Conversely, if the researcher enters a span covering 95% of the expected school means, the program will translate this into an ICC.

The program takes a similar approach to variation in effect sizes for multi-site designs. The researcher can say that 95% of the effect sizes (d) will vary over a particular span (say, 0.20 to 0.50) and the program will compute the variance of these effects. Conversely, the user can provide the variance of the effects and the program will display the corresponding span (95% interval).

In a cluster-randomized or multi-site design there are many combinations of samples sizes that will yield the same power (for example, more schools with fewer students per school, or fewer schools with more students per schools). Some of these designs can be much more cost-effective than others. The program can help the researcher to identify the most cost-effective design, taking into account the variance, covariates, and costs at each level.
In the process of identifying the most cost-effective design the program can also take account of practical constraints. For example, the most cost-effective design might call for one class in each school, but for pragmatic reasons the researcher would prefer to include four classes in each school. Conversely, the most cost-effective design might call for ten classes in each school but for pragmatic reasons the researcher needs to enroll only five per school. The researcher is able to set constraints on each level. The program will then find the most cost-effective design by modifying the number of units at other levels while honoring these constraints.

Power analysis should be an exercise in finding the best balance among competing options. For example, while it might be desirable to increase the number of schools and the number of students per school, our resources might force us to choose one or the other. To make an informed decision we need to understand how these numbers affect power and how they affect costs. To facilitate this process the program creates graphs that show power and cost as function of any two variables at a time.

Another goal of this program is to educate the researcher about the issues in power analysis for multi-level trials. To that end the program produces two detailed reports of the analysis. One is a synopsis, but the second is a lengthy report that explains all elements of the analysis in detail. The researcher can use this to learn more about the process, and also to help ensure that all values were entered correctly.

For continuous normally distributed outcomes, the power calculations are based on exact theory. For the binary outcomes, we must rely on large sample theory, but we are conducting simulations to identify the range of parameters where the computed power does (and does not) match the empirical power. In particular, when the prevalence is very low (or high) we would expect the studies to have some cells with zero counts, which would violate some assumptions of the tests and affect the validity of the algorithms.

**Significance / Novelty of study:**

The algorithms developed for and implemented in this program provide an important advance to the field in the following areas

Designs. This program extends the available methods to four levels, with option to randomize subject at any level. For the first time, it is possible to allow for covariates at all levels.

Conditional optimal design. This program allows the researcher to set constraints on the sample size at one or more levels when finding the most cost-effective design. The algorithms for finding the most cost-effective design have also been extended to four levels, to work with blocking at one or more levels, and to take account of covariates at all levels.

The simulations provide empirical evidence about the limits of the algorithms. Additionally, these empirical limits are incorporated into the program, and issue a warning if the combination of factors (mean prevalence, variation in prevalence, effect size, variation in effect size, and sample size) affect the validity of the power estimates.
Translations. While the mathematical meaning of the various parameters may be clear, there is a lot of confusion in the research community about how to enter these parameters into computer programs. The problem is exacerbated by the fact that different programs expect the user to enter different variants of a parameter. In this program we include various features to help the researcher understand what a parameter means, and that the value entered is plausible.

Statistical, Measurement, or Econometric Model:

For computing power for continuous data and binary data (using the log-odds model) we used the same approach as the OD (Optimal Design) program, but extended this approach to work with additional sampling levels and to allow for randomization at all levels of the design. To compute power for binary data using risk differences we developed new algorithms. New algorithms were developed to identify the most cost-effective design for two, three, and four level designs, while allowing for constraints at each level.

The algorithms used in this program for binary outcomes (as well as those used in OD) are based on asymptotic theory. We conducted a series of simulations to provide empirical information about the validity of the algorithms, and also help us to identify the useful limits of these algorithms. The algorithms employed for binary outcomes do not perform well outside a range of parameters (for example, they may not work when the prevalence is below .10 or above .90) and we are using the simulations to identify these parameters.

Usefulness / Applicability of Method:

Cluster-randomized (hierarchical) and multi-site (matched or randomized-block) studies play a critical role in the field of educational research, and therefore researchers need tools to perform a power analysis for these studies. This program is an important contribution to the field in that (a) it will work with two-level, three-level, and four-level designs (b) works with continuous and binary outcomes (c) allows randomization at any level (d) allows for covariates at all levels (e) includes mechanisms to help ensure that parameters such as the ICC are entered correctly (f) will help find the most cost-effective design while taking account of practical constraints (g) creates graphs that the researcher can use to find the best balance among competing options, and (h) creates detailed reports.

Research Design:

We ran simulations to identify the conditions under which the algorithms for computing power are (and are not) valid. In these simulations we systematically varied factors including the mean prevalence, variation in prevalence, mean effect size, variation in effect size, and sample size.

Data Collection and Analysis:

Clusters, sub-clusters, sub-sub-clusters, and subject outcomes are being generated for treated and control groups based on study parameters, then tested for statistical significance. The proportion
meeting the criterion (over 10,000 simulations) gives the empirical power. This value is then compared with the value reported by CRT-Power and (where appropriate) by Optimal Design.

Findings / Results:

The simulations are ongoing.

Conclusions:

The computer program affords the opportunity to compute power for a wide array of study designs, to ensure that the parameters are plausible, and to display results in a manner that allows the user to balance competing options.

The simulations are ongoing. We have confirmed that the algorithms employed by CRT-Power and OD are valid within a wide range of parameters, but not outside that range. The “Range” is actually a set of parameters which interact, and we will show how these can be used to identify study designs that exceed these limits.