

Key Vocabulary:

1. **Power:** the likelihood that, when the program has an effect, one will be able to distinguish the effect from zero given the sample size.
2. **Significance:** the likelihood that the measured effect did not occur by chance. Statistical tests are performed to determine whether one group (e.g. the experimental group) is different from another group (e.g. comparison group) on the measurable outcome variables used in the evaluation.
3. **Standard Deviation:** a standardized measure of the variation of a sample population from its mean on a given characteristic/outcome. Mathematically, the square root of the variance.
4. **Standardized Effect Size:** a standardized measure of the [expected] magnitude of the effect of a program.
5. **Cluster:** the level of observation at which a sample size is measured. Generally, observations which are highly correlated with each other should be clustered and the sample size should be measured at this clustered level.
6. **Intra-cluster Correlation Coefficient:** a measure of the correlation between observations within a cluster; i.e. the level of correlation in drinking water source for individuals in a household.

Sample size calculations

The Extra Teacher Program (ETP) case study discussed the concept of cluster randomized trials. The Balsakhi example used in the prior lecture introduced the concept of power calculations. In the latter, we were interested in measuring the effect of a treatment (balsakhis in classrooms) on outcomes measured at the individual level—child test scores. However, the randomization of balsakhis was done at the classroom level. It could be that our outcome of interest is correlated for students in the same classroom, for reasons that have nothing to do with the balsakhi. For example, all the students in a classroom will be affected by their original teacher, by whether their classroom is unusually dark, or if they have a chalkboard; these factors mean that when one student in the class does particularly well for this reason, all the students in that classroom probably also do better—which might have nothing to do with a balsakhi.

Therefore, if we sample 100 kids from 10 randomly selected schools, that sample is less representative of the population of schools in the city than if we selected 100 random kids from the whole population of schools, and therefore absorbs less variance. In effect, we have a smaller sample size than we think. This will lead to more noise in our sample, and hence larger standard error than in the usual case of independent sampling. When planning both the sample size and the best way to sample classrooms, we need to take this into account.

This exercise will help you understand how to do that. Should you sample every student in just a few schools? Should you sample a few students from many schools? How do you decide?

We will work through these questions by determining the sample size that allows us to detect a specific effect with at least 80% power. Remember power is the likelihood that when the treatment has an effect you will be able to distinguish it from zero in your sample.

In this example, “clusters” refer to “clusters of children”—in other words, “classrooms” or “schools”. This exercise shows you how the power of your sample changes with the number of clusters, the size of

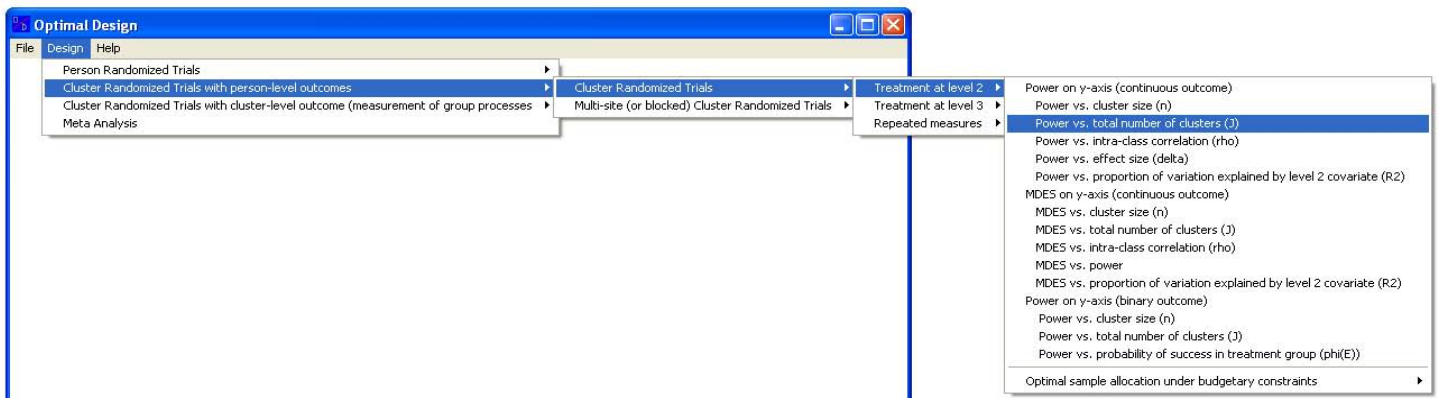
the clusters, the size of the treatment effect and the Intraclass Correlation Coefficient. We will use a software program developed by Steve Raudebush with funding from the William T. Grant Foundation. You can find additional resources on clustered designs on their web site.

Section 1: Using the OD Software

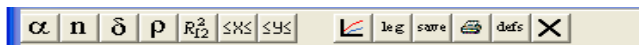
First download the OD software from the website (a software manual is also available):

http://sitemaker.umich.edu/group-based/optimal_design_software

When you open it, you will see a screen which looks like the one below. Select the menu option “Design” to see the primary menu. Select the option “Cluster Randomized Trials with person-level outcomes,” “Cluster Randomized Trials,” and then “Treatment at level 2.” You’ll see several options to generate graphs; choose “Power vs. Total number of clusters (J).”



A new window will appear:



Select α (alpha). You’ll see it is already set to 0.050 for a 95% significance level.

First let’s assume we want to test only 40 students per school. How many schools do you need to go to in order to have a statistically significant answer?

Click on n , which represents the number of students per school. Since we are testing only 40 students per school, so fill in $n(1)$ with 40 and click OK.

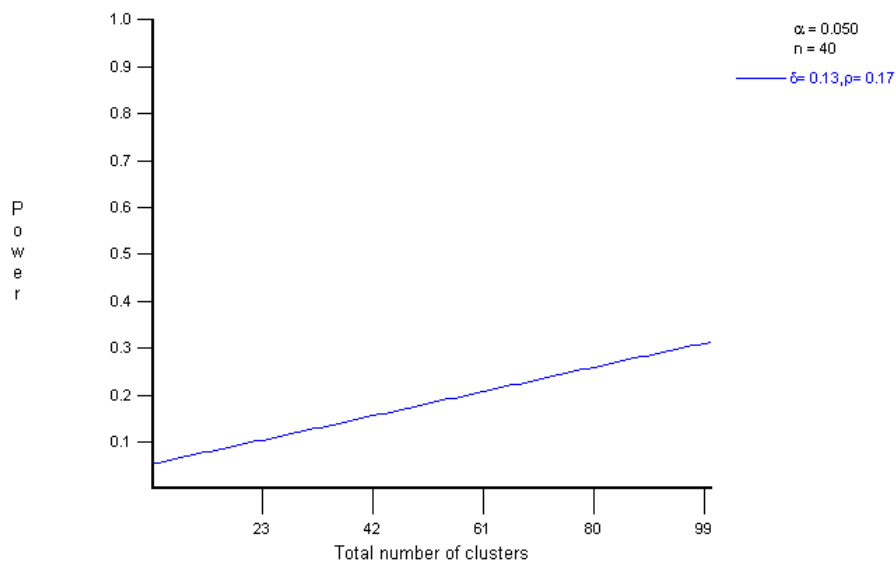
Now we have to determine δ (delta), the standard effect size (the effect size divided by the standard deviation of the variable of interest). Assume we are interested in detecting whether there is an increase of 10% in test scores. (Or more accurately, are uninterested in a detect less than 10%) Our baseline survey indicated that the average test score is 26, with a standard deviation of 20. We want to detect an effect size of 10% of 26, which is 2.6. We divide 2.6 by the standard deviation to get δ equal to 2.6/20, or 0.13.


Select δ from the menu. In the dialogue box that appears there is a prefilled value of 0.200 for delta(1). Change the value to 0.13, and change the value of delta (2) to empty. Select OK.

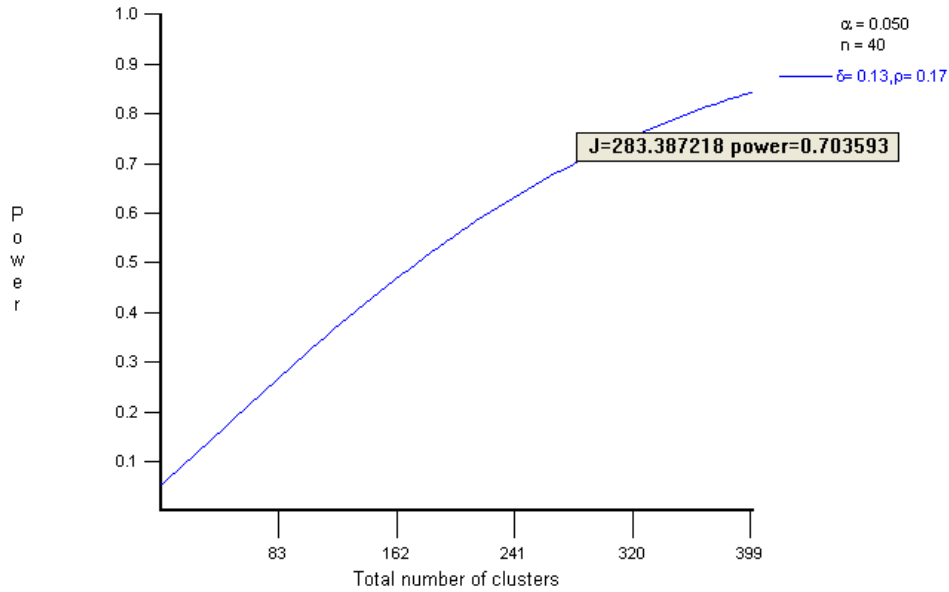
Finally we need to choose ρ (rho), which is the intra-cluster correlation. ρ tells us how strongly the outcomes are correlated for units within the same cluster. If students from the same school were clones (no variation) and all scored the same on the test, then ρ would equal 1. If, on the other hand, students from the same schools are in fact independent—and there was no differences between schools, then ρ will equal 0.

You have determined in your pilot study that ρ is 0.17. Fill in rho(1) to 0.17, and set rho (2) to be empty.

You should see a graph similar to the one below.



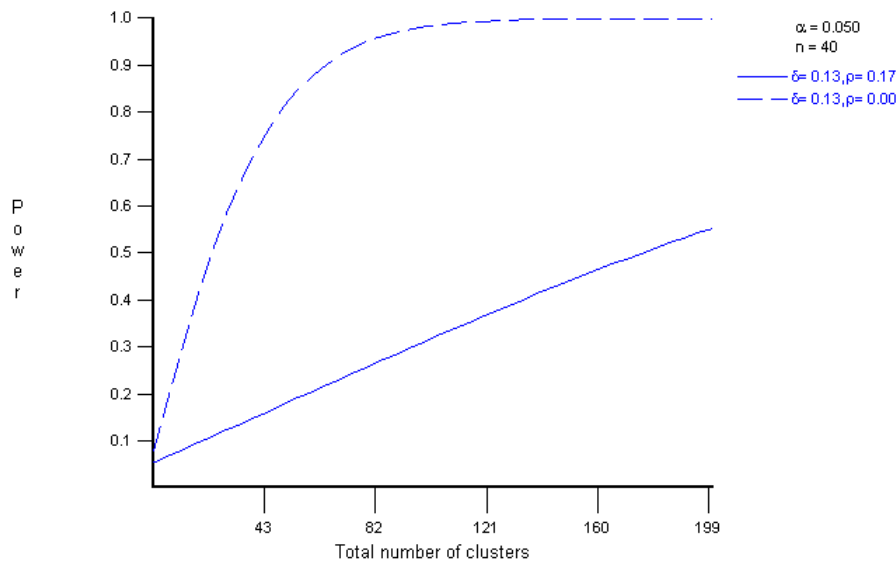
You'll notice that your x axis isn't long enough to allow you to see what number of clusters would give you 80% power. Click on the  button to set your x axis maximum to 400. Then, you can click on the graph with your mouse to see the exact power and number of clusters for a particular point.




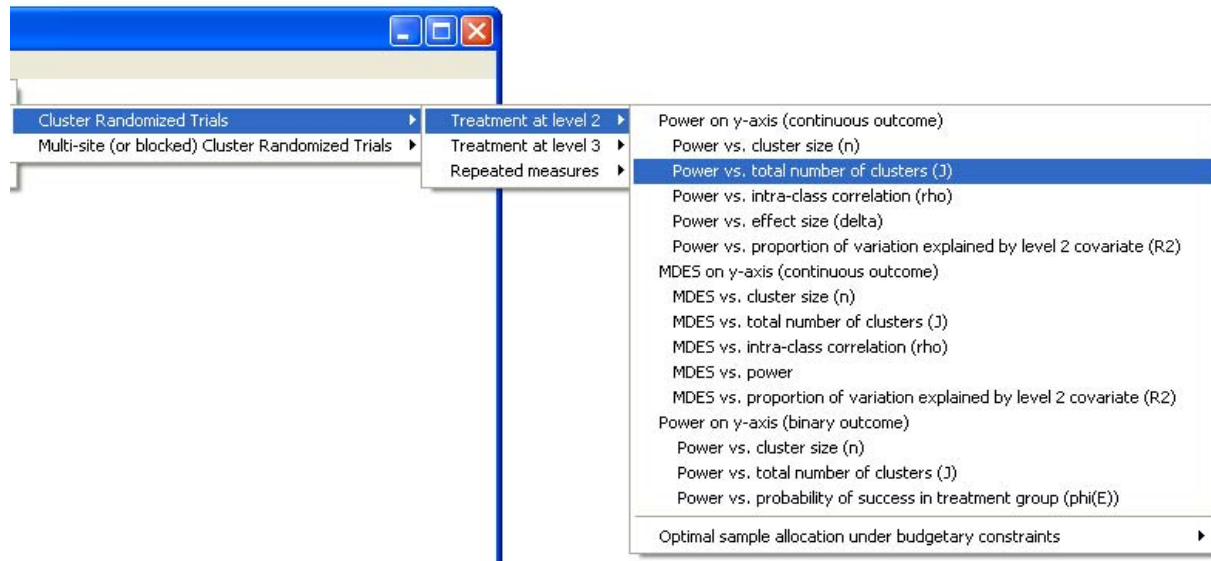
Now you have seen how many clusters you need for 80% power, sampling 40 students per school. Suppose instead that you only have the ability to go to 124 schools (this is the actual number that was sampled in the Balsakhi program).

Finally, let's see how the Intraclass Correlation Coefficient (ρ) changes power of a given sample. Leave $\rho(1)$ to be 0.17 but for comparison change $\rho(2)$ to 0.0.

You should see a graph like the one below. The solid blue curve is the one with the parameters you've set - based on your pretesting estimates of the effect of reservations for women on drinking water. The blue dashed curve is there for comparison - to see how much power you would get from your sample if ρ were zero. Look carefully at the graph.



To take a look at some of the other menu options, close the graph by clicking on the  in the top right hand corner of the inner window. Select the Cluster Randomized Trial menu again.



MIT OpenCourseWare
<http://ocw.mit.edu>

Resource: Abdul Latif Jameel Poverty Action Lab Executive Training: Evaluating Social Programs
Dr. Rachel Glennerster, Prof. Abhijit Banerjee, Prof. Esther Duflo

The following may not correspond to a particular course on MIT OpenCourseWare, but has been provided by the author as an individual learning resource.

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.