

## Online Resource 6. Estimating the required sample size

Power calculations help program managers and evaluators estimate the required sample size that is large enough to provide sufficient statistical power to detect the impact of the program. The following four steps outline the information needed to calculate the required sample size (World Bank 2007):

### 1. Desired level of confidence:

We need to determine the probability of mistakenly finding an impact when the true effect is zero. This value is commonly referred to as “alpha”. Naturally, we want this probability to be small. Common thresholds for alpha are 1%, 5% or 10%, which in turn means a confidence of 99%, 95%, or 90%. The lower the alpha, that is, the more confident we want to be to correctly identify the impact, the higher the needed sample size.

### 2. Expected effect size of the program:

We need to determine the average magnitude we expect our outcome(s) of interest to change, e.g. an increase in income by 20%; a reduction in job search time by 50%; etc. In practice, we need to specify the minimum effect we would like to see that would justify the existence of the program. Consultations with stakeholders are useful to get consensus on this minimum desired effect size. Choosing smaller effect sizes requires larger samples to detect them.

### 3. Expected variation in outcomes:

Another determinant of the required sample size is the expected variation of the outcome(s) of interest across the study population. For instance, how big may be the difference in incomes between those who benefit most (e.g. doubling their earnings) and those who benefit least (e.g. zero earnings). To estimate this, we can use existing data, e.g., national household surveys, and previously available evaluations and studies on the same topic. If this type of information is not available, a rule of thumb is to use one quarter of the total range of outcomes (e.g. between \$0 and \$100) as “standard deviation” (i.e. \$25). Standard deviation refers to how much variation there is from the average or mean value (a low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values).

### 4. Desired level of power:

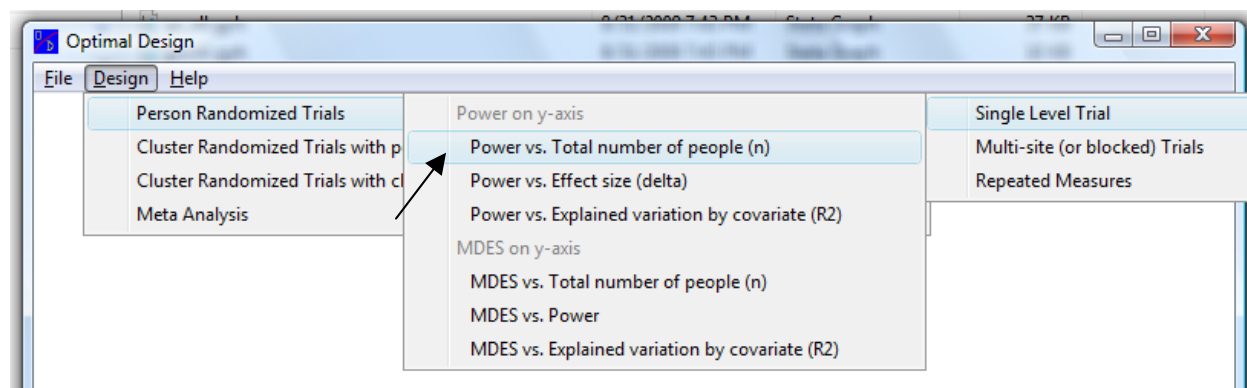
Power refers to the probability of detecting an impact when there is one. We want this probability to be high, of at least 80% or higher. For example, a power of 90% means that we will find an impact in 90% of the cases where one has occurred. The higher the power we choose (i.e. a higher chance of detecting an existing effect), the larger the sample we need.

Several software packages can be used to enter the above information and compute the estimated sample size, including the Optimal Design program which can be downloaded for free ([http://sitemaker.umich.edu/group-based/optimal\\_design\\_software](http://sitemaker.umich.edu/group-based/optimal_design_software)).

Instructions to compute the estimated sample size using Optimal Design follow below:

**Step 1:**

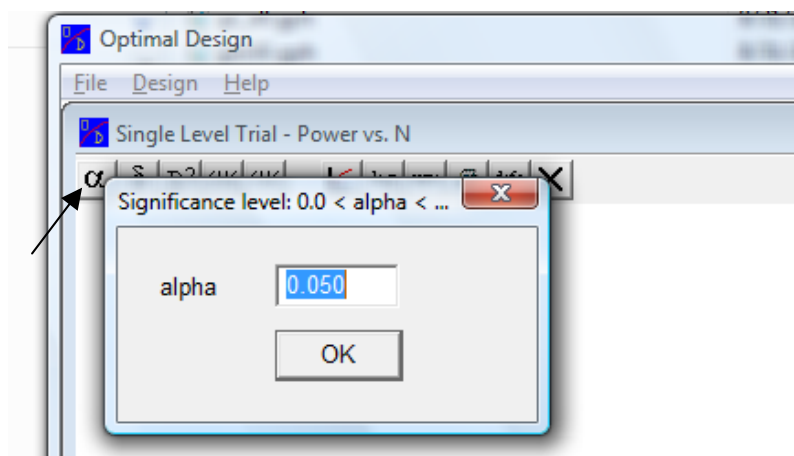
Click on: Design > Person Randomized Trials > Single Level Trial > Power vs. Total number of people (n)



This will yield a new window which will allow us to visualize the required sample size for a given level of the features described above.

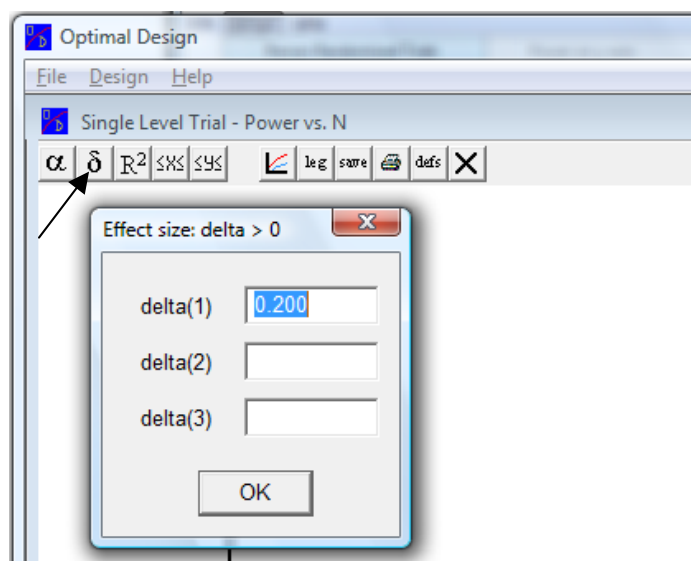
**Step 2:**

Select the desired level of *alpha* (e.g. 5%) which would represent a confidence level of 95% to not mistakenly finding an impact when the true effect is zero.



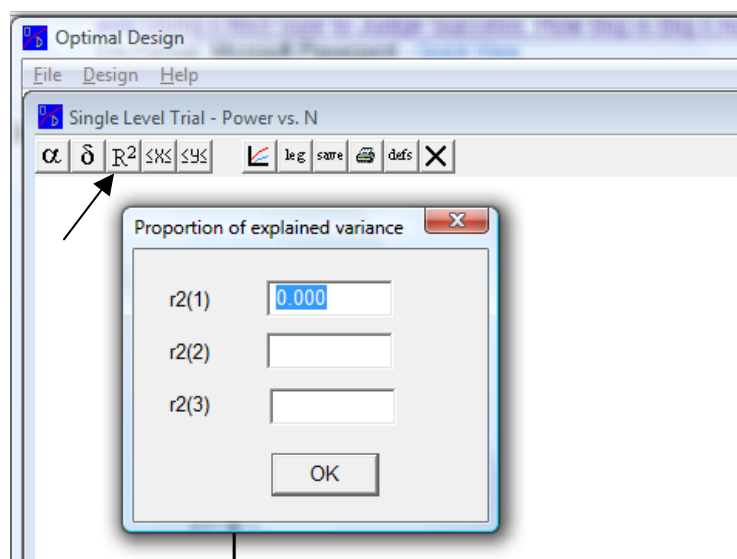
**Step 3:**

Select the expected effect size ("delta"); that is, the average percentage change in outcome(s) that we expect to see.



**Step 4:**

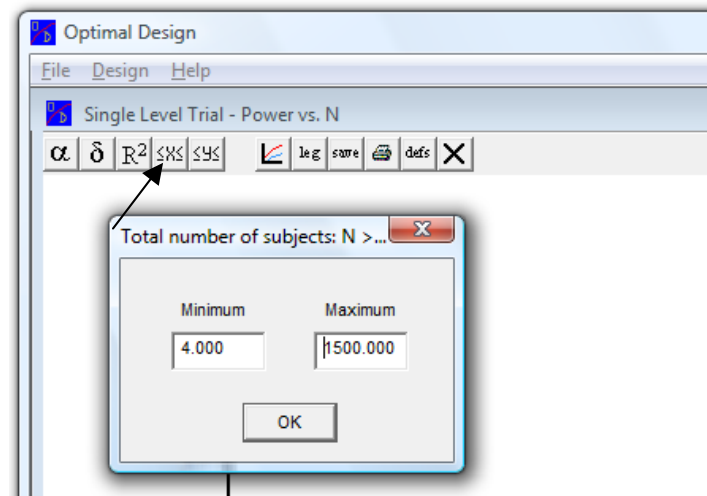
Optimal Design does not allow directly entering the variation of effect size in terms of standard deviation. Instead it captures the so called “proportion of explained variance”, or  $R^2$ , which is a related concept. It is advised to keep this value at zero unless when working with a professional evaluator who could advise a more appropriate value for the specific program.



**Step 5:**

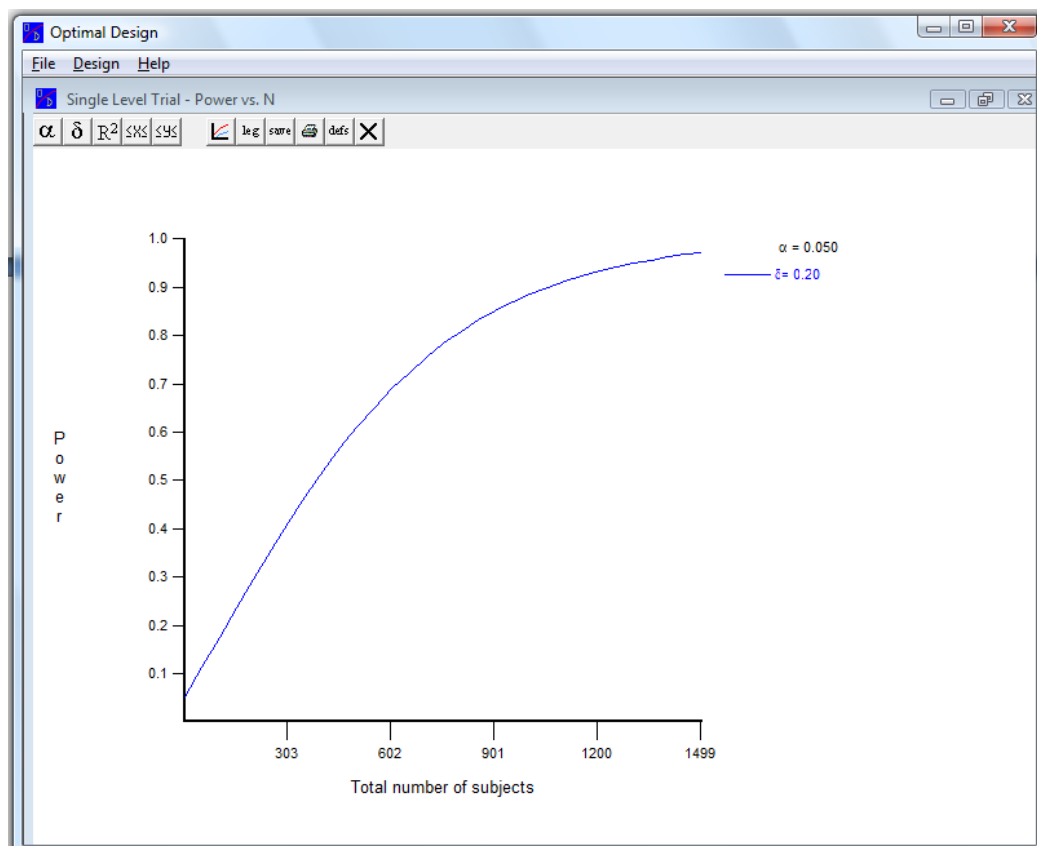
Adjust the x-axis to a large enough number (e.g. 1500) to capture the required sample size.

This version: January 31, 2012.



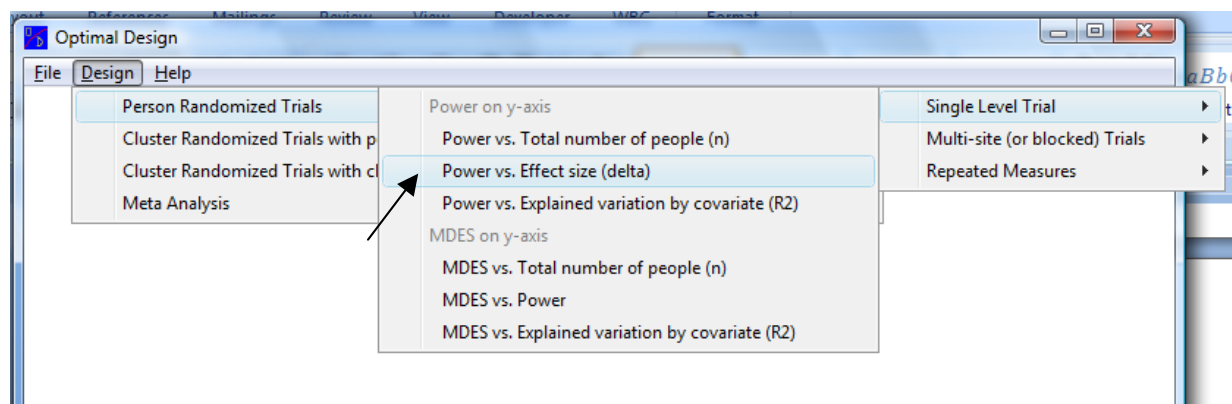
**Result:**

Optimal Design yields the minimum required sample size (on the x-axis) for different levels of power (y-axis). We are interested in the sample size for levels of power 80% (0.8) or above. In the given example, we would need a sample size of at least 793 individuals at a level of power of 80%. At a power of 90%, we would need a sample of approximately 1055 individuals.



**Note 1:** Estimates of the required sample size represent the minimum number of people that need to be included given the parameters chosen. In practice, there is a realistic possibility that during the implementation of the evaluation some observations will drop, e.g., due to the difficulty to find all members of the treatment and comparison group, non responses, or problems of data processing. Therefore, we need to estimate an expected rate of lost observations and increase the sample according to this rate (a 10-20% extra margin is common).

**Note 2:** If the size of the program and/or sample is already known, power calculations can identify the minimum effect size that would be needed to identify an impact if one has occurred.



#### Sources:

University of Michigan. 2011. *Building Capacity to Evaluate Group-Level Interventions*. Website. [http://sitemaker.umich.edu/group-based/optimal\\_design\\_software](http://sitemaker.umich.edu/group-based/optimal_design_software). Accessed: January 12, 2012.

World Bank. 2007. *Data for Impact Evaluation*. Doing Impact Evaluation No. 6. Washington, DC: The World Bank. [http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing\\_ie\\_series\\_06.pdf](http://siteresources.worldbank.org/INTISPMA/Resources/383704-1146752240884/Doing_ie_series_06.pdf)