A Worked Example for an *A Priori* Stratified Analysis of a Net Difference in a Cross-Sectional Design

The calculator will use the parameter estimates you have provided to calculate the required sample size of groups per condition x stratum cell for the intervention effect (Δ) and the level of power that you specified.

In addition, the calculator will use the other parameter estimates to calculate the detectable difference (Δ) that is available with the level of power that you specified as a function of the number of groups (g) in each condition x stratum cell and the number of members (m) in each group, using the formula below. Note: this formula is for a cross-sectional design, so that $\left(1-r_{yy_m}\right)$ has been removed from the left-hand side of the numerator.

$$\Delta=\sqrt{2\times 2\times 2\left(\frac{\left(\sigma_y^2(1-ICC)\left(1-R^2_{y\cdot x_m}\right)\left(1-r_{ys_m}\right)+m\sigma_y^2(ICC)\left(1-R^2_{y\cdot x_g}\right)\left(1-r_{ys_g}\right)\left(1-r_{yy_g}\right)\right)}{mg}\right)\left(t_{\alpha/2}+t_\beta\right)^2}$$

Here is the variance of the outcome variable ignoring any expected ICC, ICC is the intraclass correlation among the members in the same group, $R^2_{y\cdot x_m}$ is the proportion of variance explained by the member-level covariates, r_{ys_m} is the correlation between the outcome variable and the stratification variable at the level of the member, m is the number of members in each group, $R^2_{y\cdot x_g}$ is the proportion of variance explained by the group-level covariates, r_{ys_g} is the correlation between the outcome variable and the stratification variable at σ_y^2 the level of the group, r_{yy_g} is the overtime correlation for the outcome variable at the level of the group, g is the number of groups or clusters in the each condition x stratum cell, $t_{\alpha/2}$ is the t-value selected based on the two-tailed alpha level and available degrees of freedom, and t_β is the t-value selected based on the desired power and available degrees of freedom.

As an example, consider the following set of parameter estimates:

Let m = 100 and g =48 with x_m=4 df used for member-level covariates and x_g=1 df used for group-level covariates. Then

$$df=(t-1)cs(g-1)-df_g=(2-1)2\times 2(48-1)-1=187$$

For a two-tailed type 1 error rate of 5% and power of 80%

$$t_{\alpha/2}=1.9727 \text{ and } t_\beta=0.8435$$

Let σ_y^2= 1.0 and ICC=0.05.

Let $R^2_{y\cdot x_m}=0.20$ and $R^2_{y\cdot x_g}=0.00$.

Let $r_{ys_m} = 0.10$ and $r_{ys_g} = 0.00$.

Let $r_{yy_g} = 0.20$.

Then

$$\Delta = \sqrt{2 \times 2 \times 2\left(\frac{\left(1(1-0.05)(1-0.2)(1-0.1)+100(1)(0.05)(1-0)(1-0)(1-0.2)\right)}{100 \times 48}\right)(1.9727+0.8435)^2}$$
$$= 0.2488$$

The detectable difference given these parameter estimates is 0.2488 standard deviation units.