Sample size calculation

Simplest approach is to use the standard formula for detecting a difference in two proportions :

$$n = \frac{r + 1}{r} \times \frac{(\bar{p})(1 - \bar{p})(Z_\beta + Z_{\alpha/2})^2}{(p_1 - p_2)^2}$$

- n: Sample size in the *case* group
- r: Ratio of controls to cases
- $\bar{p}$: Mean proportion exposed in cases and controls
- Z_β: Desired Power (0.84 for 80%)
- $Z_{\alpha/2}$: Significance (use 1.96 for 5%)
- p_1: Proportion of cases exposed
- p_2: Proportion of controls exposed

Sample size calculation

Some simplification if we accept typical values for power (80%) and significance (5%) :

$$n = \frac{r+1}{r} \times \frac{(\bar{p})(1-\bar{p}) \times 7.84}{(p_1 - p_2)^2}$$

Since SAM is a rare condition we will find it easier to find controls than to find cases. We should, therefore, collect more controls than cases. This is little to gain from collecting more than about five controls per case. We can simplify further :

$$n = 1.2 \times \frac{(\bar{p})(1-\bar{p}) \times 7.84}{(p_1 - p_2)^2}$$

Sample size calculation

An example :

How many cases and controls are required assuming :

- 80% power and 5% significance
- We want to detect an odds ratio (OR) of 3.0 (i.e. exposure triples the odds of being a case) or greater.
- We will collect five controls for each case.
- The proportion exposed to the risk factor in the control group (p_2) is 25%.

We will used this as a worked example on the next page.

Sample size calculation

Calculate the proportion exposed in the case group :

$$p_{\text{exposed in } \mathbf{case} \text{ group}} = \frac{OR \times p_{\text{exposed in } \mathbf{control} \text{ group}}}{p_{\text{exposed in } \mathbf{control} \text{ group}} \times (OR - 1) + 1} = \frac{3(0.25)}{(0.25)(3-1)+1} = \frac{0.75}{1.5} = 0.5$$

Calculate the mean exposure proportion :

$$\textit{mean proportion exposed} = \frac{p_{\text{exposed in } \mathbf{case} \text{ group}} + p_{\text{exposed in } \mathbf{control} \text{ group}}}{2} = \frac{(0.5+0.25)}{2} = 0.375$$

Calculate the sample size :

$$n = 1.2 \times \frac{(0.375)(1-0.375) \times 7.84}{(0.5-0.25)^2} = 35$$

This is the number of cases needed. The number of controls needed is :

$$n_{controls} = 5 \times 35 = 175$$

Sample size calculation – Continuous case

The calculations shown above are for binary exposure variables. If the exposure variable is a continuous measure then the following formula should be used :

$$n = \frac{r+1}{r} \times \frac{\sigma^2 \left(Z_\beta + Z_{\alpha/2} \right)^2}{\left(\bar{x}_1 - \bar{x}_2 \right)^2}$$

n: Sample size in the *case* group

σ^2: Pooled (average) variance

Z_β: Desired Power (0.84 for 80%)

$Z_{\alpha/2}$: Significance (use 1.96 for 5%)

r: Ratio of controls to cases

$\bar{x}_1$: Mean value in cases

$\bar{x}_2$: Mean value in controls

This formulae can be simplified by assuming fixed levels of power and significance and a fixed number of controls per case (as was done to the formula for binary exposure variables).