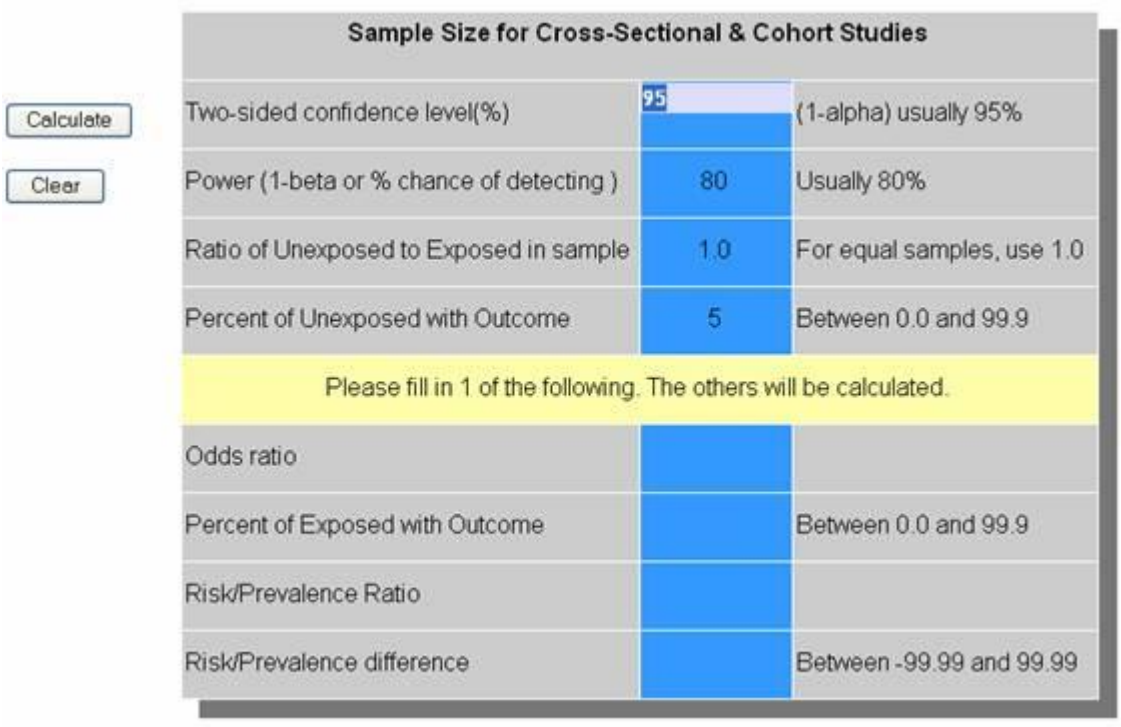
Feb 16 2007

Documentation for

Sample Size for a Cross-Sectional, Cohort, or Clinical Trial Studies

Kevin M. Sullivan, PhD, MPH, MHA: cdckms@sph.emory.edu
Minn M. Soe, MD, MPH, MCTM: msoe@sph.emory.edu

This module calculates sample size for a cross-sectional study, a cohort study, or a clinical trial. The data input screen is as follows:



The four values required for a sample size calculation are:

- **Two-sided confidence level** – most individuals would choose a 95% confidence interval, but a different confidence interval could be entered.
- **Power** – most individuals choose a power value of 80% or 90%, however, any power level can be entered.
- **Ratio of Unexposed to Exposed in sample** – place the desired ratio of unexposed individuals to exposed individuals. If there are to be an equal number of unexposed and exposed, then enter the value of 1.0; if there are to be twice as many unexposed as exposed, enter the value of 2.0. Any other ratio can be entered.
- **Percent of Unexposed with Outcome** – enter an estimate of the percentage of unexposed individuals that will develop (or have) the outcome of interest. For example, in a randomized control trial, you would estimate the percentage of those in the comparison group that will develop the outcome of interest during the trial. In a cohort study, enter the percentage of unexposed individuals who will develop the outcome of interest during the study. In a cross-sectional study, enter the estimated prevalence of disease among the unexposed.

The user has the choice of entering an odds ratio ***or*** the percent of exposed with the outcome of interest or the risk (prevalence) ratio or the risk (prevalence) difference - just enter one of these. The results using the default values for a risk ratio of 2 are below:

Sample Size for Cross-Sectional & Cohort Studies & Clinical Trials

Two-sided significance level(1-alpha):	95
Power(1-beta, % chance of detecting):	80
Ratio of sample size, Unexposed/Exposed:	1
Percent of Unexposed with Outcome:	5
Percent of Exposed with Outcome:	10
Odds Ratio:	2.1
Risk/Prevalence Ratio:	2
Risk/Prevalence difference:	5

	Kelsey	Fleiss	Fleiss with CC
Sample Size - Exposed	437	436	475
Sample Size- Nonexposed	437	436	475
Total sample size:	874	872	950

References

Kelsey et al., Methods in Observational Epidemiology 2nd Edition, Table 12-15
Fleiss, Statistical Methods for Rates and Proportions, formulas 3.18 &3.19

CC = continuity correction

The sample size formula for the method described in Kelsey et. al. is:

$$n_1 = \frac{\left(Z_{\alpha/2} + Z_{1-\beta}\right)^2 \bar{p}\bar{q}(r+1)}{r(p_1 - p_2)^2}$$

and

$$n_2 = r\,n_1$$

where

n_1 = number of exposed

n_2 = number of unexposed

$Z_{\alpha/2}$ = standard normal deviate for two-tailed test based on alpha level (relates to the confidence interval level)

Z_{β} = standard normal deviate for one-tailed test based on beta level (relates to the power level)

r = ratio of unexposed to exposed

p_1 = proportion of exposed with disease and $q_1 = 1-p_1$

p_2 = proportion of unexposed with disease and $q_2 = 1-p_2$

$$\bar{p} = \frac{p_1 + r p_2}{r+1} \quad \text{and} \quad \bar{q} = 1 - \bar{p}$$

The sample size formula *without* the correction factor by Fleiss is:

$$n_1 = \frac{\left[Z_{\alpha/2}\sqrt{(r+1)\bar{p}\bar{q}} + Z_{1-\beta}\sqrt{p_1q_1 + p_2q_2}\right]^2}{r(p_1 - p_2)^2}$$

$$n_2 = r\,n_1$$

For the Fleiss method ***with*** the correction factor, take the sample size from the uncorrected sample size formula and place into the following formula:

$$n_{1c} = \frac{n_1}{4}\left[1 + \sqrt{1 + \frac{2(r+1)}{n_1 r|p_2 - p_1|}}\right]$$

$$n_{2c} = r\,n_{1c}$$

When the input is provided as an odds ratio (OR) rather than the proportion of exposed with disease, the proportion of exposed with disease is calculated as:

$$p_1 = \frac{p_2 OR}{1 + p_2(OR - 1)}$$

When the input is provided as a risk (or prevalence) ratio (RR) rather than the proportion of exposed with disease, the proportion of exposed with disease is calculated as:

$$p_1 = p_2 RR$$

When the input is provided as a risk (or prevalence) difference (RD) rather than the proportion of exposed with disease, the proportion of exposed with disease is calculated as:

$$p_1 = RD + p_2$$

References

Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in Observational Epidemiology. Oxford University Press, 1996.

Fleiss JL. Statistical Methods for Rates and Proportions. John Wiley & Sons, 1981.

Updated Feb 16 2007: changed the "-" sign in the numerator of the Fleiss formula without a correction factor to "+".