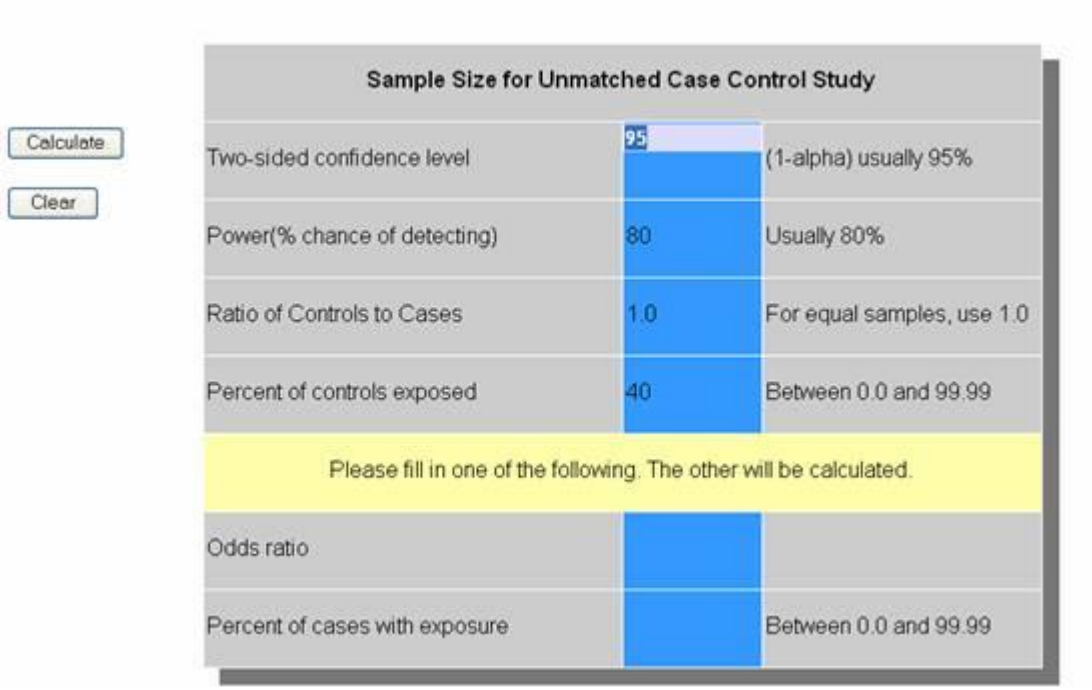
Feb 16 2007

**Documentation for**

# Sample Size for an Unmatched Case-Control Study

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

This module calculates a sample size for an unmatched case-control study. 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 Controls to Cases** – place the desired ratio of controls to cases. If there are to be an equal number of controls and cases, then enter the value of 1.0; if there are to be twice as many controls as cases, enter the value of 2.0. Any other ratio can be entered.
- **Percent of controls exposed** – enter an estimate of the percentage of controls that have (or had) the exposure of interest. For example, in a case-control study on cigarette smoking and lung cancer, among the controls (those without lung cancer), what percent would be expected to say they smoked cigarettes?

The user has the choice of entering an odds ratio ***or*** the percent of cases with the exposure of interest – just enter one of these, not both. The results using the default values for an odds ration of 2 are below:

# Sample Size for Unmatched Case-Control Study

For:

| | |
|---|---|
| Two-sided confidence level(1-alpha) | 95 |
| Power(% chance of detecting) | 80 |
| Ratio of Controls to Cases | 1 |
| Hypothetical proportion of controls with exposure | 40 |
| Hypothetical proportion of cases with exposure: | 57.14 |
| Least extreme Odds Ratio to be detected: | 2.00 |

| | Kelsey | Fleiss | Fleiss with CC |
|---|---|---|---|
| Sample Size - Cases | 134 | 133 | 144 |
| Sample Size - Controls | 134 | 133 | 144 |
| Total sample size: | 268 | 266 | 288 |

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
Results are rounded up to the nearest integer.

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 cases

$n_2$ = number of controls

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

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

r = ratio of controls to cases

$p_1$ = proportion of cases with exposure and $q_1 = 1-p_1$

$p_2$ = proportion of controls with exposure 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:

$$n_1 = \frac{\left[Z_{\alpha/2}\sqrt{(r+1)\bar{p}\bar{q}} + Z_{1-\beta}\sqrt{rp_1q_1 + p_2q_2}\right]^2}{r(p_1 - p_2)^2}$$

$$n_2 = r\,n_1$$

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 cases exposed, the proportion of cases exposed is calculated as:

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

**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 "+".