Matched Case-Control and Cohort Studies: Key Sample Size and Power Expressions

Define d_i as the number of cases (for case-control studies) or the number exposed (in cohort studies) and m_i as the number of controls in the i^{th} of N matched sets. Each matched set has $n_i = d_i + m_i$ subjects. Let z_{ij} denote the value, which can be quantitative or a 0/1 coded binary, of an exposure variable (for case-control) or an outcome (cohort studies) for the j^{th} variable in the i^{th} matched set. Let σ_i denote the standard deviation of z_{ij}. Note that for binary z_{ij} this is the square root of the Bernoulli variance $p_i(1-p_i)$. The parameter of interest, the log odds ratio, is denoted by θ . Using the conditional logistic regression model likelihood function, Lachin (2008) derives a test statistic $U(\theta)$ as the sum over the N matched sets of the differences within the sets between the sum $s_i = \sum_j^{d_i} z_{ij}$ for the actual d_i subjects with the case (case-cohort) or exposed (cohort) affiliations and the average of this sum over all arbitrary choices of d_i subjects from the matched set. Note there are a larger number of combinations, given by $\binom{n_i}{d_i}$, of selecting such choices when d_i and m_i are both > 1. Further larger d_i makes the sum s_i as well it's expectation more informative. These features help estimate the effect within each matched set more efficiently and result in a more sensitive test of effect using the aggregated test statistic. The distribution of the test statistic under general alternatives θ is derived as

$$U(\theta_0)|H_1 \sim N[\theta I(\theta_0), I(\theta)]$$

The expression for power for two sided alternatives follows from this as in the following

$$\Phi\left(|\theta|\sqrt{I(\theta)} - Z_{1-\alpha/2}\right),$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal. The null variance $I(\theta_0)$ is given by

$$I(\theta_0) = \sum_{i=1}^{N} d_i \sigma_i^2 \left[\frac{\binom{n_i}{d_i} - 1}{\binom{n_i}{d_i}}\right] = \sum_{i=1}^{N} d_i \sigma_i^2 \left[\frac{n_i! - d_i!\, m_i!}{n_i}\right]$$

When n, d, m and σ are fixed or identical across matched sets, the expression simplifies to

$$I(\theta_0) = N d \sigma^2 \left[\frac{\binom{n}{d} - 1}{\binom{n}{d}}\right] = N d \sigma^2 \left[\frac{n! - d!\, m!}{n!}\right]$$

This further simplifies for 1 to m matches to

$$I(\theta_0) = \frac{N m \sigma^2}{m+1}$$

In fixed matched sets, the number N of matched sets needed for power $(1-\beta)$ for a two-sided test at a significance level of α is given, for the d to m and the 1 to m contexts respectively, by

$$N = \frac{\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{\theta^2 d \sigma^2 \left[\frac{n! - d!\, m!}{n!}\right]} \quad and \quad N = \frac{(m+1)\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{m \theta^2 \sigma^2},$$

Consistent with the earlier note, we see the increased number of combinations $\binom{n}{d}$ and increased d leading to smaller sample sizes and more power. Our discrete calculator uses the Bernoulli variance based on the pooled proportion as the variance σ^2 . The continuous calculator uses a unit variance, as the odds ratio is based on a difference of a standard deviation – this effectively scales the continuous variable to a standard deviation and variance of 1.0.

Details on the power and sample size for alternate unconditional tests are provided in the next page.

Return to the calculator at this link.

Comparisons to Power and Sample Size for Unconditional Tests

The power of the unconditional test (expressions applicable to both binary and continuous data) for the log odds ratio, using similar notation to that used earlier, is given by (all expressions adapted from the EAST Manual)

$$\Phi\left(|\theta|\sigma\sqrt{N_1N_2/(N_1+N_2)} - Z_{1-\alpha/2}\right),$$

Where, N_1 is the total number of cases (case-control) or number exposed (cohort)and N_2 is the total number of controls. This simplifies to

$$\Phi\left(|\theta|\sigma\sqrt{Nmd/(m+d)} - Z_{1-\alpha/2}\right) \quad and \quad \Phi\left(|\theta|\sigma\sqrt{Nm/(m+1)} - Z_{1-\alpha/2}\right),$$

for fixed composition matched sets for the *d* to *m* and 1 to *m* contexts respectively. Similarly the number of matched sets is given by

$$N = \frac{(m+d)\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{md\theta^2\sigma^2} \quad and \quad N = \frac{(m+1)\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{m\theta^2\sigma^2},$$

Notice that the sample size and power for unconditional tests is identical to that for the conditional test for fixed 1 to *m* matched sets. The conditional test dominates when both *m* and *d* are > 1.

The power for the test of differences in proportions, given by $p_1 - p_2$, using a pooled proportion to compute variance, is identical to that for the chi-squared test. This is given by (all expressions adapted from the EAST Manual)

$$\Phi\left(|p_1 - p_2|\sigma^{-1}\sqrt{N_1N_2/(N_1+N_2)} - Z_{1-\alpha/2}\right)$$

For fixed matched sets for the *d* to *m* and 1 to *m* contexts respectively, power is given by

$$\Phi\left(|p_1 - p_2|\sigma^{-1}\sqrt{Nmd/(m+d)} - Z_{1-\alpha/2}\right) \quad and \quad \Phi\left(|p_1 - p_2|\sigma^{-1}\sqrt{Nm/(m+1)} - Z_{1-\alpha/2}\right)$$

Sample size is given by

$$N = \frac{(m+d)\sigma^2\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{md(p_1 - p_2)^2} \quad and \quad N = \frac{(m+1)\sigma^2\left(Z_{1-\alpha/2} + Z_{1-\beta}\right)^2}{m(p_1 - p_2)^2}$$

As with the unconditional test for the odds ratio, the sample size and power for the chi-squared test is similar to that for the conditional test for fixed 1 to *m* matched sets. The conditional test dominates when both *m* and *d* are > 1.

References:

Lachin JM (2008). Sample Size Evaluation for a Multiply Matched Case-Control Study Using the Score Test From a Conditional Logistic (Discrete Cox PH) Regression Model. *Statistics in Medicine* June 30; 27(14): 2509-2523.

Cytel EAST 6 User Manual (2014).

Return to the calculator at this link.