# Sample Size based on OR CI Width and the Delta Method

Let us use the delta method to determine an approximation to the variance of the estimate of the log odds

$$\log(\text{odds}) = \log(p/(1-p)) \tag{1}$$

$$\text{the estimate of the log odds is: } \log(\hat{p}/(1-\hat{p})) \tag{2}$$

$$\text{var(estimate of } \log(\text{odds})) \approx (d\log(\text{odds})/dp)^2 \text{var}(\hat{p}) \tag{3}$$

$$= (\frac{1}{(p(1-p)})^2 p(1-p)/n = (1/p + 1/(1-p))/n \tag{4}$$

The log odds ratio is

$$\log(\text{odds ratio}) = \log(p_1/(1-p_1)) - \log(p_0/(1-p_0) \tag{5}$$

The estimate of the log odds ratio is

$$\log(\hat{p}_1/(1-\hat{p}_1)) - \log(\hat{p}_0/(1-\hat{p}_0)) \tag{6}$$

Since $\hat{p_0}$ and $\hat{p_1}$ are assumed to be statistically independent, we see that the variance of the estimate of the log(odds ratio) is approximated by

$$(1/p_0 + 1/(1-p_0))/n_0 + (1/p_1 + 1/(1-p_1))/n_1 \tag{7}$$

which may be written as

$$\frac{1}{n_0 p_0(1-p_0)} + \frac{1}{n_1 p_1(1-p_1)} \tag{8}$$

Now, if we replace $p_0$ and $p_1$ by their estimates $\hat{p}_0$ and $\hat{p}_1$, we get

$$1/a + 1/b + 1/c + 1/d \tag{9}$$

This expression was noted by Woolf(1955). It is certainly an estimate of the variance of the estimate of the log odds ratio. The previous expression can be written as

$$\frac{1}{n_0}(\frac{1}{p_0(1-p_0)}+\frac{1}{kp_1(1-p_1)}) \tag{10}$$

where $k = n_1/n_0$ Based on the above, the expected half width of a $1-\alpha$ confidence interval for the log odds ratio would then be:

$$\frac{z_{\alpha/2}}{\sqrt{n_0}}\sqrt{(\frac{1}{p_0(1-p_0)}+\frac{1}{kp_1(1-p_1)})} \tag{11}$$

The lower limit and upper limit of the corresponding confidence interval for the odds ratio [relative to this odds ratio] would then be:

$$\exp(-\frac{z_{\alpha/2}}{\sqrt{n_0}}\sqrt{(\frac{1}{p_0(1-p_0)}+\frac{1}{kp_1(1-p_1)})}) \tag{12}$$

$$\exp(\frac{z_{\alpha/2}}{\sqrt{n_0}}\sqrt{(\frac{1}{p_0(1-p_0)}+\frac{1}{kp_1(1-p_1)})}) \tag{13}$$

The relative lower width (R) is the distance between the lower limit and the estimate of the odds ratio [relative to the odds ratio]. Therefore 1-R is this lower limit [displayed above]. It is important to remember that this interval for the odds ratio is not symmetrical about the odds ratio. If the odds ratio determination is set up so that one is anticipating an odds ratio that is greater than 1, then the lower limit of the confidence interval has more relevance than the upper limit. Hence the definition for R used here.

For sample size determination, we have:

$$\log(1-R) \leq \frac{-z_{\alpha/2}}{\sqrt{n_0}}\sqrt{(\frac{1}{p_0(1-p_0)}+\frac{1}{kp_1(1-p_1)})} \tag{14}$$

Rearranging, we get

$$n_0 \geq \frac{z_{\alpha/2}^2}{\log^2(1-R)}(\frac{1}{p_0(1-p_0)}+\frac{1}{kp_1(1-p_1)}) \tag{15}$$

And so we get

$$n_1 = kn_0 \tag{16}$$

Tables using the formula can be found at:
http://apps.who.int/iris/handle/10665/40062
A calculator that uses this formula can be found at:
http://www.select-statistics.co.uk/sample-size-calculator-odds-ratio
R code is listed below :

```
ssor<-
function (p0, p1, odds.ratio, r, k = 1, alpha = 0.05)
{
    z.alpha <- qnorm(1 - alpha/2)
    if (!missing(odds.ratio))
        p1 <- p0 * odds.ratio/(1 - p0 + p0 * odds.ratio)

    n0 <- z.alpha^2/(log(1-r))^2 * (1/p0/(1-p0) + 1/k/p1/(1-p1))
    n1 <- k * n0
    c(n0 = n0, n1 = n1)
}
```