

Notes on sample sizes required for Receiver Operator Characteristic (ROC) curves

14 November 2018

1 Calculating sample size needed to estimate area under an ROC curve given a specified CI and power

1.1 Calculating ratio of number of non-cases to cases

We first need to calculate k which is the ratio of number of non-cases to cases. This can be calculated as follows:

$$k = \frac{1 - PREV_p}{PREV_p}$$

where :

$PREV_p$ = prevalence of the cases in the population

Given that we don't have specific values for the prevalence of children demonstrating appropriate infant and young child feeding, we make an educated guess of what this prevalence can be. It is likely that the practice of appropriate infant and young child feeding is low and is probably at best at the 20% level.

Using this, we estimate k as follows:

$$\begin{aligned} k &= \frac{1 - PREV_p}{PREV_p} \\ &= \frac{1 - 0.2}{0.2} \\ &= 4 \end{aligned}$$

1.2 Calculating the binormal distribution parameter

We then need to calculate the binormal distribution parameter A as follows:

$$A = \phi^{-1}(AUC) \times 1.414$$

where :

$$\phi^{-1} = \text{inverse of the cumulative normal distribution function}$$

$$AUC = \text{expected area under the curve}$$

Since we do not have prior knowledge of the AUC from previous studies, we set AUC at 0.80 which is the AUC value we would assume that would show agreement between ICFI and indicators for nutrition and energy adequacy. This gives us the following A parameter:

$$A = \phi^{-1}(AUC) \times 1.414$$

$$= \phi^{-1}(0.80) \times 1.414$$

$$= 0.25 \times 1.414$$

$$= 1.190052$$

1.3 Calculating the variance function

To be able to continue with the calculations, we will need to calculate variance function (VF). The variance function (VF) is calculated as follows:

$$VF = 0.0099 \times e^{-A \times A/2} \times \left[(5 \times A^2 + 8) + \frac{(A^2 + 8)}{k} \right]$$

where :

$$A = 1.190052$$

$$k = 4$$

Using the values calculated for A and k previously, we calculate VF as follows:

$$\begin{aligned}
VF &= 0.0099 \times e^{-A \times A/2} \times \left[(5 \times A^2 + 8) + \frac{(A^2 + 8)}{k} \right] \\
&= 0.0099 \times 0.4925731 \times \left[(5 \times 1.416225 + 8) + \frac{(1.416225 + 8)}{4} \right] \\
&= 0.004876474 \times [15.08112 + 2.354056] \\
&= 0.004876474 \times 17.43518 \\
&= 0.0850222
\end{aligned}$$

1.4 Calculating number of cases needed in the study sample

Now we can calculate the number of cases (i.e. those who have the condition of interest). For the case of ICFI or IYCF, this will be those children who exhibit or demonstrate appropriate infant and young child feeding. This can be calculated as follows:

$$N = \frac{Z_{\alpha/2}^2 \times VF}{L^2}$$

where :

$$Z_{\frac{\alpha}{2}} = 1.96 \text{ for a 95\% CI}$$

$$L = 0.05 \text{ (desired half-width of the CI)}$$

$$VF = 0.0850222$$

Using the value for VF calculated previously, we calculate N as follows:

$$\begin{aligned}
N &= \frac{Z_{\alpha/2}^2 \times VF}{L^2} \\
&= \frac{1.96^2 \times 0.0850222}{0.05^2} \\
&= 130.6485 \approx 131
\end{aligned}$$

1.5 Total sample size

The total sample size n needed for the ROC analysis can then be calculated as follows:

$$n = N \times (1 + k)$$

where :

$$N = 131$$

$$k = 4$$

Using the N and k values calculated earlier, n is calculated as follows:

$$n = N \times (1 + k)$$

$$= 131 \times 5$$

$$= 655$$

We will need a sample size of about **655** children about **131** of which will be children who are practising appropriate infant and young child feeding.