#### Comparing prevalence to threshold

Consider the case where there is a threshold value of prevalence, , and we wish to use a cluster pool survey to demonstrate that the prevalence is above or below this threshold. Using asymptotic normality of the MLE we can calculate minimum sample sizes required for target power, , confidence level, , and a design prevalence, . We assume that the pooling design will consist of some number, , of replicates of a pooling design. In a non-cluster survey this we could summarise this with pooling design . For a cluster survey this might require locations each with pooling design . In either case the total number of units sampled is . The standard error at the design prevalence, , and standard error at the threshold prevalence, , are

The power to reject one-sided and two-sided null hypotheses can be approximated as

where is the quantile function of the standard normal distribution. For one-sided hypothesis tests, if the power is to exceed , must be at least

#### Comparing prevalence in two samples

Now consider the case where we conduct pooled cluster surveys in two populations; for instance, surveys in two separate regions, or surveys at two time points before and after an intervention in a single region. Here we assume that the samples are statistically independent. There are now two design prevalences, and , and two pooling designs consisting of and replicates of represented by and . In the cluster sampling case there may also be two correlations and . In both the simple random and cluster sampling cases, the total number of units across both samples is . Using the independence of the samples and the asymptotic normality of each MLE of prevalence, the difference in the two MLEs is approximately normal with standard error,

where and are defined similarly to the previous section, with the generalisation to different pooling designs: