Sample size calculation to assess the prevalence of Scary Disease-x! at the herd-level

April 11, 2014

To perform a smaple size calculation, we need some references about the number of herds in the population and the expected prevalence of Scary Disease-x! at the herd-level. These assumptions are included in table 1.

Variable	Assumption	Justification
Number of herds in the population	1000	
Among-herd prevalence	30%	
Statistical confidence	95%	
Precision of posterior estimate	10%	
Test Sensitivity	80%	
Herd Sensitivity	95%	
Herd Specificity	100%	
Average herd size measured in sampling units	25	
Number of samples per farm	5	
Within-herd prevalence	50%	

Table 1: Description of assumptions that were made in the sample size calculations and the justification for these assumptions

We can start with a simple calculation of sample size based on the assumption of an infinite population and our expected prevalence

$$n = \left[\frac{Z_\alpha}{L}\right]^2 \cdot pq \tag{1a}$$

Where n is the sample size, Z_α is a z-score of 1.96 for 95% confidence, p is the expected prevalence, q = 1-p, and L is the acceptable error surrounding the point estimate of the prevalence

Assuming that the expected prevalence at the herd level is 30%, then we can estimate the crude herd level prevalence within 10% of the true value in 95% of study replications if we sample 81 herds.

We can make this more precise by accounting for an imperfect test like this:

$$n = \left[\frac{Z_\alpha}{L}\right]^2 \cdot \frac{[HSe \cdot p + (1 - HSp) \cdot q] \cdot [1 - HSe \cdot p - (1 - HSp) \cdot q]}{(HSe + HSp - 1)^2} \tag{2a}$$

$$HSe_{\substack{finite\\population}} = 1 - \left(1 - \frac{n_{withinherd} \cdot Se}{N_{withinherd}}\right)^{p_{withinherd} \cdot N_{withinherd}} \tag{3a}$$

$$HSe_{\substack{infinite\\population}} = 1 - (1 - p_{withinherd} \cdot Se)^{n_{withinherd}}) \tag{3b}$$

If we assume that Se (Test sensitivity) is 80% and in each herd we test 5 units and the average herd has 25 and that positive herds have a within-herd prevalence of 50% then this results in a HSe of 88.7% and we will need to test 95 in order to be 95% confident of estimating the true prevalance within 10% of the true value.

Finally, we can adjust the sample size for a finite population if greater than 10 % of herds are being sampled as follows:

$$n_{adjusted} = \frac{n}{1 + \frac{n-1}{N}} \tag{4a}$$

Because 9.5% of the herds are being sampled then we should assume an infinite population and still sample the 95 herds.