

## **ANNEX L**

### **SAMPLE SIZE CALCULATION**

## Sample size calculation

To estimate the sample size, we simplify the variance estimator shown in Annex J, equations (4) and (5), by assuming a two stage simple random sample design, which is less efficient than our actual sample design; hence yielding a conservative estimate of the variance. To get an analytic closed form approximation for the variance estimator, we assume that each community has the same size of population of potential respondents ( $M$ ) and that we draw a sample of  $m$  respondents from each community. We also conservatively assume that the number of communities in the frame ( $N$ ) is very large, so that terms such as  $1/N$  and  $n/N$  can be neglected. Under this simpler design, the variance estimator is approximately:

$$1) \hat{V}(\hat{y}) \cong \left[ \frac{1+(m-1)\rho}{nm} \right] \bar{y}(1-\bar{y}),$$

where  $\bar{y}$  is the population proportion being measured,  $n$  is the number of sampled communities, and  $\rho$  is the intraclass (within community) correlation for variable  $y$ .

Applying equation (1), the estimated number of sample communities ( $n$ ) necessary to achieve a given margin of error ( $e$ ), assuming a 95% confidence interval, is given by:

$$2) n \cong \left[ \frac{1+(m-1)\rho}{m} \right] \bar{y}(1-\bar{y}) \frac{(1.96)^2}{e^2}$$

Figure L.1 shows values of  $n$  for various values of  $e$  and  $\rho$ , assuming conservatively that  $\bar{y} = 0.5$ . We assume that  $m = 8$ , as in our sample design (whenever  $M \geq 8$ ). If we target a margin of error of 0.05 (10% of  $\bar{y}$ ), the estimated number of sample communities required ranges from 48 if  $\rho = 0$  to 384 if  $\rho = 1$ . We expect  $\rho$  to be less than 0.5, and take 0.3 as a reasonable value. For this value of  $\rho$ , the estimated sample size required is 149 communities. Based on this, we selected 150 communities as our sample size.

If our assumptions prove to be incorrect, the actual margin of error will be different than that estimated. The sensitivity of  $e$  to variations in our assumptions is not very large if  $n = 150$ ; ranging from 0.028 if  $\rho = 0$  to 0.08 if  $\rho = 1$ . However, the range of  $e$  for different values of  $\rho$  becomes larger for smaller values of  $n$ ; especially for sample sizes smaller than 100. Given lack of prior information about the true value of  $\rho$ , a community sample size of at least 150 appears prudent.

Figure L.1. Required sample size as a function of margin of error and intraclass correlation

