Appendix A. Example of Alternative Strategy for Follow-Up Sampling

BMP Ground Verification Considerations

Overview

This appendix provides a statistics-based approach for post-implementation verification that BMPs are still on the ground (or otherwise continue to be implemented) and still performing as expected based on engineering specifications or other applicable criteria. This approach does not address initial implementation of BMPs.

The measure of choice for this approach is the proportion (percentage) of implemented BMPs (1) still in place or (2) still performing in accordance with expectations. The approach described here addresses how to compute the sample size necessary to estimate these proportions (i.e., "p" or proportion of "Yes" responses and "q" or proportion of "No" responses) with the desired degree of confidence and a specified acceptable error (±X%) using simple random sampling. No hypothesis testing, comparison of proportions, or trend analysis is considered.

The following are data requirements for the sample size (n) calculations described in this appendix:

- An initial estimate of both the percent of BMPs still in place and the percent of BMPs still performing as expected. This can be based on previous studies or assumed to be 50% (p=0.5) for a conservative (high) estimate of sample size.
- An allowable error (e.g. ±5% or 0.05). This error (d) can be different for different BMPs based on considerations of BMP importance, risk of BMP abandonment, failure, cost, or other factors.
- A confidence level (e.g., 90% or α=0.10). This is used to determine the 2-sided Z score from the standard normal distribution ($Z_{1-\alpha/2}$), e.g., $Z_{1-\alpha/2}$ is equal to 1.645 for α = 0.10.
- An estimate of the total population (N) from which the sample is taken (e.g., how many BMPs were installed). This can be based on records of BMP implementation.

States may choose to perform stratified random sampling where differences in p values are likely to exist due to such factors as the mechanism or program through which the BMP is implemented (e.g., cost-shared, permits, voluntary with no cost-share), location (e.g., critical areas, cropping region), or other factors. Random sampling is then performed within the different strata.

Considerations for Estimating Sample Size

Important considerations that must be considered before computing sample size include the following:

Define Objective:

- i.e., Quantify proportion of implemented BMPs that are:
  - Still in place (yes/no)
  - Performing in accordance with expectations (yes/no)
  - Define precision requirements (i.e., the true proportion lies between *p-d* and *p+d* with a $1-\alpha$ confidence level)

Define population and sampling units

- e.g., All BMPs implemented during a specified timeframe (e.g., this year)

Define stratification

- e.g., Are there important differences between cost-shared voluntary, regulatory, and non-cost-shared voluntary implementation?
- e.g., Are there important geographic differences?

Sample size calculations

- See below for simple random sampling discussion.

Random selection process

- Are there certain BMPs that cannot be field verified; will this create a bias?
- Can the population be enumerated (i.e., can all BMPs be labeled and randomly selected from)?
- What is the level of effort (i.e., cost) to do a ground verification?

Field methodology

- How will field information be collected?
- What will be done to ensure consistency between different field technicians?
- How will judgments be made; re: whether answer is "yes" or "no"? (States may want to collect/retain more specific information for other purposes, but this approach requires a simple "yes" or "no".)
- Will there be an opportunity to field test and refine the approach?

Probabilistic Sampling

Most study designs that are appropriate for ground verification are based on a probabilistic approach because visiting each site is not cost-effective. In a probabilistic approach, individuals are randomly selected from the entire group. The selected individuals are evaluated, and the results from the individuals provide an unbiased assessment about the entire group. Applying the results from randomly selected individuals to the entire group is *statistical inference*. Statistical inference enables one to determine, in terms of probability, for example, the percentage of implemented multi-year BMPs that are still in place without visiting every site.

The group about which inferences are made is the population or *target population*, which consists of *population units*. The *sample population* is the set of population units that are directly available for measurement. Statistical inferences can be made only about the target population available for sampling. For example, if only a certain class of BMPs can be ground verified (e.g., cost-shared BMPs), then inferences cannot be made about other classes of BMPs that could not be ground verified (e.g., voluntarily implemented BMPs with no cost-share). States will need to consider carefully how they define their population units for each BMP. For example, structural practices such as animal waste lagoons are easily defined whereas nutrient management plans that apply to multiple fields on the same farm are more complex. With nutrient management one might consider that the population unit is defined as the farm-wide nutrient management plan, or as the collective of individual fields upon which nutrient management is practiced. These decisions must be made before the target population can be quantified.

The most common types of sampling that should be used are either simple random sampling or stratified random sampling. *Simple random sampling* is the most elementary type of sampling. Each unit of the target population has an equal chance of being selected. This type of sampling is appropriate when there are no major trends, cycles, or patterns in the target population. If the pattern of BMP presence or performance is expected to be uniform across the geographic area of interest (e.g., state), simple random sampling is appropriate to estimate the proportion of BMP presence or performance. If, however, implementation is homogeneous only within certain categories (e.g., region of state, cost-shared vs. non-cost-shared), stratified random sampling should be used.

In *stratified random sampling*, the target population is divided into groups called strata for the purpose of obtaining a better estimate of the mean or total for the entire population. Simple random sampling is then used within each stratum. Stratification involves the use of categorical variables to group observations into more units (e.g., cost-shared vs. non-cost-shared), thereby reducing the variability of observations within each unit. In general, a larger number of samples should be taken in a stratum if the stratum is more variable, larger, or less costly to sample than other strata.

If the state believes that there will be a difference between two or more subsets of the sites, the sites can first be stratified into these subsets and a random sample taken within each subset. The goal of stratification is to increase the accuracy of the estimated mean values over what could have been obtained using simple random sampling of the entire population. The method makes use of prior information to divide the target population into subgroups that are internally homogeneous. There are a number of ways to "select" sites to be certain that important information will not be lost, or that results will not be misrepresented. One current approach is Generalized Random Tessellation Stratified (GRTS) survey design (Stevens and Olsen 2004).

Sample Size Calculations with Simple Random Sampling

In simple random sampling, we presume that the sample population is relatively homogeneous and we would not expect a difference in sampling costs or variability. If the cost or variability of any group within the sample population were different, it might be more appropriate to consider a stratified random sampling approach.

To estimate the proportion of BMPs still in place or still performing as expected (p), such that the allowable error, *d*, meets the study precision requirements (i.e., the true proportion lies between *p-d* and *p+d* with a *1-α* confidence level), a preliminary estimate of sample size (n_0) can be computed with the following equation assuming a large population from which to sample (Snedecor and Cochran, 1980):

$$n_o = \frac{(Z_{1-\alpha/2})^2 pq}{d^2} \tag{1}$$

In many applications, the number of population units in the sample population (*N*) is large in comparison to the population units sampled (*n*) and the *finite population correction term* ($1-\varphi$) can be ignored. However, depending on the number of units (e.g., expensive or unique BMPs) in a particular population, *N* can become quite small. *N* is determined by the definition of the sample population and the corresponding population units. If φ is greater than 0.1, the finite population correction factor should not be ignored (Cochran, 1977). Thus, the final sample size (n) is then estimated as (Snedecor and Cochran, 1980)

$$n = \begin{cases} \dfrac{n_0}{1+\varphi} & \text{for } \varphi > 0.1 \\ n_o & \text{otherwise} \end{cases} \tag{2}$$

where φ is equal to n_o/N.

Terms:

N = total number of population units in sample population

n = number of samples

p = proportion of “yes” responses

q = proportion of “no” responses (i.e., 1-p)

n_0 = preliminary estimate of sample size

$\phi = n_0/N$ unless otherwise stated

$Z_{1-\alpha/2}$ = value corresponding to cumulative area of $1-\alpha/2$ using the normal distribution

d = allowable error

Example: If a previous analysis indicated that 91% of a specific BMP installed via a state cost-share program was still in place at a specified point in time (e.g., a year or a number of years after installation), then one could assume p=0.91 and q=0.09 to determine sample size for the next assessment of that BMP. Assuming that an error of ±10% is acceptable for this estimate (i.e., d=0.10), the preliminary sample size estimate (n_0) can then be directly computed from the above equation. It may be that a given BMP might not meet the requirements for a large population assumption and that the sample size needs to be adjusted by a finite population correction term.

Estimates of required sample size for a range of values for both α and N are given in Table 1. Note that the finite population correction term was applied to some values in Table 1 (e.g., values for N=100 and 250 with α=0.05). So for example, if there were 250 installed BMPs and the assumption of a 91% continued presence or operation were observed, a 9% sampling (22/250) would allow the Chesapeake Bay program to state with 90% confidence that the true proportion of BMPs still in operation ranged from 81 to 100%. Larger sample sizes would be needed if the true proportion of BMPs still in operation approaches 50% or a smaller value of d is needed.

Table 1. Number of samples required as a function of α and population size (N).

α	Number of Population Units (N)						
	100	250	500	750	1000	2000	Large N
0.05	24	28	31	31	31	31	31
0.10	18	22	22	22	22	22	22
0.20	12	13	13	13	13	13	13
Assumes p=0.91, q=0.09, and d=0.10							

References

Cochran, W.G. 1977. Sampling Techniques. 3rd ed. John Wiley and Sons, New York, New York.

Snedecor, G.W. and W.G. Cochran. 1980. *Statistical methods*. 7th ed. The Iowa State University Press, Ames. Iowa.

Stevens, D. L., Jr. and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99:262-278.