Why Small Sample Sizes Are Accurate in Analyzing Large Areas

John N. Lovett, Jr., Ph.D., P.E., QuantiTech, Inc.

It is a basic principle of statistical sampling that a conclusion may be drawn from a large population of data based on a relatively small sample taken from that data, with a certain degree of statistical confidence. This principle has been applied to polling, for example, for many years. By sampling the opinions of a small group of voters (say 1,000), representing the voting public, a reasonable conclusion may be drawn about an election in which tens of millions of voters participate. Typically, the percentage of voters sampled may be less than 0.002% of the total, though the pollster may be 95% confident that the conclusion is correct.

There are some assumptions that must hold for the conclusion to be valid statistically. In terms of sampling an ordnance site, the sample "areas of opportunity" for the discovery of an ordnance item (say unexploded ordnance, or UXO) are usually grids or transect areas. Using grids as an example, an assumption is that the grids are placed randomly throughout the Area of Investigation (AOI). Also, the ordnance is assumed to occur homogeneously within the AOI. This does *not* mean the ordnance is uniformly distributed; it only means the occurrence of ordnance items follows a Poisson statistical distribution. Therefore, there may be certain areas of high density and low density ordnance, as long as it is not "clustered."

This principle is illustrated in many statistical texts through the use of operating characteristic curves. These probability curves show the changes in the confidence level of a conclusion based on increasing levels of UXO, for example. They may be developed for any population size (total area of the AOI, usually expressed in acres) and sample size (total area of the sample grids or transects used in the field work). The curves will exhibit very little variation for a given confidence level as the AOI area increases dramatically.

The statistical tool used for several years to determine the minimum amount of sampling for an AOI (assuming no UXO are found) is **UXO Calculator**. Its second generation counterpart is now termed **UXO Estimator**. They may also be employed if UXO are found, but in that case slightly more sampling will be required for the same degree of confidence. Using either of these tools, the following example illustrates the insensitivity of the sample size to the AOI area, assuming a 95% confidence level for the conclusion (which is typical), and a "target UXO density" of 0.1 UXO per acre:

AOI Area (acres)	Sample Area (acres)
1,000	29.514
5,000	29.868
10,000	29.913
100,000	29.952

Notice that, even though the AOI acreage changes one hundred fold, the sample area increases only slightly for the same degree of confidence. The pollsters therefore use this same principle to arrive at their predictions.