

## April 2003 Using Statistics in Sampling?

Questions sometimes asked about swab sampling are “How many swab sites must I sample?” and “Can I determine the number of swab sites statistically?” While statistics is critical for many sampling purposes, one must be careful about applying statistical principles to the sampling of different locations in a manufacturing vessel. The purpose of picking swab locations is not to determine exactly the total amount of residue in the equipment. Rather swab sites are picked based on “worst-case” principles. For example, those locations in the vessel that are most difficult to clean are selected and swabbed. If those locations are within the acceptance limits, then other locations (that are easier to clean, and hence should have lower residue) should also be within the acceptance limits. The number of locations that are “most difficult to clean” are not determined statistically, but are determined by a sound understanding of the equipment, of the manufacturing process and of the cleaning process. Experimental data may support the selection process, as well as prior experience and good scientific judgment. A 5000-liter bioreactor may have the same number of worst-case locations as a similar scaled-down 1000-liter bioreactor.

Statistics is not applicable because the point is not to statistically sample the equipment. Size is not applicable unless there is something about a different size that involves more (or fewer) worst-case locations for cleaning. But the point about size is that one cannot base the number of swab samples (or the surface area sampled) strictly based on size (surface area, for example). Here is a practical example. Suppose I have a 2000-liter simple storage vessel, and also have a 1000-liter bioreactor. Is it not likely that there will be more worst-case locations to perform swab sampling on the smaller bioreactor than on the larger storage vessel?

Although one could use either statistical principles or a certain percentage of surface area for picking the number of swab locations, its unclear what value that adds to a cleaning validation protocol. That said, one should also realize that it may be appropriate to select locations other than worst-case locations. This might include representative materials of construction and/or representative functional locations. Sampling these sites might add value as a diagnostic tool in a cleaning validation failure; if one obtains a failing result on an “easy-to-clean”, but representative location (for example, the sidewall of a tank well below the liquid /air interface), it may suggest a serious cleaning failure.

In contrast to other Cleaning Memos, this one is designed to suggest that certain practices, while not wrong, probably don't add significant value in terms of time and effort. Selecting the number of swab sample sites (when those sample sites are not equivalent) based on statistics or based on a percentage of surface area is one of these cases.