November 2005
Averaging Swab Sample Results?

In cleaning validation, limits are set for swab samples (usually based on a carryover calculation). The question is often asked, "Am I allowed to average different swab samples and compare that average to my calculated established acceptance limit?" This is usually asked in the context of having one (or more) swab samples exceed the established swab limit. Does the average mean that the overall result satisfies the acceptance criteria?

Well, if the protocol is written that such averaging is allowed, then the average does meet the acceptance criteria. However, in general such averaging of different swab locations should not be permitted in a cleaning validation protocol. Why? There are several concerns with such a practice. One is that the general practice in cleaning validation is to select those swab sampling sites which are most "difficult to clean", that is, most likely to have higher levels of residue following cleaning. The reason for "difficulty of cleaning" may include accumulation of product on the specific surface prior to cleaning (such as in a drain), drying of product on the surface prior to cleaning (such as at the air/liquid interface), or based on the cleaning process itself (such as the underside of an agitator blade). If the "difficult to clean" sites are chosen properly, and if the results for each of those locations is below the acceptance limit, then a reasonable conclusion is that any site sampled would be below the acceptance limit, and therefore the equipment would be appropriately cleaned.

Another reason for not averaging sites is that generally there are a number of other sites sampled which may not be "difficult to clean", but which represent different functional locations, different materials of construction, and surfaces where residues may be preferentially transferred to one portion of the next manufactured product. If this is the case, doesn't averaging all the swab results give an overall picture of residues in the equipment? The answer is NO. Picking swab samples based on the criteria discussed does not assure that averaging gives an overall picture. If the intent was truly to pick swab samples which "mapped" the entire surface of the equipment, one would need a statistician to determine the number and locations of sampling sites to provide this overall picture, and the number of sampling sites would increase exponentially. Unless a statistician were involved to select the appropriate number and locations of sampling sites, it would be easy to manipulate the process by sampling a large number of "easy to clean" locations (such as vessel sidewalls) so that no matter how many "failing" results I obtained on "difficult to clean" locations, my overall average would still allow me to be below my acceptance criterion.

Furthermore, the general principle in cleaning validation is that one determines the swab acceptance limit, and then designs a cleaning process such that the most "difficult to clean" sampling sites are below that acceptance criterion. Pre-validation work is to be done to establish that the acceptance criteria can be met before I execute my validation protocol. If I truly follow that scenario (having designed my cleaning process to reduce residues at those worst-case locations to an acceptable level), and then have failures at specific sites in my validation protocol, then something is amiss. I have clearly done something wrong in my process.

The objection is sometimes made that "Rinse sampling essentially is averaging residues over all surfaces. Why is it allowed in that case?" The reason is that in rinse sampling, one is truly sampling all surfaces, so that the rinse process in essence integrates data from the entire surface. In one sense, rinse sampling is the ultimate of having that statistical evaluation of swab sampling sites such that an overall picture is obtained. This is

assuming, of course, that rinse sampling limits are established appropriately, and that sampling recovery studies are done for rinse sampling.
For clarification, this discussion of rinse sampling above applies only to rinse sampling for cleaning validation protocol acceptance criteria, not to rinse sampling as a routine monitoring process for an already-validated cleaning process.

Furthermore, when a "swab sample" from a specific location is injected into an analytical instrument (such as HPLC) multiple times, one is allowed to average those results. In other words, I can inject a portion of the same sample three times, getting results of 5.1 ppm, 5.0 ppm, and 4.9 ppm, with an average result of 5.0 ppm. If my acceptance limit were ≤5.0 ppm, then that particular sample would meet my acceptance criterion. This is not a pretty situation, and I would hope that my results were not that borderline.

This Cleaning Memo is not designed to promote or to discourage the use of swab sampling. The intent is help insure that if swab sampling is performed, the results are handled appropriately both from a scientific and compliance viewpoint.