

**July 2009**  
**Revisiting Linearity of Swab Recovery Results**

Those of you familiar with my writings and seminars know that I am not of fan of doing sampling *recovery* studies (either swab sampling or rinse sampling) using *multiple* spiked levels to establish linearity of results. In part my view is based on the high variability of any individual recovery experiment. In other words, if one person does three replicates and gets recoveries of 60%, 66% and 72%, that most likely is the same data (it is not *significantly* different). So, what expectation is there, if I spike at five different levels, that a plot of “percent recovery” versus “spiked amount” would be linear? For example, with an acceptance criterion in a protocol of 2  $\mu\text{g}/\text{cm}^2$ , one might spike at levels of levels of 1.0, 1.5, 2.0, 2.5 and 3.0  $\mu\text{g}/\text{cm}^2$  (corresponding to 50%, 75%, 100%, 125% and 150% of the residue level at the acceptance limit). And for those expecting (or even requiring linearity), there should be a nice linear relationship between percent recovery and amount spiked.

Now one reason I don't like this approach is that spiking above the acceptance criteria is not *useful* information (unless I think I might sample for the same residue in a different study where the acceptance limit is higher). However, that's not the focus of this Cleaning Memo. One could reply to my objection by stating that the five spiked levels were going to be 0.4, 0.8, 1.2, 1.6 and 2.0  $\mu\text{g}/\text{cm}^2$ , corresponding to 20%, 40%, 60%, 80%, and 100% of the acceptance limit. That makes sure all spiked levels are at or below the acceptance limit (where one expects the data in a protocol to be). However, is it reasonable to assume, with all the variability of a swab method, that the relationship between “percent recovery” and “amount spiked” should be linear?

Part of the reason I didn't like that approach was that I had seen data where people tried to show a relationship that was sloped (either positively, with increasing percent recovery with increasing spiked level, or else negatively, with decreasing percent recovery with increasing spike level). I believe that *if there is any slope*, it is a negative slope. However this would be in a study where the spiked levels would be *much greater* than the range typically used in cleaning validation studies involving about five spiked levels. In other words, if one were to spike at levels of 1, 5, 25, 125, and 250  $\mu\text{g}/\text{cm}^2$ , my expectation is that recoveries at the higher end would be significantly lower (hence the negative slope that I referred to). The reason for this difference would be loading of residue onto the swab and/or rate of removal of residue from the surface. However, in the levels typically utilized for cleaning validation studies (where the range of the different spiked levels is generally no more than 5:1), I would not expect any significant differences in recovery percentages as a function of the applied residue amount.

But, the results in a recent publication where swab recoveries were performed forced me to conclude that perhaps when the claim is made that percentage recoveries should be in a linear relationship with spiked amount, what is actually meant is that within a relatively narrow range the recovery percentages should be *all the same* (essentially meaning in a plot, the slope is zero). Does that suggest that I am wrong, and that one should perform sampling recoveries at five different levels to demonstrate linearity? I would argue just the opposite. Obtaining the same recovery percentages over a relatively narrow range of spiked levels is a good argument for saying one should only perform a recovery study at *one* spiked level (at the residue limit). In other words, if all recoveries are going to be same over that narrow range, then why do multiple spiked levels? Some may argue that it demonstrates *consistency* of the sampling process. That may be true, but couldn't the same consistency be demonstrated by performing *more replicates at one spiked level* as opposed to spiking at

five levels. For example, I would expect to have a clearer picture of recovery consistency by performing 15 replicates at one level as compared to performing three replicates at each of five spiked levels. I would probably go further and argue that only nine replicates at one spiked level might be a better measure of consistency of the swabbing procedure. But, it is important to remember that swabbing is a type of manual cleaning, and it has the inherent variability of a *manual* cleaning process.

For those of you still convinced that spiking five levels is required, here is what I would suggest. For several residues (such as actives or cleaning agents) that are sampled by the same swabbing process, perform your recovery studies at five levels for the first three or four times you do recovery studies. Then once you have demonstrated that the swabbing procedure is consistent over the spiked range, write a rationale for only spiking at one level (at the acceptance limit) for subsequent recovery studies using the same swabbing procedure. This is merely leveraging what we have learned from previous studies, a position consistent with recent FDA guidance.

This Cleaning Memo is designed to further explore issues in sampling recovery studies. As in any scientific work, an understanding of the sampling process is a key to designing an appropriate and defensible sampling recovery program.