

July 2007
Spiking Amounts for Sampling Recovery Studies

If a sampling method is utilized in a cleaning validation protocol, it is important that a sampling recovery study be performed. This need is covered in the FDA's cleaning validation guidance. One issue to consider is the amount of residue to be spiked onto the coupons (coupons of the same materials of construction as the production equipment are usually used for these laboratory studies).

Typically what is done is to spike an amount of material per surface area corresponding to the MAC (maximum allowable carryover) calculation for the surface limit (usually in $\mu\text{g}/\text{cm}^2$), a limit I typically call the L3 limit). In other words, if my calculated limit of the active were $1.7 \mu\text{g}/\text{cm}^2$, I would want to spike at that level. If the swabbed (or sampled) area were 25 cm^2 , I would spike $42.5 \mu\text{g}$ of active onto the coupon. Could I also perform spiking studies at other levels, such as levels above or below $1.7 \mu\text{g}/\text{cm}^2$? My rationale for not doing levels above the L3 value is that if we were to find those levels in my protocol execution, I would certainly fail the protocol. I might want to perform spiking studies at levels above the L3 value to cover future situations where the L3 value was higher; however, that is a business justification and not a scientific justification.

What about levels below my L3 value? Should I spike at levels below the L3, because certainly that is where I expect my residue values to be in protocol execution (if that is not the case, I had better redesign my cleaning process!). In other words, if my L3 value is $1.7 \mu\text{g}/\text{cm}^2$, in my protocol I am expecting (or at least would like to see) values of less than $0.8 \mu\text{g}/\text{cm}^2$ (so that I am not on the ragged edge of failure). Does it make sense to additionally spike at such lower levels? My answer is that it makes more sense than spiking at levels above the L3 value. However, it is not required, because recovery percentages associated with spiking at a higher level generally represent a worst case as compared with lower levels. What I mean by this is that recovery percentages generally decrease with increasing spiked load. So, if I determine the recovery at the L3 value, the percentage will be lower than the recovery at 0.5 (or half) of the L3 value. An illustration I sometimes use in my training seminars is that of a pile of snow and a snow shovel (although a pile of sand and a sand shovel also works). If I have a snow drift four feet high and am allowed one shovelful, the percentage of snow picked up in one try will be relatively low (<10%). On the other hand, if the snow is only four inches high, I am more likely to pick up a greater percentage (>50%) of the snow that is present in one shovelful. (I can certainly carry this to an extreme, and assume a situation where the snow is a monomolecular layer; in that case I will have essentially no recovery from one shovelful, because the shovel will pass right over the layer. However, I do not believe that most situations in cleaning validation represent this extreme case.)

It may also be the case that at very low spiked levels, I do get lower recoveries due to variability in the analytical method at those low levels. This also doesn't mean that I might not perform a study and find a recovery of 80% at the L3 spiked level and 75% at 0.5 (half) of the L3. In that case I would argue that the difference represents the inherent variability in the recovery method. Next time I run it, I might find the percentages reversed. However, it makes sense as a general principle, within the residue levels considered for cleaning validation, that recovery percentages will decrease as loadings (spiked amounts) increase. This is the basis for my preference in cleaning validation that if only one level is chosen for spiking (my preference), it be at the L3 level. If an additional level is chosen, it should be in the range of 20-25% of the L3 value. The rationale for this level is that, in general, these represent levels where I am expecting my data to be for

protocol execution.

Furthermore, I would never do more than two spiking levels. While some companies like to spike at five or six levels to demonstrate linearity in recoveries, this is not a requirement, nor is it an expectation. I assume the justification used for this practice is the linearity studies for analytical method validation. While it is reasonable to expect a certain linear response over a limited range for an analytical method, it is not reasonable or expected that I should get a linear response for recoveries at different spiked levels. The main reason is the variability of the sampling method itself. It adds so much variability that linearity is not expected. The analogy I sometimes use is manual cleaning. Most people agree that manual cleaning is highly variable. Well, swab sampling is just another case of manual cleaning. Yes, it is probably more controlled than manual cleaning of process equipment, but it still has the variability associated with manual cleaning. In my opinion, doing more than two levels for spiking studies is not a good use of resources (unless there is a justified business reason or scientific rationale, such as covering situations where limits for the same residue may be significantly different in two different protocols, and I want to cover all the possibilities in one study rather than several studies). Note that I am not advocating two levels; my preference is still to spike at only one level (the L3 level).

Whatever is done for spiking amounts should be well defined in a cleaning validation policy, plan or procedure, to assure consistency within a given program.