July 2010
Statistics for Visual Limits

This makes two months in a row that I have discussed statistics in the context of cleaning validation, because it is that important.

This Cleaning Memo is an evaluation of a recent *Pharmaceutical Technology* article entitled "Statistically Justifiable Visible Residue Limits", by M. Ovais (March 2010 issue, pages 58-71). The author asserts that "current methods for establishing visible residue limits are not statistically justifiable". The author presents an example of determining a "visual residue limit" by spiking studies. In a spiking study, coupons are spiked at different levels, and a panel of observers looks at each panel under defined viewing conditions to determine the nature of the residue. Without going into the detail, the author provides a "logistic-regression" model to determine the "probability of detection" of the residue at that selected level. Needless to say, what this results in is a higher limit (a worst-case) than what would be determined by a consensus of multiple observers.

There are several questions to ask. Is this statistically correct? And, is this statistical evaluation really necessary? I can't answer the first question; I'll leave that up to the statisticians. While one answer to the second question is "you can certainly do it because it results in a higher visual limit" (a higher visual limit, contrary to what is often thought, is actually a worst-case), I'll give my answer to the second question below.

However, to do that it is necessary to clarify a few things. One is that many publications (apparently including this *Pharm Tech* article) list the "visual limit" as the *lowest* spiked level at which observers can consistently see *any* residue on the spiked surface (of course this is under defined viewing conditions and for a defined residue and a defined surface, but that will be assumed throughout this discussion). In other words, if I spike a surface of 25 cm^2 at different levels, then the spiked level at which all observers see even a speck on the spiked surface is the visual limit. While that may be one definition of visual limit, it is not a useful definition for cleaning validation purposes.

Why do I say it is not useful? The main reason is that the purpose of a visual limit is to say any surface viewed (under the same viewing conditions) that is visually clean has residue below that defined visual limit. Unfortunately, doing spiking studies and determining the lowest spiked level which has any residue on the spiked surface can't be used in that way. Why? First, remember that the worst case for a visual limit is a high value, not a low value. Defining the visual limit in this way presents an artificially low visual limit, which will allow one to state that the residue is below the specified value without a sound scientific basis.

The issue here is that when I spike at a fixed level (let's say 1.0 µg/cm^2), and only see a small speck of residue in a corner of the area spiked, I cannot really say that any surface that is visually clean has a residue level of less than 1.0 µg/cm^2. If I spiked at 1.0 µg/cm^2, and the surface was evenly covered (an ideal situation), then it would be appropriate to say that the spiked surface truly represents 1.0 µg/cm^2, and therefore any surface which is visually clean has a residue level below 1.0 µg/cm^2.

What happens in the real world when I do spiking studies is that the residue is not evenly spread over the spiked surface. Instead, due to the drying effects (difference in drying between the edges of the spiked residue solution and the center of the spiked residue solution), I will typically see a "donut hole" effect, with differing

amounts of residues on different parts of the spiked area. Therefore, if I spike at 1.0 μg/cm^2, it is possible that some portions of the spiked area may have concentrations of 0.8 μg/cm^2, while other portions have residue levels of 1.2 μg/cm^2. Perhaps I can see the residue in the spiked areas where the surface concentration is 1.2 μg/cm^2, but not see it at a surface concentration of 0.8 or 1.0 μg/cm^2. In that case, I will say the visual limit is 1.0 μg/cm^2, which would be misleading.

The "correct" way (or at least one correct way) to determine the visual limit is slightly different. The same spiking coupons are prepared. However, the visual limit is then defined as the lowest concentration (in μg/cm^2) in which the entire spiked area is visually dirty or soiled (that is, the lowest level at which residue is seen across the entire spiked area). Defined in this way, there is then a scientific or rational justification for saying any surface observed that is visually clean has a residue below the spiked level. Note that in this case, there may be (or better stated, there will be) variations in amounts of residue on different parts of the spiked coupon. That is inevitable, because of the drying effect. However, this approach is one that should be used (and not the approach of defining the visual limit based on the lowest spiked level where any residue is seen).

Why am I explaining this to address a statistical question? First, there is a certain level of "safety" already built into the determination of the visual limit. When I spike at 1.0 μg/cm^2, and state that the visual limit is 1.0 μg/cm^2, the true visual limit is lower than that (due to the drying effects mentioned above). How much lower, I can't say for certain; however, I suspect that the "true" visual limit is probably 0.8 μg/cm^2 or below.

Secondly, defining a visual limit is not necessarily an exercise where I need to get that visual limit as low as possible. My preference in using "visually clean alone" is not to do a series of coupons spiked at different levels. I prefer to first calculate the residue limit (using traditional maximum allowable carryover calculations, for example) to determine the limit per surface area (for those of you who follow my writings, this is the L3 limit in μg/cm^2). If my calculated residue limit is 4.0 μg/cm^2, why do I need to establish a visual limit that may be as low as 1.0 μg/cm^2? In this situation, I prefer first do a spiking study at 4.0 μg/cm^2. If at that spiked level all observers were not able to see residue across the entire spiked area, then who cares what the visual limit is? I clearly cannot use visually clean alone in a protocol to establish that the residue is below the calculated limit. On the other hand, if I spike at 4.0 μg/cm^2 and can readily see residue across the entire spiked area (albeit uneven amounts on different portions of the spiked area), then I have a rationale for saying that surfaces observed that are visually clean are, in fact, below the calculated limit.

Note that this last situation (of spiking at 4.0 μg/cm^2) already has some (undefined) safety margin in that the "true" visual limit is somewhat lower (because of the uneven concentrations across the spiked surface). That said, my preference is to add an extra margin of safety. If the calculated residue limit is 4.0 μg/cm^2, my preference is to spike at an additional lower level, such as 3.0 or 3.5 μg/cm^2. If I can see residue across the entire spiked area at those lower levels, I have an additional margin of safety in my determination of a visual limit.

If what I have discussed is the proper way to implement determinations of visual limits for a use of "visually clean alone" (that is, without swab or rinse sampling), then it would appear that there are significant safety

margins built into the evaluation, and that a statistical evaluation of the "probability of detection" may be nice to have, but is not necessary.

This leads me to my last point. In that same issue of *Pharm Tech* was a short article by Lynn Torbeck (my favorite statistician, as revealed in last month's Cleaning Memo) entitled "The Role of Statistical Tests". In it, Mr. Torbeck points out that statistical significance tests should only be used after one first determines that there is a practical difference between two data sets. If there is no practical difference, don't perform the statistical tests.

While the published article on statistics for visual limits is not strictly on statistical significance, it does invoke statistical principles to determine whether future observers would get the same result (or better said, to set a limit such that there is a higher probability that future observers would also report the same visual limit). I would put forth that with the determination of visual limits properly done (that is, defining the visual limit as the lowest spiked level where all observers see residue across the entire spiked area) has sufficient safety margins (either inherent in the process or which can be added to the process) such that extensive statistical analysis adds little or no value.

Note that it is certainly possible to use the statistical approach to further make the visual limit higher (which is a worst case). However, I think a good understanding of what is involved in determining visual limits suggests that there are practical safeguards already built into the visual limit determination.