Acceptance Limits for APIs

This article reviews the history of Cleaning Validation Acceptance Limits for Active Pharmaceutical Ingredients (APIs) and identifies where the currently used industry limits came from.

Cleaning Validation for the 21st Century: Acceptance Limits for Active Pharmaceutical Ingredients (APIs): Part I

by Andy Walsh

Part I of this article reviews the history of Cleaning Validation Acceptance Limits for Active Pharmaceutical Ingredients (APIs) and identifies the origins of currently used industry limits. The current approaches to setting acceptance limits will be analyzed and some of the problems and weaknesses of these approaches will be discussed.

Early Ideas on Cleaning Validation Acceptance Limits

In the early 1980s, most companies were just beginning to grapple with the FDA’s shift to validation. This activity was of course more focused on Process Validation, but some companies took a wider view and began looking at cleaning processes as well. In 1984, Samuel Harder published an article, “The Validation of Cleaning Procedures,” and discussed many aspects of what he saw would be required to validate a cleaning process. Most, if not all, of the ideas he expressed can still be seen in practice in cleaning validation today. Concerning the setting of Acceptance Limits, Harder wrote that limits

“...must be practical and achievable by a reasonable cleaning procedure... ...must be verifiable by analytical methodology existing in the company... ...and must be safe and acceptable and in line with residual limits set for various substances in foods.”

(Note: The phrase “Practical, Achievable, and Verifiable” began to appear in many guidelines and literature subsequently.)

Harder goes on to reference 21 CFR 193 “Tolerances for Pesticides in Food Administered by the Environmental Protection Agency” and showed a table of limits for a variety of hazardous pesticides and herbicides. These limits ranged from a low of 0.5 ppm for Diquat to a high of 200 ppm for hydrogen cyanide. Harder pointed out that the amount of drug products ingested by an individual is much lower than the amount of food ingested. Therefore, he suggested that Acceptance Limits for drug substances comparable to those used for pesticides would be reasonable.

In 1988, the FDA had its first major experience with cross contamination traceable to inadequate cleaning and cleaning validation. A supplier of the API Cholestyramine Resin USP had to recall the product, due to contamination with low levels of intermediates and degradants from the production of agricultural pesticides. This cross-contamination was believed to have occurred from the use of drums in the manufacture of the API that had been used to recover solvents from the manufacture of agricultural pesticides at another location. The drums were not properly cleaned leading to agricultural pesticides entering the API manufacturing process. This incident raised the FDA’s awareness to the potential for cross contamination from unvalidated cleaning processes and heightened the concern in the industry that validation would become required for cleaning as well as for process.

In 1989, Doug Mendenhall of Abbott Labora-
torsies published a chapter “Cleaning Validation” in Drug Development and Industrial Pharmacy.³ Mendenhall expanded upon the ideas presented by Harder adding ideas, such as using a matrix approach, testing for cleaning agents, placebo batches, and most interestingly, pointed out the potential use of visual inspection. In addition, for Acceptance Limits, he suggested to “…establish in collaboration with tox and medical authorities an effect threshold” or alternatively, to “…superimpose an appropriate safety factor, e.g., 10X, or 100X.” Although not cited as such, this suggestion was probably connected to a much older article by Lehmann and Fitzugh on the use of a 100-fold safety factor.⁴ The 100-fold safety factor consists of two 10-fold factors to allow for human variability and interspecies differences. Mendenhall proposed that limits for surface residue levels be calculated based on a smallest batch size/maximum dose combination. Mendenhall goes on to state that based on these calculated surface limits, visual inspection should be adequate in many cases. Mendenhall writes:

“Alternatively, one could use a very pragmatic approach: visual cleanliness. While the latter may sound far too unsophisticated and non-quantitative, in our experience quantitative calculations have almost universally yielded tolerable levels of residues which were readily apparent visually, i.e., the visual cleanliness criteria was (sic) more rigid and clearly adequate. Clearly, however, for extremely potent or toxic substances, calculated tolerable residuals may be well below visual detectability. Hence, common sense and judicious use of quantitative methods when appropriate are called for depending on the nature of the possible residues involved.”

(It is interesting to note that the use of visual inspection is now being seriously investigated, as of this writing, more than 20 years later.)

These two industry articles laid the foundations from which most cleaning validation acceptance criteria were derived and are the origins of many cleaning validation activities practiced up until today (2011). Shortly after these publications, a major event began to unfold that shaped the direction of cleaning validation and many other industry practices. This event was...

**The U.S. vs. Barr Laboratories and the Wolin Decision**

From 1989 through 1992, the US FDA inspected several Barr Laboratories, Inc. facilities and became engaged in a struggle with the company over many issues concerning the FDA’s interpretation of the GMPs. This struggle resulted in the issuance of multiple FDA Form 483s with increasing numbers of observations. The FDA finally sued Barr Labs in June of 1992 with District Judge Alfred M. Wolin presiding over the trial.⁵ The trial ended in February of 1993 with the decision by Judge Wolin that resulted in a major loss for, and an injunction against, Barr Laboratories. The signifi-

<table>
<thead>
<tr>
<th>Dose-based Limits</th>
<th>“Default” Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/10ⁿ of a therapeutic dose</td>
<td>Not more than 1 ppm</td>
</tr>
<tr>
<td>1/50ⁿ of the maximum therapeutic dose</td>
<td>3 ppm (based on Arsenic Limit)</td>
</tr>
<tr>
<td>1/1,000ⁿ of the lowest dose</td>
<td>10 ppm</td>
</tr>
<tr>
<td>Less than the smallest therapeutic dose</td>
<td>22 ppm</td>
</tr>
<tr>
<td>Some Other Limits</td>
<td></td>
</tr>
<tr>
<td>Detection Limit of Method</td>
<td>&lt; 5 mg/swab</td>
</tr>
</tbody>
</table>

Table A. 1992 Pharmaceutical Manufacturers Association survey on acceptance limits.

cance of this trial to the pharmaceutical industry cannot be overstated. Up until this time, the industry had been in an ongoing debate with the FDA’s interpretation that compliance with the GMPs required process validation. Now, the FDA’s process validation requirement was upheld in a court of law. In addition, among several other things, cleaning validation was now required.

The action in the Barr Laboratories case was closely followed by the pharmaceutical industry, and one trade organization, Pharmaceutical Manufacturers Association (PMA now PhRMA), sent out a survey to all of its members asking many questions on what they were doing in regard to cleaning validation. In particular to this discussion, they asked how companies were setting acceptance limits. In the survey results, PMA listed 44 unique acceptance criteria from the responses. A representative sampling of these limits is shown in Table A.

As can be seen in Table A, the acceptance limits in use at that time were inconsistent from company to company, and in many cases, arbitrarily selected. There were two patterns that appeared; the use of limits based in some way on the therapeutic dose and limits set arbitrarily (or as some workers call them today, “default” limits). At least one company tied their default limit to the USP Arsenic limit reasoning that if Arsenic could be present at 3 ppm, then residues of their compounds at lower levels should be considered safe. Note that both types of limits are of the kinds suggested by Harder or Mendenhall.

Concurrent with the Barr Laboratories trial and the PMA Industry Survey, the foundations for acceptance limits were further expanded by another publication known in the industry as...

**The Fourman and Mullen Article**

At the time of the Barr Labs trial, another company, Eli Lilly, was also involved in a number of issues with the FDA over cleaning validation. In particular was the setting of acceptance limits. In 1993, Gary Fourman and Dr. Mike Mullen of Eli Lilly published an article where they proposed the use of a combination of limits.⁶ These authors suggested that “any carryover of product residue meet the following criteria:”

- No more than 0.001 dose of any product will appear in the maximum daily dose of another product.
- No more than 10 ppm of a product will appear in another product.

Continued on page 76.
• No quantity of residue will be visible on the equipment after cleaning procedures are performed.

In the article, the authors explain that these 0.001 and 10 ppm criteria were selected in the following manner.

For the 0.001 dose criterion, the three factors of 10 were based on: “The first is that pharmaceuticals are often considered to be non-active at 0.1 of their normal prescribed dosage; the second is a safety factor; and the third is that the cleaning validation program must be robust, i.e., be vigorous enough that it would be considered acceptable for quite some time in a world with ever tightening standards.”

For the 10 ppm criterion, the article states that it has its roots in the regulations that apply to food products: “In those regulations, certain levels of hazardous substances are considered acceptable in animal tissues and poultry products that enter the human food chain.” Unfortunately, no regulatory reference was cited for this criterion nor was it explained how the value of 10 was selected. However, the origin of this criterion is very likely related to the article by Sam Harder.

For the visible residue criterion, the authors point out that some products meet the first two criteria, but there could still be residue visible and “...it does not seem appropriate that residues could be visible on GMP equipment even though it is labeled clean.” It seems appropriate to note here that 21 CFR 211.67(b) always required cleaned equipment to be inspected for cleanliness immediately before use. Theoretically, such inspection should already be in place and would not allow equipment with visual residues to be used making this criterion redundant.

At the time, this was a landmark article in the world of cleaning validation as it was the first publication to lay out specific criteria for determining cleaning validation acceptance limits. Underscoring its importance, this article was cited in almost every subsequent article on cleaning validation for years afterward. Pharmaceutical companies now had something to point to in how they had derived acceptance limits for their cleaning validation program.

### After Effects of Barr Laboratories Decision

There was significant fallout from the Barr Laboratories case for the industry. Judge Wolin described the Barr case as a “confrontation between a humorless warden and its uncooperative prisoner” and a product of “an industry mired in uncertainty and conflict, guided by vague regulations which produce tugs-of-war of varying intensity.” During the course of the trial, Judge Wolin observed that the GMP regulations were vague and not very detailed — certainly not detailed enough to expect companies to easily understand what the FDA’s interpretations and expectations were. Judge Wolin criticized the GMPs for their lack of detail and clarity. The FDA responded to these criticisms in a number of important ways. One by initiating a rewrite of the GMPs, which added much more detail. In this rewrite, they called for validation of cleaning (Proposed §211.220(a)) and the dedication of manufacturing facilities from just Penicillin to potentially many more compounds, such as cytotoxic agents, steroids, etc. And two, they also responded in the MidAtlantic region by issuing in 1992 the following.

### Guide to Inspections: Validation of Cleaning Processes

As the case was proceeding against Barr Laboratories, a number of inspectors in the MidAtlantic region put together a guide clarifying what their expectations were for Cleaning Validation. This was a Guide for FDA MidAtlantic region inspectors on what to look for in a cleaning validation program, including what is and what is not acceptable. The effort was led by a well known inspector at the time, Henry (Hank) Avalone. This Guide was very detailed and specific and included many of the elements that are typically found today in company cleaning validation programs. One year later, the Guide developed by the MidAtlantic region inspectors was adopted by the national center for use by all FDA inspectors.

The Guide also contained content that can be traced back to the publications we have already discussed; the Harder, Mendenhall, and Fourman and Mullin articles. For the establishment of limits, the Guide states that:

“The firm’s rationale for the residue limits established should be logical based on the manufacturer’s knowledge of the materials involved and be practical, achievable, and verifiable. It is important to define the sensitivity of the analytical methods in order to set reasonable limits. Some limits that have been mentioned by industry representatives in the literature or in presentations include analytical detection levels, such as 10 PPM, biological activity levels such as 1/1000 of the normal therapeutic dose, and organoleptic levels such as no visible residue.”

While at the start of the section on establishment of limits, the Guide makes a point of stating that they do not intend to

---

**Table B. 2006 Parenteral Drug Association survey on acceptance limits.**

<table>
<thead>
<tr>
<th>Type of Limits Used (APIs)</th>
<th>Dose-based only</th>
<th>Dose-based/ default</th>
<th>Dose-based/ default/Visual</th>
<th>Default Value</th>
<th>Process Capability</th>
<th>Limits of Detection</th>
<th>Other Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>49%</td>
<td>22%</td>
<td>45%</td>
<td>4%</td>
<td>10%</td>
<td>6%</td>
<td>16%</td>
<td></td>
</tr>
</tbody>
</table>

**Range of Dose-based Limits Used (APIs)**

<table>
<thead>
<tr>
<th>1/1,000&lt;sup&gt;a&lt;/sup&gt;</th>
<th>1/100&lt;sup&gt;a&lt;/sup&gt;</th>
<th>1/2,000&lt;sup&gt;b&lt;/sup&gt;</th>
<th>1/10,000&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>85%</td>
<td>13%</td>
<td>2%</td>
<td>6%</td>
</tr>
</tbody>
</table>

**Range of Default Limits Used (APIs)**

<table>
<thead>
<tr>
<th>10 ppm</th>
<th>1 ppm</th>
<th>&lt; 15 ppm</th>
<th>&lt; 1 ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>89%</td>
<td>6%</td>
<td>6%</td>
<td>3%</td>
</tr>
</tbody>
</table>
set acceptance specifications, the fact that they mention limits, such as 10 ppm and 1/100 of the normal therapeutic dosage as examples has been understood by many in the industry as tacit approval of these limits. However, it is worth highlighting that in concluding the discussion of the establishment of limits, the FDA sets a very clear criterion for inspection:

“The objective of the inspection is to ensure that the basis for any limits is scientifically justifiable.” So despite the mention of these limits and regardless of any industry adoption of them, the FDA still clearly expects that companies will put some thought and analysis into the selection of their cleaning validation acceptance limits. Simple adoption of the three Fourman and Mullin criteria is not satisfactory without a scientific justification for using these limits.

As mentioned earlier, the Fourman and Mullin article turned into a landmark publication for cleaning validation and the impact of this article on the industry also can be understood by considering how many regulatory guidelines subsequently adopted the three criteria they presented. Examples of the use of the Fourman and Mullin approach can be found in:

- PIC/S 2001 Guideline (European)
- CEFIC/APIC* 2000 Guideline
- TPP 2000 CV Guidelines (Health Canada)
- ICH (Q7A)
- Pan American Health Organization
- WHO Guide to GMPs
  *European Chemical Industry Council/Active Pharmaceutical Ingredients Committee

Finally, in August of 1998, The Parenteral Drug Association (PDA) released their PDA Technical Report No. 29 “Points to Consider for Cleaning Validation,” which summarized the current thinking and outlined the elements considered fundamental to a compliant cleaning validation program, including the setting of acceptance limits, at that time. Concerning the selection of limits the Report states: “It is very important that cleaning limits not be selected arbitrarily, but rather, that there be a logical and scientific basis for the numerical limit selected.”

More recently, in 2006, PDA conducted a survey on acceptance limits. The results can be seen in Table B. The types of limits reported look nearly the same as the ones listed in Table A. Unlike the 1992 PMA survey, the percentages of companies using the same limit are shown; however, the percentages in each case do not add up to 100 (Type of Limit = 152%; Range of Dose-based Limits = 106% and Range of Default Limits = 104%) so it is impossible to clearly interpret these data. But what can be taken away is that the industry is not following the limits stated in the regulatory guidelines mentioned above, they are still inconsistent in the application of limits, and they are still apparently using arbitrarily set limits. One important change is that approximately 10% (true % not known) report using Process Capability to set their acceptance limits. This is an important development that will be discussed in Part I of this article.

Continued on page 78.
Analysis of Current Approaches to Setting Acceptance Limits

The author of this article started working in cleaning validation in 1994 just as all these events were unfolding. All of the previously mentioned articles were eagerly read and incorporated in the new cleaning program being created as part of the author’s responsibilities. However, after implementing these limits, a number of issues began to surface.

For one thing, when calculating limits using these approaches, it seemed that the limits were everywhere since they will vary based on the value of the lowest doses, the batch size, and equipment surface area, etc. In some cases, the calculated limits seemed to leave large amounts of residues and in other cases, the calculated limits could be below analytical capabilities and impossible to meet; this situation has already been reported in the literature.

Table C demonstrates this issue through a simulation of the calculation.

In Table C, two drug products with widely different minimum doses are shown; a high of 500 mg and a low of 1 μg. Limits are calculated based using high/low combinations for Batch Size, Maximum Daily Dose, and Total Equipment Surface Area. Every possible combination of these factors was calculated (basically using a 2^4 Full Factorial Design) and the data then sorted by their Maximum Allowable Carryover (MAC) values from lowest to highest. Although some of these combinations may not actually occur in every manufacturing operation, this simulation is meant to demonstrate how widely varying the limits can be. These data clearly show that acceptance limits can vary by many orders of magnitude for the same drug compound depending on the Batch Sizes and Maximum Daily Doses of the next drug product and the Total Surface Area of the equipment used. It should be evident that different drugs manufactured in the same equipment will have different limits; the same drug manufactured in different equipment will have a different limit; and different drugs manufactured in the different equipment will have greatly different limits. For example, a drug with a higher minimum dose manufactured in a Granulator/Wet Mill/Fluid Bed Dryer Train (larger surface area) may have much lower limits than a drug with a lower minimum dose manufactured in a Single Pot Processor (smaller surface area). Or if the tablets are sent to a Contract Packaging Organization, the limits for a Thermoformer, which has much smaller product contact areas by comparison, will be much higher. This is inconsistent; why does a Granulator/Wet Mill/Fluid Bed Dryer Train need to be “cleaner” (have much lower limits) than a Single Pot Processor or even lower for a Thermoformer line for the exact same product?

Another difficulty with this calculation (Batch Size/Maximum Daily Dose) is that it is primarily applicable to tablets and capsules where dose sizes are fixed and known. Some workers have pointed out the problem of employing

Calculation for a Swab Sample:

1. Lowest Dose (Product A) \times \text{Batch Size} = \text{Maximum Allowable Carryover (MAC)}

   (Product A is the product being cleaned; Product B is the subsequent Product)

2. MAC/Total Surface Area = Surface Residue μg/cm^2

3. Surface Residue/μg/cm^2 × Area Swabbed = Residue on Swab (μg)

4. Residue on Swab/μg/Dilution Volume (ml) = Residue level in swab sample (ppm)

The table below shows the MAC values, Surface Residues, and PPM values in an Analytical Sample for all combinations of Lowest Dose (0.001 mg and 500 mg), Batch Size (15 kg and 50 kg), Maximum Daily Dose (0.05 gm and 5 gm), and total Surface Areas (1,000 cm^2 and 100,000 cm^2). A swab area of 100 cm^2 and a swab recovery of 100% are assumed.

<table>
<thead>
<tr>
<th>Lowest Dose (mg)</th>
<th>Batch Size (kg)</th>
<th>Max Daily Dose (gm)</th>
<th>MAC (mg)</th>
<th>Total Surface Area (cm^2)</th>
<th>Surface Residue (μg/cm^2)</th>
<th>ppm in 20 mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>15</td>
<td>5</td>
<td>0.003</td>
<td>10000</td>
<td>0.00003</td>
<td>0.002</td>
</tr>
<tr>
<td>0.001</td>
<td>15</td>
<td>5</td>
<td>0.003</td>
<td>1000</td>
<td>0.003</td>
<td>0.02</td>
</tr>
<tr>
<td>0.001</td>
<td>1200</td>
<td>5</td>
<td>0.24</td>
<td>100000</td>
<td>0.0024</td>
<td>0.1</td>
</tr>
<tr>
<td>0.001</td>
<td>1200</td>
<td>5</td>
<td>0.24</td>
<td>1000</td>
<td>0.24</td>
<td>1</td>
</tr>
<tr>
<td>0.001</td>
<td>15</td>
<td>0.05</td>
<td>0.3</td>
<td>100000</td>
<td>0.003</td>
<td>0.02</td>
</tr>
<tr>
<td>0.001</td>
<td>15</td>
<td>0.05</td>
<td>0.3</td>
<td>1000</td>
<td>0.3</td>
<td>2</td>
</tr>
<tr>
<td>0.001</td>
<td>1200</td>
<td>0.05</td>
<td>24</td>
<td>100000</td>
<td>0.24</td>
<td>1</td>
</tr>
<tr>
<td>0.001</td>
<td>1200</td>
<td>0.05</td>
<td>24</td>
<td>1000</td>
<td>24</td>
<td>120</td>
</tr>
<tr>
<td>500</td>
<td>15</td>
<td>5</td>
<td>1,500</td>
<td>100000</td>
<td>15</td>
<td>75</td>
</tr>
<tr>
<td>500</td>
<td>15</td>
<td>5</td>
<td>1,500</td>
<td>1000</td>
<td>1,500</td>
<td>7,500</td>
</tr>
<tr>
<td>500</td>
<td>1200</td>
<td>5</td>
<td>120,000</td>
<td>100000</td>
<td>1,200</td>
<td>6,000</td>
</tr>
<tr>
<td>500</td>
<td>1200</td>
<td>5</td>
<td>120,000</td>
<td>1000</td>
<td>120,000</td>
<td>600,000</td>
</tr>
<tr>
<td>500</td>
<td>15</td>
<td>0.05</td>
<td>150,000</td>
<td>100000</td>
<td>1,500</td>
<td>7,500</td>
</tr>
<tr>
<td>500</td>
<td>15</td>
<td>0.05</td>
<td>150,000</td>
<td>1000</td>
<td>150,000</td>
<td>750,000</td>
</tr>
<tr>
<td>500</td>
<td>1200</td>
<td>0.05</td>
<td>12,000,000</td>
<td>100000</td>
<td>120,000</td>
<td>600,000</td>
</tr>
<tr>
<td>500</td>
<td>1200</td>
<td>0.05</td>
<td>12,000,000</td>
<td>1000</td>
<td>12,000,000</td>
<td>60,000,000</td>
</tr>
</tbody>
</table>

Table C. Calculation for a swab sample limit and simulated results.
Acceptance limits for APIs

For example, what are the Minimum Dose and Maximum Daily Dose of a topical such as a sun screen? In the case of a sunscreen, the Maximum Daily Dose may be based on the total body surface not covered by the swimwear. Conversely, what is the Minimum Dose of a sunscreen or any Topical product for that matter? The use of a Finger Tip Application (FTA) as a standard minimum dose has been proposed by Ovais, et. al. Since Topicals are typically manufactured as a percentage of the formulation calculating the limits as a fraction of the percentage is a simpler approach. For example, the allowable level of the API of a 1% Topical would be 0.001% in another Topical. This approach avoids this issue and greatly simplifies the calculations. Clinical and Development batches face similar issues with minimum dose. The minimum and maximum doses may not be known or established yet.

The author has had many discussions with coworkers involved in Cleaning Validation on how the calculations of these limits can be confusing and are time consuming. A great deal of time is spent on calculations and in some cases, the calculated limits can not always be met; either the residues after cleaning are higher than the calculated limit or the swab analysis is not capable of detecting the levels dictated by the calculations. In practice, it has been found necessary to rework and/or rethink the calculations, sometimes in very creative ways, to arrive at limits that could be passed. In so doing, these calculations raised some questions: for example, why should a Topical product have to meet limits as low as those for an Injectable product?

PDA’s Technical Report No. 29 in Section 8.5 “The Basis for Quantitative Limits” contains a table showing ranges of safety factors to be used based on the type of product - Table D.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Approach Typically Applicable To</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/10th to 1/100th of a normal daily dose</td>
<td>topical products</td>
</tr>
<tr>
<td>1/100th to 1/1,000th of a normal daily dose</td>
<td>oral products</td>
</tr>
<tr>
<td>1/1,000th to 1/10,000th of a normal daily dose</td>
<td>injections, ophthalmic products</td>
</tr>
<tr>
<td>1/10,000th to 1/100,000th of a normal daily dose</td>
<td>research, investigational products</td>
</tr>
</tbody>
</table>

Table D. PDA’s Technical Report No. 29 in Section 8.5 “The Basis for Quantitative Limits.”

There is no reference to this table anywhere in the text and no rationale was given for the choices. An example calculation for an oral product only shows 1/1,000th as the safety factor to use. Regardless, it is clear that the intent was to provide lower safety factors for products with apparently lower risks. For example, why should a topical ointment for treating athlete’s foot have the same limits for an injectable solution for treating cancer? While superficially these ranges seem to make sense when they are looked at in detail, this approach only raises more questions.

Let’s say we are manufacturing a Topical product that is a low risk so we choose 1/10th as our “safety factor” to calculate a MAC value. So, if we are manufacturing a 400 kg batch of

Continued on page 80.
this Topical product and we are following it with another 400 kg batch of another product, by definition we can leave 39.99 kg of product behind and pass this limit. This is clearly a substantial amount of material that should easily be visible so clearly not what we could consider GMP; notwithstanding the implied yield loss. So is this limit really of any use?

At the other end of the spectrum, let’s consider a Clinical Manufacturing scenario where we are manufacturing clinical batches of a Topical product and we choose to use 1/100,000th to calculate our MAC value. In this scenario, the calculated limits may very well not be passable. I am aware of several situations where equipment was used once and then relegated to a corner since the analytical method was not capable of detecting to the level calculated while a debate raged on what to do. Repeated washings, exhaustive rinsing, and even re-passivation could not release this equipment from its prison. The use of such “safety factors” has even driven a perceived need for disposable equipment. Simple hand scoops for weighing raw materials are now disposable since they cannot be demonstrated to be “clean enough.” Taking this scenario even further, our new Topical product would use the 1/100,000th “safety factor” when in Clinical Manufacture, requiring disposable equipment, but when we move our product to launch in Commercial Manufacture the 1/10th “safety factor” would now be applied? All these situations prompted this author to ask the question:

Does Adding Safety Factors Actually Increase Safety?

My answer to this question, at the present state of setting cleaning validation limits, is no. So why is adding safety factors a problem? Let’s look at the actual effect of adding safety factors. Figure 1 shows the relationships between a drug’s “safe” limits, the drug’s residue levels after cleaning, and the Limit of Detection (LOD) of the method. On the left, there is a clear separation between the residue data and the acceptance limit. This is shown by the green arrow labeled “Margin of Safety.” It should be axiomatic that the greater the separation between the residue data and the acceptance limit, the greater the “Margin of Safety” is and the lower the risk to the patient.

The result of the application of additional “safety factors” to acceptance limits can be seen in the graph on the right. Here the acceptance limit (red line) is now lower. This has significantly decreased the apparent “Margin of Safety,” but with no improvement in the residue data. In both cases, the patient is exposed to the same level of residue. If the original acceptance limit was already safe, the addition of more safety factors did not provide any additional safety to the patient; but it may certainly cause the manufacturer to fail the cleaning validation and may result in restrictive and unnecessarily costly measures in operations. In many cases, we are being overly restrictive with cleaning validation limits and causing many operational difficulties.

As shown in Figure 1, improvement in patient safety is not achieved through adding extra safety factors to acceptance limits. If patient safety is the goal, a more appropriate approach is to focus on reducing the residue data rather than adding safety factors to acceptance limits. In Figure 2, we see the same graph on the left as in Figure 1, but the graph on the right shows the effect of providing a better developed cleaning process. The “true” residue levels are much lower (shown below the Limits of Detection) and the “Margin of Safety” is much higher since the acceptance limit was not arbitrarily lowered. This clearly improves patient safety and is a superior situation. Note also that in this example, the LOD of the method also was improved to provide a greater measure of how large the “Margin of Safety” actually is.

A firm should always start by designing quality into its cleaning procedures. The focus should be on developing robust cleaning processes; not on simply adding safety factors. On top of the issues concerning the calculation of acceptance limits and the use of safety factors there are...

Other Limitations to Using Dose-Based Limits

In addition to the difficulties with the MAC calculations and the application of safety factors discussed above, there are other fundamental difficulties with the concept of calculating acceptance limits based simply on the compound’s...
therapeutic dose. Remembering that according to Harder the goal of setting Acceptance Limits is that they “...must be safe” and that FDA also expects limits to be safe and “...to be scientifically justified” we need to look critically at whether these simple fractions of the therapeutic dose actually can achieve these criteria. There are many instances where the simple application of a fraction of the therapeutic dose may not be considered safe.

**Teratogens**

One aspect of a drug’s safety profile is whether the drug has teratogenic effects, that is, it can cause birth defects. The degree of teratogenicity is dependent on the dose, but it is not at all associated with the therapeutic dose. Let’s look at a few examples. In Table E, there are three drugs with the same lowest therapeutic dose; Difenoxin HCL, Coumadin, and Diethylstilbestrol. Consequently, their acceptance limits based on $1/1,000^{th}$ of the lowest therapeutic dose are all the same (0.001 mg). Although, in the fourth column, the doses shown to cause teratogenic effects for these compounds are vastly different. Difenoxin HCL is not teratogenic even at doses more than 60 times higher than its normal dose. Coumadin is teratogenic at $4$ mg/kg, while Diethylstilbestrol is teratogenic at a dose as low as 0.006 mg/kg. Clearly, these compounds present greatly different risks to patients, yet their cleaning validation acceptance limits are the same. So once again products that have the same dosage will have the same acceptance limits although the products may have greatly different safety profiles. Below these three, the drug Ribavirin has a lowest therapeutic dose 600 times higher and its corresponding acceptance limit also 600 times higher than the other three compounds, including Difenoxin HCL which demonstrates no teratogenicity. Clearly, this important aspect of drug safety is not factored into the calculation of limits considered when simply using the therapeutic dose.

**Therapeutic Index**

All drugs are evaluated for their Therapeutic Index, that is, the ratio of their toxic dose to their effective dose ($LD_{50}/ED_{50}$). This ratio is the lethal dose of a drug for 50% of the population divided by the minimum effective dose for 50% of the population. This ratio shows the relative safety of the drug; the higher the value the safer the drug is. Figure 3 compares two drug compounds that have the same therapeutic dose (1.0 mg), but have very different Therapeutic Indices.

So for calculating limits, if the products have the same dosage they will have the same acceptance limits although the products may have greatly different safety profiles. Drugs with very low ratios are known as “Narrow Therapeutic Index.” For example, warfarin, a drug with a notoriously Narrow Therapeutic Index is metabolized by cytochrome P450 2C9. Up to 35% of the population suffers from a deficiency in this enzyme. Since members of this population do not metabolize Warfarin as quickly as the normal population, they can be harmed if their blood levels are too high. While this is normally seen as a clinician’s problem for dosing, what about the effect on this population if Warfarin residues are carried over into another product manufactured on the same equipment? What is a safe level for them? Is $1/1,000^{th}$ of a dose sufficiently safe?

<table>
<thead>
<tr>
<th>Drug Compound</th>
<th>Lowest Therapeutic Dose</th>
<th>$1/1,000^{th}$ of Therapeutic Dose</th>
<th>Lowest Reported Teratogenic Dose (mg/kg)</th>
<th>Limit (ADE) based on Teratogenicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difenoxin HCL</td>
<td>1 mg</td>
<td>0.001 mg</td>
<td>No teratogenic effects at up to 61X therapeutic dose</td>
<td>N/A*</td>
</tr>
<tr>
<td>Coumadin</td>
<td>1 mg</td>
<td>0.001 mg</td>
<td>4</td>
<td>1.9</td>
</tr>
<tr>
<td>Diethylstilbestrol</td>
<td>1 mg</td>
<td>0.001 mg</td>
<td>0.006</td>
<td>0.003</td>
</tr>
<tr>
<td>Ribavirin</td>
<td>600 mg</td>
<td>0.6 mg</td>
<td>0.12 – 0.14</td>
<td>0.06</td>
</tr>
</tbody>
</table>

*Acceptable Daily Exposure values supplied by Dr. Bruce Naumann of Merck, Inc.

*(ADE must be set using other safety data)

Table E. Comparison of $1/1,000^{th}$ limits to ADE limits for teratogens.
Using the 1/1,000th approach, a low risk drug such as Aspirin can have much lower limits than some drugs that are much riskier. This doesn’t seem science-based and certainly not logical at all. How then should a drug company go about determining what is safe?

As it turns out, all drug companies have ready answers at their disposal. The pharmaceutical industry is unique in that it has the most extensive safety data on their products than any other industry, and in particular, data in humans. Extensive pharmacological and toxicological testing is performed on all drugs prior to marketing. All drug companies have already established comprehensive safety profiles of their products. All it takes is consulting the individuals in their companies who developed this information and have knowledge of the drugs’ safety data. Indeed, Hank Avallone wrote as early as 1989 that “It is also recommended that a medically qualified responsible person, such as a physician, have input in the establishment of a limit.”

Ironically, drug companies have always had everything they need for establishing science-based limits.

Part II of this article will discuss how to establish science-based limits for APIs using data from clinical and toxicological studies, a risk-based approach to evaluating cleaning validation data, and guidance on setting statistical process control limits from that data.

**References**

2. FDA Guide To Inspections: Validation Of Cleaning Processes.
Acceptance Limits for APIs


Acknowledgements
The author wishes to thank Dr. Richard Berkof, Nick Haycocks, Robert Kowal, Dr. Bruce Naumann, Mohammed Ovais, and Joel Young for reviewing this article and for their insightful comments and helpful suggestions.

About the Author
Andy Walsh is an Industry Professor at Stevens Institute of Technology in their Pharmaceutical Manufacturing Program where he teaches courses on Validation and Lean Six Sigma. In 2009, Walsh founded the Stevens Pharmaceutical Research Center (SPRC), a research lab focusing Pharmaceutical Manufacturing topics, such as cleaning process development, Total Organic Carbon analysis and method development, visual inspection method development and automation of GMP systems. A current chair of an international task team to write a cleaning Guide for ISPE and ASTM, he was one of the contributors to the ISPE Risk-Based Manufacture of Pharmaceutical Products (Risk-MaPP) Baseline Guide. He has more than 20 years of diverse validation experience in pharmaceutical and biotech companies, including Johnson & Johnson, Schering-Plough, and Hoffmann-La Roche. Walsh has given numerous presentations over the past 15 years with IIR, Barnett, WorldPharm, IPA, IVT and ISPE. He can be contacted by telephone: +1-908-707-1765 or email: andywalsh@clean6sigma.com.

Clean6Sigma, LLC, 50 W. Cliff St., Somerville, New Jersey 08876, USA.