Validation of Alternative Microbiology Methods for Product Testing

Quantitative and Qualitative Assays

Scott Sutton

The validation of alternative microbiological testing is an opportunity for a manufacturer to decrease the amount of time required for laboratory results. To properly validate these alternatives, a practitioner must first identify what is being studied. The regulatory effect on established product and process specifications and levels must be completely evaluated, as changing the method of analysis may well change the apparent number in the sample.

Table I: Sample-size requirements for comparison between two methods of the sterility test.

Failure rate 1 (compendial sterility test)	Failure rate 2 (sterility test with alternate method)	Sample size
0.001	0.01	1179
0.001	0.02	492
0.001	0.04	223
0.001	0.06	144
0.001	0.1	83

Scott Sutton, PhD, is a pharmaceutical consultant, microbiology, at Vectech Pharmaceutical Consultants Inc., tel. 800.966.8832, fax 248.442.0060. scott.sutton@microbiol.org.

The validation of alternative microbiological methods for pharmaceutical testing is a very confusing topic, the more so as people use phrases such as "alternate sterility test" or "alternate microbial limits test." What we have to agree to from the outset is that in fact we are not looking for "alternate" compendial tests. What most people actually mean is that we are looking for an alternative (and usually more rapid) means to determine whether a sample would meet the requirements of the compendial test. In most cases, the major components of the compendial test will be respected, it is only one or two particular aspects of the test that will be changed to increase the reliability or speed of the test.

The sterility test by membrane filtration is an appropriate example. To simplify this discussion, the compendial test is assumed to be the standard against which others are to be compared. But what does it mean to have an alternate sterility test? Should 50 samples be taken rather than 20? Increasing the sample size will result in a more sensitive test but also a fundamental change in the compendial test. However, this is not supportive of the goal of finding a real-time test that will provide a high degree of confidence that the sample will pass the compendial test. One could change the nominal porosity of the filter, arguing that a smaller pore will retain "small bacteria" and so presumptively result in a more sensitive test. Few people are arguing for this modification. Finally, one could change the method used to determine whether the filter has retained viable cells.

The compendial test answers this question by requiring the contents of those samples to be filtered through a 0.45-μM filter and the filter to be placed in two different growth media (to increase the range of microorganisms potentially recovered), followed by an incubation period of 14 days. The recovery conditions (*e.g.,* specific media, temperature, and incubation period) provide the parameters that determine the ability to recognize the presence of viable cells. An alternate test method may determine the presence of viable cells by testing for the presence of DNA, or the presence of ATP, a specific antigen, or the presence of an integral cell membrane and esterase activity (*i.e.,* a vital dye). What is important to remember is that the specific question that must be answered is, Can we devise a better way to demonstrate the presence of viable cells on a filter than by growth to turbidity under specified conditions?

An overwhelming advantage to rephrasing the question in this manner is that it provides us with testable predictions. For

example, if we were to come across an antigen that was of particular use, then a serological test would need to demonstrate only that it recognized live (whole) cells rather than fragments and that it was able to "see" as many cells on the filter as does the compendial test. Similarly, a test using ATP bioluminescence, a vital stain, or polymerase chain reaction would only need to demonstrate that it was as sensitive to low numbers of cells as was the compendial test (in this case turbidity from a low inoculum on a membrane filter). Of course, it would be necessary for all these tests to demonstrate that the method was only measuring true events and not false positives from poor design.

Using this approach, three questions relate to this discussion:

- Is something there? (qualitative assay)
- How many are there? (quantitative assay)
- What are they? (identification assay)

These are the questions we wish to answer by an alternative means. It will serve us all well to remember that we can realize the benefits of a rapid answer while minimizing regulatory risk by clearly focusing our validation plans on the specific question of interest.

Microbiology is not chemistry

The other difficulty in the discussion of "alternative microbiological methods" is the unfortunate reliance on *USP* ⟨1225⟩ "Validation of Alternative Methods" to shape our thinking on the subject. This is a difficulty with which the PDA Task Force writing Technical Report #33 wrestled, as did the US Pharmacopeia (1). Both documents seek to use the familiar terms such as *linearity, ruggedness, robustness,* and *limit of detection* but immediately run into problems. The definition of these terms in the *USP* Guidance Chapter ⟨1225⟩ works well for analytical chemistry assays but not for microbiology. The solution that both the proposed *USP* chapter ⟨1223⟩ and PDA Technical Report #33 pursued was to redefine the terms to fit the microbiological data (2).

In hindsight, it has become apparent that this approach was flawed for two reasons. First, it created two sets of definitions for very specific terms, thereby creating even more confusion in the minds of microbiologists and management. Second, it implied that the current compendial methods needed to be measured by these terms to lay the groundwork to demonstrate the equivalency of the new method. As anyone who has tried to perform these studies can attest, the current state of microbiology (established in the 1800s) does not lend itself well to this analysis. Consequently, data are generated that are sometimes difficult to interpret. This creates a second set of problems for the microbiologists and their management who are naturally reluctant to put anything in front of regulators they fear might slow down product approval. Review and interpretation of unfamiliar and complex data takes time.

One size does not fit all

The desire to find a validation scheme that is applicable to all compendial tests and to all potential alternative methods is understandable but doomed to failure. Each technology offers its own challenges and each compendial assay provides its own peculiarities. Consider, as examples, the sterility test and the microbial limits test (bioburden). On the surface, both tests seem to present the practitioner with similar challenges. Method suitability must be demonstrated. In this, one must demonstrate the ability of a recovery scheme to recover any microorganisms that might be present in the background of a potentially antimicrobial product. However, once this laboratory assay is done, equivalence to the compendial assay must be demonstrated.

It is tempting to try to demonstrate this property of equivalence between the two methods by means of parallel testing. For example, testing the finished product by the compendial method and by the alternate method to show equivalent (or superior) results by the alternate seems a reasonable and direct approach to establishing that the new method provides results equivalent to those obtained by the compendial method. Although this might have utility in a situation where there are a significant number of counts expected, this approach falls apart where the normal response is no recovery. This is certainly the case for the sterility tests, in which the failure rate should not exceed 0.1%. Clinical statisticians have well-characterized methods for determining the sample sizes for studies in which the investigator wishes to distinguish between two treatments using a rare event as the measure of the difference (3). An example of this in practice is provided in Table I.

In other words, almost 1200 parallel sterility tests would be needed to show that a new method was even ten-fold worse than the compendial method in determining failures (3). This is clearly unworkable, and another approach would need to be taken to evaluate equivalence.

If the intent is to use the alternative method in an assay designed to find large numbers of cells, however, then the situation is very different. In this situation, depending on the application, parallel testing in a strong study design might well provide excellent data.

False negatives and false positives

The regulator's concern with changes in compendial methods is that the use of a new method, although conferring some advantage to the manufacturer, cannot result in a lessening of the sensitivity of the analytical method. The new test must find every problem that the compendial assay would have found. Again using the example of the sterility test, the new method should not allow a lot to pass that the traditional method would have failed. This is a reasonable concern and one supported by every responsible manufacturer.

Continuing to use the sterility test as an example, a manufacturer has an additional concern, which is the corollary to the question of false negatives. Although a manufacturer certainly does not want to improperly release lots that would have failed the compendial sterility test, no one can stay in business if a large number of lots are improperly designated as failing the test and therefore cannot be released to market. The alternate test must not yield a large number of false positives. This concern is particularly pressing for technologies that do not require microbial growth because they use other characteristics of the microbial organism to demonstrate viability.

As early as 1956, Bryce described the two critical limitations of the compendial sterility test. First, it can only recognize organisms able to grow under the conditions of the test. Second, the

sample size is so restricted that it provides only a gross estimate of the state of "sterility" of the product lot (4). As noted previously, the issue is not changing the number of samples used, but rather that several vendors have technologies on the market that eliminate the need for microbial growth. Eliminating the need for microbial growth to recognize viable cells could potentially lead to the recognition of contaminants that are viable but nonculturable (VBNC) (5). It is an open question whether these previously unknown (or unseen) microorganisms matter clinically in the slightest. It is undeniable, however, that the obligations inherent in the manufacture of sterile products are not relaxed by an improvement in test methods.

When is a specification not a specification?

The issue of VBNC microorganisms raises a host of other questions. For example, water limits, presterilized bulk bioburden limits, and even environmental monitoring alert and action levels are established in submissions based on current technology. It is well-established that many, if not most, of the microorganisms in the environment are beyond the ability of traditional methods (6). Clearly, changing the method will change the results.

This question will be particularly interesting for established products. FDA's Center for Drug Evaluation and Research is proposing the use of comparability protocols to address these changes with a change in methodology (7). The key aspect of this approach is the inclusion of FDA reviewers in the design and the evaluation of the protocols and studies to reduce confusion at the time of filing.

Another source of potential confusion arises from the temptation to over-apply a new method. From the manufacturer's perspective, the investment in a new method usually is a significant one (especially for a microbiology unit's budget!). Once acquired, it seems sensible to apply this new technology in all possible locations. From the regulator's side, the new technology offers some advantage (*e.g.,* sensitivity and speed) and so it seems reasonable to insist that this technology be used in all microbiological process control tests possible. On the other hand, just as no one would require an analytical chemistry lab to use high-performance liquid chromatography in all assays (sometimes a pH meter is the proper tool), different microbiological assays will have different strengths and weaknesses in examining various aspects of a microbe. Therefore, given that diverse methods may be desirable in specific settings and that different methods will yield different results, it is entirely possible that the alternate method best suited to a process control will not be one used for finished-product testing. In fact, it is entirely possible that various manufacturers will propose a variety of methods for the same application.

Environmental monitoring levels will be especially affected by these changes because various manufacturers may propose different levels for identically classified rooms, on the basis of a range of methodologies.

A proposal

I propose that the guidance in PDA Technical Report #33 and in the proposed *USP* chapter ⟨1223⟩ be used only as a general guide. It can be very useful when designing operational qualification studies for a new method, but the critical consideration in a new method is in comparison with the established compendial method and neither document addresses this concern adequately. A more fruitful approach would follow the general guidelines below:

- Clearly define the aspect of the compendial test that is under investigation.
- Determine relevant measures to establish the equivalence of the new method to that used in the compendial test for that particular product. These measures will usually yield to statistical analysis. Both the USP and the PDA have begun steps in this direction (1, 2).
- Demonstrate the equivalence of the new method to the established method in the absence of the product sample. This will be specific to the particular portion of the compendial test method to be replaced by the alternate. For example, if you wish to use the alternate method to test for viable cells in the Antimicrobial Efficacy Test, you would evaluate the method against standard plate-count methods.
- Demonstrate the equivalence of the new method to the established method in the presence of the sample.
- Demonstrate the ability of your organization to perform adequate technology transfer (*e.g.,* transfer the method to the laboratory where the testing will be occurring).

Conclusion

The current discussion about alternative methods in microbiology is plagued with terms such as "alternate sterility test" or "alternate antimicrobial efficacy" test, when in fact few if any of the currently proposed methods are anything of the sort. A more useful characterization would be to recognize that there are now some alternatives available to the methodology of growth in liquid media or growth in solid media that has been the mainstay and the bane of pharmaceutical microbiology in the manufacturing arena. These new methods offer many advantages and must be evaluated individually for their application to a specific concern or assay. The regulatory effect on established product and process specifications and levels must be completely evaluated because the apparent number of microorganisms in a sample may be affected by the method of analysis.

References

1. Parenteral Drug Association, *Technical Report #33: Evaluation, Validation, and Implementation of New Microbiological Testing Methods* (PDA, Bethesda, MD, 2000).
2. "*USP* ⟨1223⟩: Validation of Alternative Microbiological Methods," *Pharm. Forum* **29** (1), 256–264 (Jan.–Feb. 2003).
3. Gordon and Watson, "A Note on Sample Size Determination for Comparison of Small Probabilities," *Controlled Clinical Trials* **15,** 77–79 (1994)
4. D.M. Bryce, "Tests for the Sterility of Pharmaceutical Preparations," *J. Pharm. Pharmacol.* 8, 561(1956).
5. M. Rappe *et al.,* "The Uncultured Microbial Majority," *Ann. Rev. Microbiol.* **57,** 369–394 (2003).
6. Falkowski and Vargas, "Genomics and Evolution: Shotgun Sequencing in the Sea: A Blast from the Past?" *Science* **304,** 58–60 (2004).
7. Food and Drug Administration, *Guidance for Industry: Comparability Protocols—Protein Drug Products and Biological Products; Chemistry, Manufacturing, and Controls Information* (FDA, Rockville, MD, 2003). **PT**