dx.doi.org/10.14227/DT140307P8



# Proposed Change to Acceptance Criteria for Dissolution Performance Verification Testing

e-mail: wh@usp.org.

**Walter W. Hauck[1], Ronald G. Manning, Todd L. Cecil, William Brown, Roger L. Williams**

*US Pharmacopeia, 12601 Twinbrook Parkway, Rockville, MD*

## Abstract

*USP Dissolution* <711> specifies performance verification testing (PVT) of dissolution Apparatus 1 and 2. Acceptance criteria are determined from a collaborative study and apply per tablet; i.e., each of the six tablets tested must fall within the specified acceptance criteria in order to pass. In this *Stimuli* article, USP proposes changing the form of the acceptance criteria to one that is consistent with the International Organization for Standardization's (ISO's) recommendations for proficiency testing. The new criteria would apply to the laboratory's average and standard deviation of the tablets tested. The article explains the rationale and shows the criteria that would be applied to USP Lot P Prednisone Reference Standard (RS) Tablets and USP Lot Q Salicylic Acid RS Tablets.

## Introduction

In either the private regulatory specification or the public monograph, the USP Performance test can be satisfied with approaches described in General Chapters *Disintegration* <701> or *Dissolution* <711>. As described in <711>, dissolution testing of non-solution oral drug products is complex, involving a preparatory apparatus, an analyst, and an analytical procedure. In 1979 at the request of industry and FDA, USP introduced reference standard (RS) tablets, formerly termed calibrators tablets, for use in periodic performance verification testing (PVT), formerly termed an apparatus suitability test. This is typically conducted every six months (*1*).

The acceptance criteria for the PVT are established on the basis of data from a collaborative study (Reference *2,* for example) conducted for each new lot of reference standards tablets and are provided to the laboratory on the information sheet accompanying the tablets. The acceptance criterion is set per tablet; i.e., the criterion is an interval and all six tested tablets must fall within that interval to be considered passing. Historically, the acceptance interval was determined as $\overline{X} \pm 2SD_R$, where $\overline{X}$ is the average (assigned value) and $SD_R$ the reproducibility standard deviation for a single determination from the collaborative study.

Recently, this formula has been modified in two ways. First, the statistical analysis is now done in the natural log scale to better satisfy the assumption of normality. Second, the factor 2 corresponds, approximately, to 95% coverage and up to a 5% false error rate *per tablet*. Because six tablets are tested, the actual false positive rate is higher than the nominal 5%. To correct for this multiple testing, a 1% value is now used, so the current formula is $\exp(\overline{X} \pm 2.576SD_R)$, where $\overline{X}$ and $SD_R$ are determined in the natural log scale.

Performance verification testing has the character of proficiency testing as described in guides published by the International Organization for Standardization (ISO). ISO Guide 43-1, *Proficiency Testing by Interlaboratory Comparisons—Development and Operation of Proficiency Testing Schemes* (*3*), describes proficiency testing as the use of interlaboratory comparisons to "determine the performance of individual laboratories for specific tests or measurements and to monitor laboratories' continuing performance." The USP PVT operates at the interlaboratory level because the acceptance limits are set from a collaborative study. Each laboratory conducting a PVT is thus comparing itself to the laboratories in the collaborative study. USP PVT differs from a typical proficiency test as described by Guide 43.1 because the PVT is conducted in a single laboratory for comparison to an assigned value from a collaborative study. ISO Guide 43-1 continues, "Participation in proficiency testing schemes provides laboratories with an objective means of assessing and demonstrating the reliability of the data they are producing," and "One of the main uses of proficiency testing schemes is to assess laboratories' ability to perform tests competently … It thus supplements laboratories' own internal quality control procedures by providing an additional external measure of their testing capability." This describes well the intent of USP PVTs. Although proficiency testing to an external sample is not usually performed in pharmaceutical QC laboratories, it is more common in official medicine control laboratories (e.g., check sample testing) and in other sectors (e.g., clinical chemistry laboratories). Its value in ensuring the integrity of the dissolution procedure is important for at least two reasons: first, well-known sensitivity of the dissolution procedure to various experimental variables; and second, the importance of the dissolution procedure itself in ensuring product performance over time. USP notes that

[1]*Corresponding author.*

the general ISO objectives of interlaboratory comparisons are not satisfied by a manufacturer-specific tablet, which is sometimes proposed as a solution to the wide acceptance criteria arising from USP's collaborative studies. Individual laboratories of course may set their own acceptance criteria for USP's reference standard tablets if they believe that the acceptance criteria set by USP are too wide.

Other ISO guides address acceptance criteria. Table 3 of ISO Guide 5725-6 titled *Accuracy* (*Trueness and Precision) of Measurement Methods and Results—Use in Practice of Accuracy Values* (*4*) lists the difference of the laboratory's mean versus the consensus value from a collaborative study and the laboratory's reliability standard deviation as important parameters for checking trueness and precision. Both ISO Guides 5725-6 and 21748, the latter titled *Guidance for the Use of Repeatability, Reproducibility, and Trueness Estimation in Measurement Uncertainty Estimation* (*5*), give suggested acceptance criteria. Those from Guide 21748 are:

- A laboratory's mean should be within two standard deviations of the assigned value, where the standard deviation is based on the reproducibility standard deviation of the mean; and
- Compare the laboratory's reliability variability to the value from the collaborative study using an *F* test with 95% confidence.

For the variance comparison, Guide 21748 sets a minimum of 15 degrees of freedom for the laboratory. The Guide also requires that the number of determinations (tablets) be sufficiently large that the repeatability standard error of the laboratory's mean is not more than 0.2 times the reproducibility standard deviation for a single measurement. The criteria from Guide 5725-6 differ from those of Guide 21748 only in using a chi-square rather than an *F* test to assess the within-laboratory variability.

## Preliminary Investigation

ISO 5725-6 leaves some unanswered questions regarding the USP PVT as specified in <711>:

- The criterion for the mean is an approximate tolerance interval, but the criterion for the variance is a statistical hypothesis test. Should these be more consistent in structure? A priori, one expects that the difference is not large. The approximate tolerance interval for acceptance criteria corresponds to the test if the variability in the collaborative study is disregarded. Thus, for example, the chi-square test of Guide 5725-6 is a tolerance interval, in contrast to the *F*-test of Guide 21748.
- The variance test of Guide 21748 does not specify whether the test should be one- or two-sided. Which should it be? The criterion in Guide 5725-6 is one-sided.
- Do we need full reliability standard deviations for the PVT, or can we use the between-tablet variability?
- If we compare between-tablet variability, the ISO requirement for degrees of freedom corresponds to three sets of six tablets (18 total) instead of the single set currently. Does the PVT need to be that much larger?
- Because there are two acceptance criteria (mean and standard deviation), should we adjust for multiple testing?

To address these questions and gain some understanding of the likely interactions of the criteria, we determined variations on the ISO criteria for USP Lot P Prednisone RS Tablets and USP Lot Q Salicylic Acid RS Tablets.

The first variation was to consider a tolerance interval for the within-laboratory variability instead of the *F*-test. The ISO criterion for the laboratory average is an approximate 95% tolerance interval for a laboratory average based on results from the collaborative study. It is approximate in using a factor of 2 rather than an exact tolerance interval factor, although the difference is not large with the degrees of freedom for the within-laboratory error from the collaborative study. The corresponding approach for the standard deviation would be a tolerance interval based on a chi-square distribution, as in Guide 5725-6.

The second variation was to consider the multiple testing issue. There are two acceptance criteria to meet. If each is at 5%, then the maximum false positive rate would be approximately 10%. To control the maximum false fail rate at 5%, each of the two tests would be performed at 2.5%.

For acceptance limits for the standard deviation, low variability is desirable, so only one-sided limits are considered; i.e., failure is considered only as too large a standard deviation.

ISO specifies reliability as the particular variability to compare in proficiency testing. USP's experience from its collaborative studies (see reference *2,* for example) is that the predominant component of reliability is the tablet-to-tablet variability in results. The tablet-to-tablet variability includes variability from the assay, position within the apparatus, location of the tablet within the vessel, the vessels, and any variability in the tablets themselves. Reliability would add additional variability associated with multiple runs conducted in a short period of time by the same analyst using the same apparatus.

For the number of tablets, the ISO criteria are determined based on one, two, and three sets of six tablets. With more than one set, the intralaboratory variability is pooled across sets.

***Table 1. Acceptance Limits, USP Lot P Prednisone RS Tablets, Apparatus 1.***

| | | | Limits for Xbar | | %CV Upper Limits | |
|---|---|---|---|---|---|---|
| Approach | Number of Tablets | Limits for Tablets | 5% Error Rate | 2.5% Error Rate | 5% Error Rate | 2.5% Error Rate |
| Current *USP* ISO | 6 | 47–82 | | | | |
| | 6 | | 53.5–72.3 | 52.6–73.6 | 12.3% | 13.2% |
| | 12 | | 53.9–71.8 | 53.0–73.0 | 11.2% | 11.9% |
| | 18 | | 54.1–71.6 | 53.2–72.8 | 10.7% | 11.2% |
| Modified ISO | 6 | | | | 12.1% | 13.1% |
| | 12 | | | | 11.0% | 11.7% |
| | 18 | | | | 10.5% | 11.0% |

Collaborative study results were a mean of 62.2 and an intralaboratory (residual) CV of 8.1%.

Results for the various options and sample sizes are shown in *Tables 1* and *2* for Lot P Prednisone RS Tablets and in Tables 3 and 4 for Lot Q Salicylic Acid RS Tablets. The limits shown as "Current USP" are the limits approved by the USP Reference Standard and Biopharmaceutics Expert Committees for these lots. *Figures 1* to *4* show the Prednisone limits with the data from the collaborative study.

There is little difference between the ISO (*F* test) and modified ISO (tolerance interval) limits for the percent coefficient of variation (%CV). Increasing the number of tablets from 6 to12 does noticeably change the limits for the CV, and the further change to 18 tablets has little effect. Changing the number of tablets also changes the power of the statistical test. This is the probability that a lab whose variability is greater than that of the collaborative study will fail (i.e., obtain a %CV in the PVT outside the limits). Again, the largest change is observed in power moving from six to 12 tablets. Although the further increase to 18 tablets does increase the power further, it has much less an effect than increasing from 6 to 12.

Changing the number of tablets has little effect for the mean, but that partly depends on an assumption about how the PVT would be conducted. The reproducibility standard deviation includes a component for experiment, corresponding to the intermediate precision components of analyst and equipment. The limits for the mean in Tables 1–4 assume that all two or three sets would be done by the same analyst on the same equipment, so the intermediate precision components are not reduced with the additional testing.

## Proposal

To the extent possible, USP is interested in being consistent with practices set in ISO guides. From this

***Table 2. Acceptance Limits, USP Lot P Prednisone RS Tablets, Apparatus 2.***

| | | | Limits for Xbar | | %CV Upper Limits | |
|---|---|---|---|---|---|---|
| Approach | Number of Tablets | Limits for Tablets | 5% Error Rate | 2.5% Error Rate | 5% Error Rate | 2.5% Error Rate |
| Current *USP* ISO | 6 | 37–70 | | | | |
| | 6 | | 42.3–61.8 | 41.4–63.3 | 12.7% | 13.7% |
| | 12 | | 42.6–61.4 | 41.7–62.8 | 11.6% | 12.3% |
| | 18 | | 42.7–61.3 | 41.8–62.6 | 11.1% | 11.7% |
| Modified ISO | 6 | | | | 12.6% | 13.6% |
| | 12 | | | | 11.5% | 12.1% |
| | 18 | | | | 10.9% | 11.5% |

Collaborative study results were a mean of 51.2 and an intralaboratory (residual) CV of 8.5%.

***Table 3. Acceptance limits, USP Lot Q Salicylic Acid RS Tablets, Apparatus 1.***

| | | | Limits for Xbar | | %CV Upper Limits | |
|---|---|---|---|---|---|---|
| Approach | Number of Tablets | Limits for Tablets | 5% Error Rate | 2.5% Error Rate | 5% Error Rate | 2.5% Error Rate |
| Current *USP* ISO | 6 | 23–30 | | | | |
| | 6 | | 24.2–28.9 | 24.0–29.2 | 3.1% | 3.4% |
| | 12 | | 24.2–28.9 | 24.0–29.2 | 2.8% | 3.0% |
| | 18 | | 24.2–28.9 | 24.0–29.2 | 2.7% | 2.9% |
| Modified ISO | 6 | | | | 3.1% | 3.3% |
| | 12 | | | | 2.8% | 3.0% |
| | 17 | | | | 2.7% | 2.8% |

Collaborative study results were a mean of 26.5 and an intralaboratory (residual) CV of 2.1%.

***Table 4. Acceptance limits, USP Lot Q Salicylic Acid RS Tablets, Apparatus 2.***

| | | | Limits for Xbar | | %CV Upper Limits | |
|---|---|---|---|---|---|---|
| Approach | Number of Tablets | Limits for Tablets | 5% Error Rate | 2.5% Error Rate | 5% Error Rate | 2.5% Error Rate |
| Current *USP* ISO | 6 | 17–25 | | | | |
| | 6 | | 18.4–23.0 | 18.1–23.3 | 7.8% | 8.5% |
| | 12 | | 18.5–22.9 | 18.2–23.2 | 7.2% | 7.6% |
| | 18 | | 18.5–22.9 | 18.2–23.2 | 6.8% | 7.2% |
| Modified ISO | 6 | | | | 7.8% | 8.4% |
| | 12 | | | | 7.1% | 7.5% |
| | 18 | | | | 6.7% | 7.1% |

Collaborative study results were a mean of 20.6 and an intralaboratory (residual) CV of 5.2%.

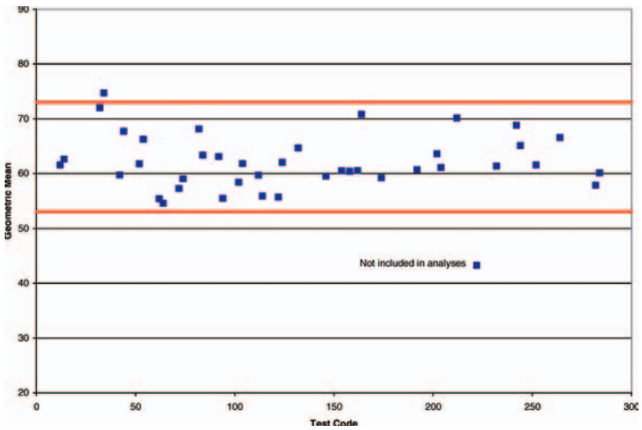


*Figure 1. USP Lot P Prednisone RS Tablets, Laboratory Means from the Collaborative Study, Apparatus 1.*

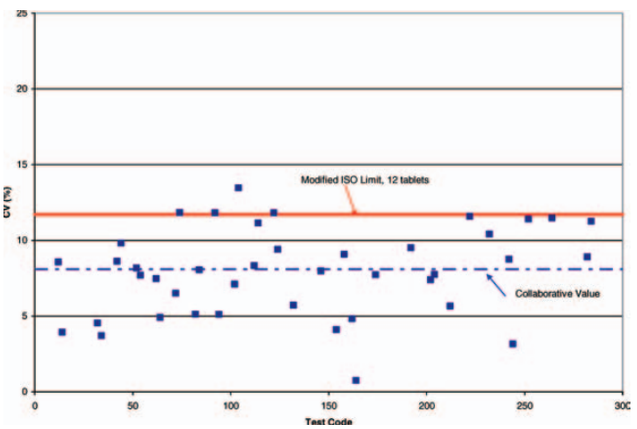


*Figure 2. USP Lot P Prednisone RS Tablets, Laboratory %CVs from the Collaborative Study, Apparatus 1.*

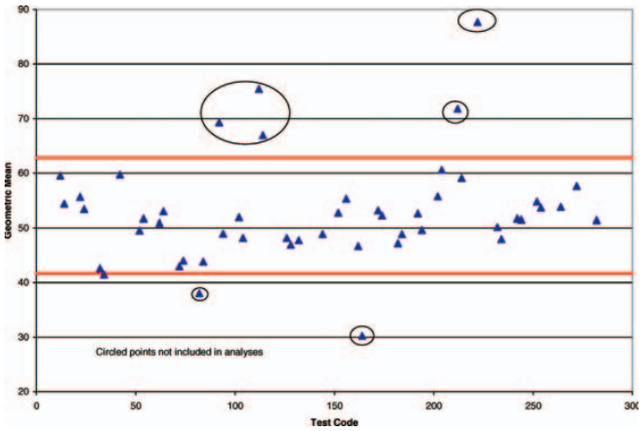


*Figure 3. USP Lot P Prednisone RS Tablets, Laboratory Means from the Collaborative Study, Apparatus 2.*

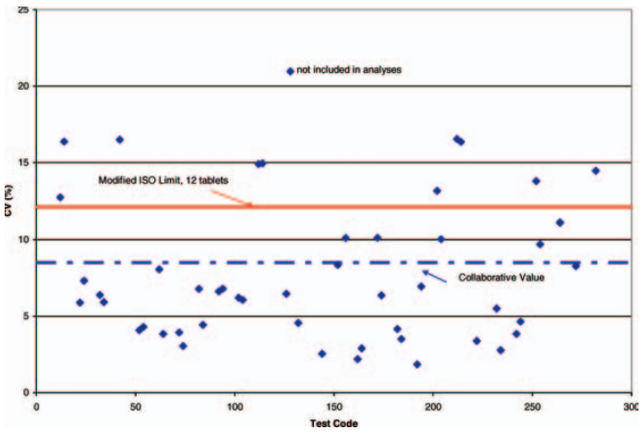


*Figure 4. USP Lot P Prednisone RS Tablets, Laboratory %CVs from the Collaborative Study, Apparatus 2.*

perspective, USP and its Biopharmaceutics Expert Committee envision changing the PVT acceptance criteria to align them with recommendations in cited ISO guides. Specifically, we propose to use the general form of the criteria of Guide 21748, modified to compare between-tablet variability based on 12 tablets.

Another consideration is that when six tablets are tested and fail, often that failure is one tablet of the six falling just outside the acceptance range although the other five pass. Criteria based on the mean and standard deviation should eliminate many instances of this type of failure.

ISO is unambiguous about the need for a sufficient sample size—three sets of six tablets—for the variance comparison. Preliminary analyses in this report suggest little added benefit of 18 compared to 12 tablets. Twelve tablets would be a doubling of effort for dissolution PVT if nothing else changes. One solution to this added burden might be to require for compliance in execution of the PVT use of only one USP tablet (e.g., Salicylic Acid or Prednisone) per apparatus. The USP Biopharmaceutics Expert Committee is exploring this option via the Dissolution Advisory Panel.

## Conclusion

USP believes that metrological understanding of ISO guides can improve manufacturers' approaches to ensure integrity of the dissolution procedure, including use of a PVT and publicly available USP RS tablets that support interlaboratory comparisons. As this article suggests, metrological approaches to measuring variation may call for increasing the number of tablets tested but may reduce the number of cases in which a test "fails" when only one of six tablets falls outside specifications. Further work at USP is being done to determine the proper sample size for sound metrological examination of testing tablets and acceptance criteria. In addition, USP will conduct further collaborative testing of its Lot P Prednisone RS Tablets using increased stringency in execution to determine if the high interlaboratory variance can be reduced.

USP welcomes comments on the approaches discussed in this *Stimuli* article. They should be sent to Walter W. Hauck, PhD, at wh@usp.org no later than August 15, 2007.

## References

1. USP Council of Experts Biopharmaceutics Expert Committee, Dissolution Advisory Panel, and USP Staff. The USP performance verification test, part I: proficiency testing and the USP dissolution procedure. Submitted for publication. 2006.
2. Glasgow M, Dressman S, Brown W, et al. The USP performance verification test, part II: collaborative study of USP's Lot P Prednisone Tablets. *J Pharm Sci*. In press.
3. ISO. *Guide 43-1, Proficiency Testing by Interlaboratory Comparisons—Development and Operation of Proficiency Testing Schemes*. 2nd ed. Geneva, Switzerland: ISO; 1997.
4. ISO. *Guide 5725-6, Accuracy (Trueness and Precision) of Measurement Methods and Results—Use in Practice of Accuracy Values*. Geneva, Switzerland: ISO; 1994.
5. ISO. *Technical Specification 21748, Guidance for the Use of Repeatability, Reproducibility, and Trueness Estimates in Measurement Uncertainty Estimation*. Geneva, Switzerland: ISO; 2004.