AAPS PharmSciTech, Vol. 10, No. 1, March 2009 (© 2009)
DOI: 10.1208/s12249-008-9169-y

Research Article

Change in Criteria for USP Dissolution Performance Verification Tests

Walter W. Hauck,[1,2] Anthony J. DeStefano,[1] William E. Brown,[1] Erika S. Stippler,[1] Darrell R. Abernethy,[1] William F. Koch,[1] and Roger L. Williams[1]

Received 18 September 2008; accepted 3 November 2008; published online 9 January 2009

Abstract. The US Pharmacopeial Convention has been evaluating its performance verification tests (PVT) for several years. These tests help ensure the integrity of the *US Pharmacopeia* performance test when a dissolution procedure, as described in General Chapter *Dissolution* <711>, is relied upon to test a nonsolution orally administered dosage form. One result of the evaluation is a change in the PVT criterion from one based on individual tablet results to one based on the mean and variability of a set of tablets. This paper describes the new PVT and its criterion and how its acceptance limits are derived from results of a collaborative study, explains a two-stage option for the test, and presents operating characteristics.

KEY WORDS: acceptance criteria; data analysis; dissolution test; performance verification test; reference standards.

INTRODUCTION

The *United States Pharmacopeia* (*USP*) in General Chapter *Dissolution* <711> includes performance verification tests (PVTs) for dissolution Apparatus 1 and 2 (1). As currently conducted, each of Apparatus 1 and 2 dissolution assemblies is tested periodically with one set of Prednisone Reference Standard (RS) Tablets and one set of Salicylic Acid RS Tablets. Accompanying purchase of a set of PVT RS tablets is a data sheet with acceptance criteria and other information to support analysts conducting the periodic test. In addition, the United States Pharmacopeial Convention (USPC) supplies a web-based toolkit that explains details of the dissolution procedure and adds information about both mechanical calibration and the PVT (2). The acceptance limits for a USP PVT are determined from results of a USPC-conducted collaborative study of performance of the tablets. For the assembly to pass, all tablets' results now must fall within the acceptance limits for the apparatus.

Over the past several years, USPC staff, working closely with the Biopharmaceutics Expert Committee of the Council of Experts, has conducted extensive evaluations of USP's PVTs (3,4). These evaluations have clarified the purpose and value of the PVT and have led specifically to two changes. First, based on a decision of the Expert Committee, Salicylic Acid Tablets RS will be discontinued as a PVT tablet on the grounds that the Prednisone Tablet RS is sufficient to the purpose of a PVT. This change is planned for Supplement 2 to *USP 32* and thus will be official December 1, 2009 (5). The second change is to alter the criterion for a dissolution PVT from a per-tablet approach to one based on the mean and coefficient of variation of a set of tablet results (6,7) (*Criterion* is defined here as the algorithm for analyzing PVT results together with the acceptance limits to which results are compared.). There are two primary reasons for this change. First, it brings *USP* into alignment with International Organization for Standardization (ISO) International Standard 5725 (8). Second, it resolves a scientific concern regarding the current criterion. Passing results might occur at the two extremes of the acceptance limits (for example, two values at the lowest value of the acceptance limits and the remainder at the highest) even though such results suggest a problem with the assembly. The revision to the PVT data analysis will allow analysts the option of conducting a PVT in two stages. That is, after one set of tablets, if results meet appropriate criteria the PVT can be stopped, and the assembly can be said to meet the acceptance criteria. Otherwise, the PVT continues with an additional set of tablets. This paper describes the statistical rationale and determination of the acceptance criteria for the two-stage option.

USPC intends that the advance to a mean/variance approach will be applied to all *USP* PVTs, where applicable. However, for simplicity of presentation, this paper will restrict its attention to dissolution Apparatus 2 (paddle). Marketed Apparatus 2 assemblies vary in the number of positions from 6 to 12. This paper will further restrict attention to assemblies with six positions, with the understanding that the acceptance criteria at stages 1 and 2 (still drawn from the USPC collaborative study) will depend on the number of positions in the assembly.

[1] US Pharmacopeial Convention, 12601 Twinbrook Parkway, Rockville, Maryland 20852, USA.
[2] To whom correspondence should be addressed. (e-mail: wh@usp.org)

METHODS

We assume that the percent dissolved values are normally distributed after log transformation. This assumption is based on experience with USPC collaborative studies of dissolution PVT reference standards, where the normality assumption was better satisfied after log transformation (2). We further assume a random-effects model where there are contributions to variability from laboratory, intermediate precision factors (particularly analyst and equipment within laboratory), and between positions within a given assembly. The between-position variability includes any contribution from the tablets and the assay as well as any differences among positions within an assembly. With these assumptions, we identify three variance components of interest: σ_L^2 , the variance component that includes between-laboratory and all intermediate precision contributions; σ_R^2 , the between-run variance component; and σ_P^2 , the between-position variance. The reproducibility variance is then $\sigma_L^2 + \sigma_R^2 + \sigma_P^2$.

For determining the operating characteristics of the acceptance limits at stages 1 and 2, we determined the probability of passing for specified values of a laboratory mean and variability. That is, the operating characteristics assume that all intermediate precision factors are held fixed. This is appropriate for the PVT, where interest is in a given assembly within a given laboratory, and the set of runs will be done by one analyst in a short period of time. The mean, $\overline{X}_6$, of results from 6 tablets in the log scale and the mean, $\overline{X}_{12}$, from 12 tablets that includes the first 6, will then be correlated and jointly normally distributed, with mean for both $\overline{X}_6$ and $\overline{X}_{12}$ of μ_L, the actual mean for that assembly–analyst combination at that time in that laboratory. The variances are $\sigma_6^2 = \sigma_R^2 + \sigma_P^2/6$ for $\overline{X}_6$ and $\sigma_{12}^2 = \sigma_R^2/2 + \sigma_P^2/12$ for $\overline{X}_{12}$. The correlation is $\rho = 1/\sqrt{2} = 0.7071$. There is a correlation because the data that determine $\overline{X}_6$ are also used in the determination of X_{12} . The corresponding between-position variance estimates, S_6^2 and S_{12}^2, will be distributed proportional to chi-squares with 5 and 10 degrees of freedom (*df*), respectively. The variance based on 12 tablets, S_{12}^2, is calculated by pooling (averaging in this case because the two sets of tablets are of equal size) the two intra-run variances. For converting an $\overline{X}$ and S^2 back to the percent dissolved scale, the geometric mean is found as $\exp(\overline{X})$, and the coefficient of variation (%CV) is found using the log-normal formula, $100\%\sqrt{\exp(S^2) - 1}$.

The probability of meeting the acceptance limits for the single-stage test option of 12 tablets (two runs) is found directly from the normal and chi-square distributions of $\overline{X}_{12}$ and S_{12}^2 , respectively. Similarly, the probability of meeting the acceptance limits on the first stage (six tablets) of the optional two-stage test is found directly from the distributions of $\overline{X}_6$ and S_6^2 . The probability of passing the two-stage test is found most easily as the probability of not failing. To fail the two-stage test, one must not meet the acceptance limits at both stages. For the mean, this is a bivariate normal probability,

$$\Pr\left[\left(\overline{X}_6 > U_6 \text{ and } \overline{X}_{12} > U_{12}\right) \text{ or } \left(\overline{X}_6 < L_6 \text{ and } \overline{X}_{12} < L_{12}\right)\right], \quad (1)$$

where the *L*s and *U*s are the lower and upper limits of the acceptance limits. The probability of falling outside the limits to the high side at the first stage and then to the low side after the second stage (and vice versa) is vanishingly small and is not needed in the calculation above. The two bivariate normal probabilities in Eq. 1 are then calculated from the following series expansion, using six terms in the summation (9):

$$\Pr\left[\left(\overline{X}_6 > U_6 \text{ and } \overline{X}_{12} > U_{12}\right)\right] = [1 - P(h)] \times [1 - P(k)] + \sum_{j=0}^{\infty} \frac{Z^{(j)}(h) \times Z^{(j)}(k)}{(j+1)!} \rho^{j+1}$$

where P and Z are the standard normal cumulative distribution and density functions, respectively, and

$$h = (U_6 - \mu_L)/\sigma_6 \text{ and } k = (U_{12} - \mu_L)/\sigma_{12}.$$

A similar formula is used for values that fall below the lower limits.

For the between-position variance, the probability of failing the two-stage test,

$$\Pr\left[S_6^2 > C_6 \text{ and } S_{12}^2 > C_{12}\right],$$

where the *C*s are the upper acceptance limits for the variances, was determined by simulation. One hundred thousand random gamma variates were simulated for each stage and converted to S^2s. This was done separately for the variance corresponding to the first stage and then with a new set of 100,000 random variates for the second stage. The two variances were then averaged to obtain S_{12}^2 . With 100,000 sets of simulated variances, the standard errors of the estimated probabilities of passing do not exceed 0.0016.

All calculations except the simulations were done in Excel 2002 (Microsoft Corp., Redmond, WA, USA). The simulations results were obtained using SAS 9.1 (SAS Inc., Cary, NC, USA).

RESULTS

Single-Stage Test

ISO International Standard 5725-6 (8) specifies the form of data analysis for proficiency tests. The context is that there has been a collaborative study to determine an overall mean and components of variability. The acceptance limits that allow a pass/fail decision for a PVT are drawn from this study. When conducting a PVT, an analyst computes a mean and repeatability variance and compares them to the acceptance limits derived from the collaborative study. The ISO acceptance limits for the mean are the collaborative study mean±2 standard deviations (SDs), where the SD is the reproducibility SD for the mean. For two sets of 6 tablets (12 tablets total), for example, this would be: $\mu \pm 2\sqrt{\sigma_L^2 + \sigma_R^2/2 + \sigma_P^2/12}$, where the mean and variances are estimated from the collaborative study. For the laboratory's repeatability variance, the ISO criterion is not to be statistically significantly different from the collaborative variance at the 5% level. The upper limit on the between-position variance is then $18.31 \times \sigma_P^2/10 = 1.831 \times \sigma_P^2$ for two sets of six tablets, where 18.31 is the

Table I. Example Apparatus 2 Data

							Geometric mean	Percent CV
Run 1	36.4	37.3	37.2	36.3	36.4	39.3	37.1	3.0
Run 2[a]	37.3	34.0	35.5	36.4	35.7	37.3		
Combined							36.6	3.3

[a] No geometric mean and %CV are shown for Run 2 because only the results of Run 1 and Combined are compared to the acceptance limits

upper fifth percentile of a chi-square distribution with 10 *df*. These criteria are approximate 95% coverage tolerance intervals for the laboratory means and variances.

For application to a USP PVT, we made three clarifications and/or observations. First, the mean and SD acceptance limits in the log scale correspond to limits for the geometric mean (GM) and %CV following transformation back to the original scale. Second, we interpreted the ISO criterion for the repeatability SD to be one-sided. That is, only a laboratory SD that is too much larger than the collaborative value will fail. Third, for dissolution, the focus is on position-to-position variability. This variability is a portion of the repeatability variability. The between-position variability should be determined for each set of six and then pooled across sets.

For Apparatus 2, we can then determine the acceptance criteria for the Lot P Prednisone RS Tablets where we have, from the collaborative study, a geometric mean percent dissolved of 41.4% and a between-position %CV of 8.5%. For a single-stage study, the acceptance criteria would then be that the laboratory geometric mean of 12 tablets must be on or within 34.5% to 49.7%, and the between-position %CV must be not more than 11.5%. In contrast, the current per-tablet acceptance limits for Prednisone in Apparatus 2 are 30% to 57%.

Table I presents data obtained for USPC Lot P Prednisone RS Tablets using Apparatus 2. The two runs were conducted on the same day, on the same equipment, and by the same analyst. For the one-stage test, only the combined results are applicable. The results, as shown in Table I, meet the acceptance limits from the collaborative study.

In Technical Specification 21748 (10), ISO recommended a minimum of 15 *df* for the laboratory's variability. USPC elected to increase the number of tablets in the PVT from that currently required (i.e., six) but to a lesser number than that called for by the 15-*df* recommendation. Figure 1 shows the probability of passing for 1, 2, and 3 sets of six tablets (5, 10, and 15 *df*). The vertical line at 8.5% shows the %CV from the Lot P collaborative study. What led to the choice of 12 instead of 6 or 18 is the comparison of the curves at the right end of the graph. With only six tablets, very large CVs still have substantial probabilities of passing. Although the study of 18 tablets yielded improved characteristics by comparison with the study of 12 tablets, the study of the additional 6 tablets shows decreasing return for the extra effort. Most of the additional value arose from advancing from 6 to 12 tablets. This approach allows only two runs of an assembly with six vessels rather than three—and if the data allow an early stop, then only one (see two-stage test below).

Focus on the mean and SD, rather than on the individual tablet results, raises a specific concern. Will the new approach be sensitive to the possibility that one position in the assembly is not performing correctly although the other five are? For Apparatus 2 using the current per-tablet approach, any individual result either less than 30% or greater than 57% will fail the assembly regardless of the values at the other positions. To address this concern, we considered a scenario in which a PVT is conducted with two sets of six tablets and one position of each of the two sets of six gives a result that is discordant with data from the other five positions. We considered that the nondiscordant results have some geometric mean and %CV. We can then vary the two discordant results and observe what happens to the geometric mean and %CV of the full set of 12. Figure 2a shows the results for a geometric mean of the nondiscordant results of 41.4% dissolved (the mean from the Prednisone Lot P collaborative study) and a between-position %CV of 5%, typical of values obtained in USPC's Dosage Form Performance Laboratory (11). As the discordant value ranges from 20% to 70%, the geometric mean of the 12 tablets (upper slanted line) stays within the acceptance limits (upper pair of horizontal lines). The %CV of the 12 results (lower curved line), however, exceeds the 11.5% limit (lower horizontal line) when the discordant value is either less than 32% or greater than 53% and thus is a more stringent criterion than 30% to 57% (vertical dashed lines). Combined, then, the range of discordant values for which the acceptance limits are met is 32% to 53%. Figure 2b shows a case with a lower geometric mean for the 10 tablets, 35%. Again, the mean and %CV acceptance criteria are more stringent than the per-tablet acceptance criteria, yielding a combined interval of discordant values for which the acceptance limits are met at 32% to 46%. In contrast to Fig. 2a, the acceptance limits for

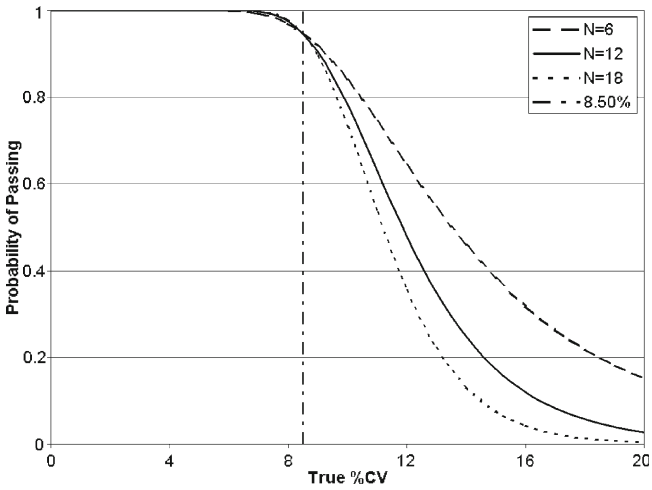

Fig. 1. Probability of passing new acceptance criteria for 1, 2, and 3 sets of six tablets

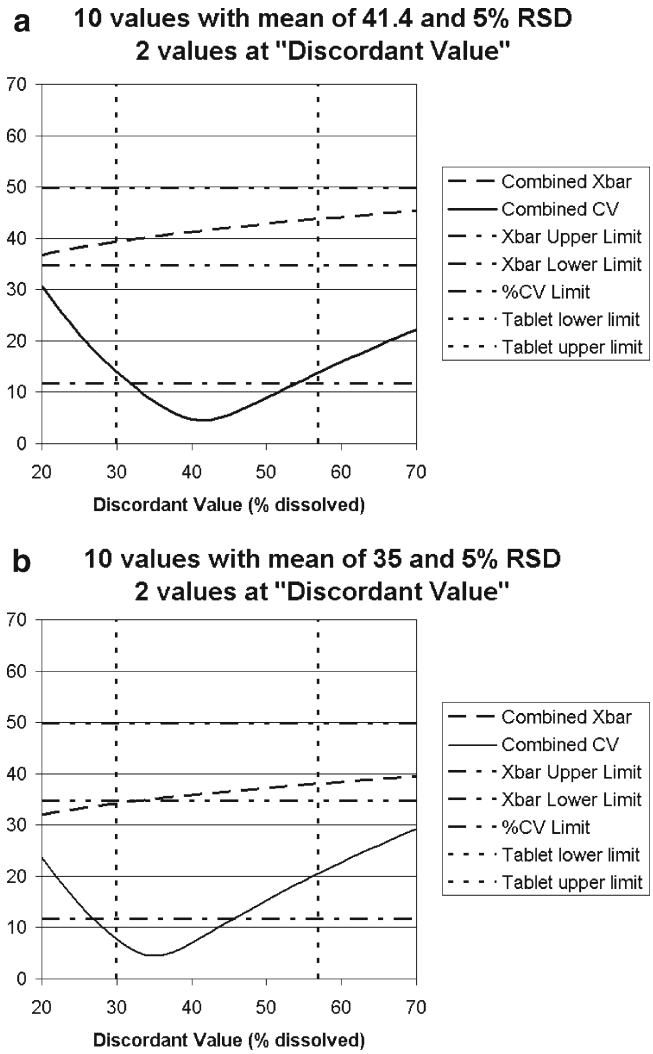

Fig. 2. **a** Determination of passing range of percent dissolved of two "discordant values" in a set of 12 results: laboratory mean=41.4%. **b** Determination of passing range of percent dissolved of two "discordant values" in a set of 12 results: laboratory mean=35.0%

the mean in Fig. 2b is not met for low values at the discordant positions. These two and additional cases are summarized in Table II. We concluded that the criterion based on the geometric mean and %CV is sensitive to the scenario of aberrant results from a single position.

Two-Stage Test

One of the suggestions made in the public comment following the original proposal to change the criterion (6,7) was to allow a two-stage test. A multistage test is already part of the dissolution procedure itself of *USP* General Chapter <711> and of the content uniformity test of *USP* General Chapter *Uniformity of Dosage Units* <905>. USPC and its Biopharmaceutics Expert Committee agreed with this suggestion as an option for the *USP* PVT.

The two-stage test is motivated by group sequential designs in clinical trials (12). The specific test chosen for the USP dissolution PVT is similar in concept to that of the group sequential test of O'Brien and Fleming (13). Group sequential designs require a "price" to be paid for the option to stop the test early and claim "pass." That price is that the acceptance limits after both the first and second stages will be narrower than those at the end of the single-stage test. The magnitude of the price in terms of the narrowing of the second-stage limits depends in substantial part on how stringent the limits are after the first stage. The wider the first-stage limits, the greater is the price at the end of the second stage. The general goal we adopted was to make acceptance criteria after the second stage similar to—though necessarily more stringent than—those of the single-stage test.

To implement this choice, the first-stage acceptance limits were determined as approximate tolerance intervals, as for the single-stage test, but with 60% coverage instead of 95%. With 60% coverage, the limits are narrower than with 95% coverage. Then, the second-stage acceptance limits were determined in a manner that approximately matched the operating characteristics of the single-stage test, i.e., retaining approximately the same probabilities of passing. The resulting criteria for Prednisone Lot P RS Tablets and Apparatus 2 are given in Table III. The geometric mean acceptance limits after the second stage of the two-stage test in this case are the same as after the single-stage test. It happens that the "price" to be paid for the two-stage test is sufficiently small that it disappears once results are rounded; i.e., the difference between the limits for the geometric mean for the single-stage test and those for the second stage of the two-stage test is smaller than the rounding. For the %CV, however, there is some price (i.e., difference) seen in Table III. The acceptance limits following the second stage of the two-stage test are somewhat more stringent (tighter) than after the single-stage test.

Table II. Passing Range of Percent Dissolved of Two "Discordant Values" in a Set of 12 Results[a]

Mean (%)	Percent CV	X Bar pass range (%)	Percent CV pass range (%)	Combined pass range (%)
41.4	0	At least 14	31–55	31–55
41.4	5	At least 14	32–53	32–53
41.4	10	At least 14	34–49	34–49
32	5	At least 51	24–42	None
35	5	At least 32	27–46	32–46
37	5	At least 24	28–48	28–48
41.4	5	At least 14	32–53	32–53
50	5	5–48	38–64	38–48

[a] The current per-tablet passing range is 30–57% for all rows

Table III. Acceptance Limits for Two-Stage Test for Apparatus 2

After 1st stage		After 2nd stage	
GM on or within	Percent CV nmt[a]	GM on or within	Percent CV nmt
38.2–44.9	8.6	34.5–49.7	11.2

[a] *nmt* not more than

For the data of Table I, as an example, after the first run, the data do not meet the acceptance limits; the geometric mean, 37.1, is below the lower acceptance limit, 38.2. Then, after the second-run data are obtained and combined with those of the first run, the second-stage limits are met, and the assembly passes.

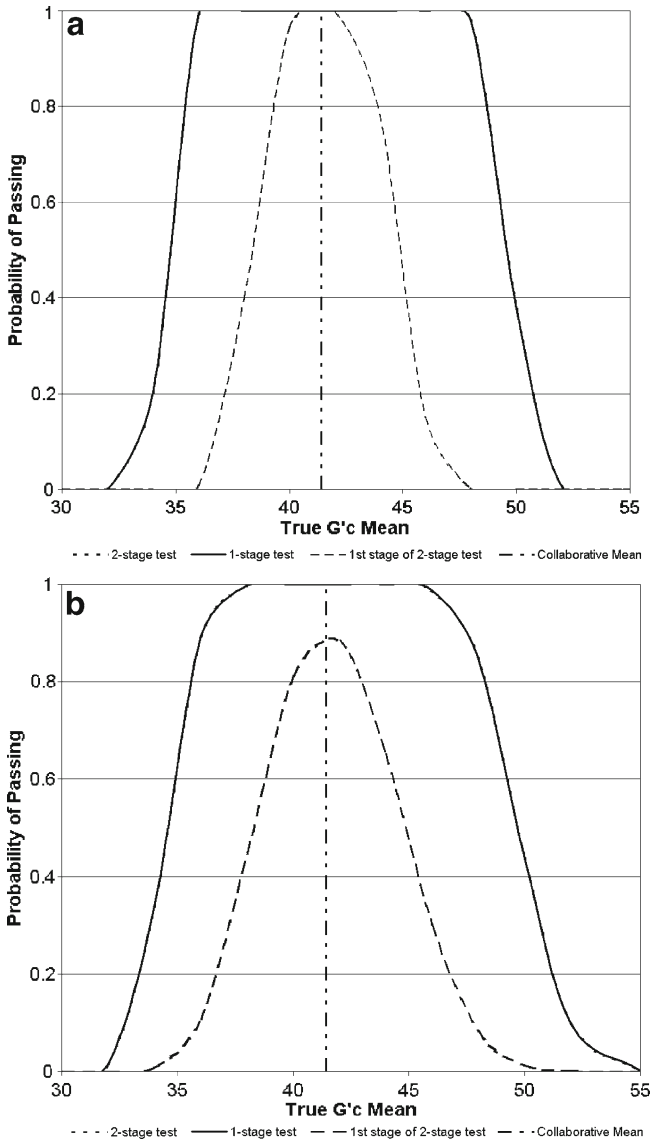

Fig. 3. a Comparison of operating characteristics (probability of passing) of one- and two-stage tests as a function of the laboratory's geometric mean: laboratory between-position %CV=6%. **b** Comparison of operating characteristics (probability of passing) of one- and two-stage tests as a function of the laboratory's geometric mean: laboratory between-position %CV=12%

Figures 3a and b show the operating characteristic curves (probabilities of passing) for the geometric mean acceptance limits for two sample choices of laboratory between-position %CV (%CV=6% and 12%, respectively). The curves for passing the two-stage test are not seen because they overlap the corresponding curves for passing the single-stage test. The inner curve is the probability of passing after the first stage of the two-stage test. Figure 4 shows the corresponding results for the %CV acceptance limits.

DISCUSSION AND CONCLUSIONS

USPC has conducted extensive evaluations of its PVTs. One consequence of this work is that USPC will change the data analytics approach for the dissolution PVT to bring it into agreement with ISO International Standard 5725 and to improve its operating characteristics (6,7). Of particular concern with the current limits that are set per tablet is that the results from a set of tablets can be very variable (at both extremes of the acceptance limits) and yet are considered passing results. With the new approach, results are assessed for consistency (%CV limits) and then for comparability to PVT results from a collaborative study of the reference material (geometric mean limits). This paper has shown how acceptance criteria from a USP collaborative study of reference material will be determined based on the new approach, as well as the advantages of using this approach. More details on what the limits would look like for dissolution Apparatus 1, 2, and 3 and step-by-step instructions

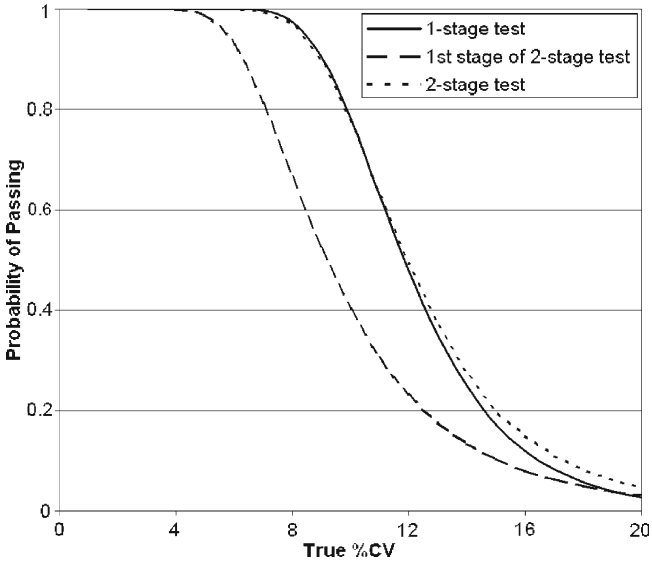

Fig. 4. Comparison of operating characteristics (probability of passing) of one- and two-stage tests as a function of the laboratory's %CV

are given in (14). The official date for the change is expected in late 2009.

With this effort and some concluding activities that will be presented elsewhere, USPC expects to conclude several years of activity on the dissolution PVT, which focused on justification, clarification of test requirements, elimination of the Salicylic Acid Tablet RS, and—in this article—improvement in PVT data analytics. Future efforts will focus on helping analysts make a smooth transition to the new data analysis procedures for the dissolution PVT. USPC remains convinced that a PVT is needed periodically to ensure the integrity of the USP Performance test when dissolution is relied upon. Further effort is being considered at USPC toward making the PVT study and the allied Prednisone Tablet into a true calibrator i.e., results from a PVT would be used to calibrate results across assemblies. At the same time, USPC is working on sound performance tests for drug products administered via other routes of administration, and these tests also are likely to need PVTs. This is the area in which USPC expects to devote its resources in the coming years.

ACKNOWLEDGMENTS

The authors thank Mario Gonzalez, Vivian Gray, Eli Shefter, and Charles Tan for their comments on an earlier version of this paper.

REFERENCES

1. US Pharmacopeial Convention. *USP–NF, Dissolution* <711>, US Pharmacopeial Convention, Rockville, MD, 2008, pp. 267–274.
2. US Pharmacopeial Convention. Compendial Tools: Dissolution Toolkit. 2008. http://www.usp.org/USPNF/compendialTools.html. Accessed 08 September 2008.
3. G. Deng, A. J. Ashley, W. E. Brown, *et al*. The USP performance verification test, part I: quality attributes and experimental variables contributing to dissolution variance. *Pharm. Res*. **25** (5):1100–1109 (2008).
4. M. Glasgow, S. Dressman, W. E. Brown, *et al*. The USP performance verification test, part II: collaborative study of USP's Lot P Prednisone Tablets. *Pharm. Res*. **25**(5):1110–1115 (2008).
5. USPC. In-process revision: Dissolution <711>. *Pharm. Forum*. **34** (5):1243–1251 (2008).
6. W. W. Hauck, R. G. Manning, T. L. Cecil, W. E. Brown, and R. L. Williams. Proposed change to acceptance criteria for dissolution performance verification testing." *Pharm. Forum*. **33** (3):574–579 (2007).
7. W. W. Hauck, T. L. Cecil, W. E. Brown, D. R. Abernethy, W. F. Koch, and R. L. Williams. USP responses to comments on *Stimuli* article, "Proposed change to acceptance criteria for dissolution performance verification testing. *Pharm. Forum*. **34** (2):474–476 (2008).
8. ISO. *ISO 5725 1–6: Accuracy (Trueness and Precision) of Measurement Methods and Results*, ISO, Geneva, Switzerland, 1994.
9. M. Abramowitz, and I. A. Stegun, eds. *Handbook of Mathematical Functions*. New York, NY: Dover; 1970:equation 26.3.29.
10. ISO. *Technical Specification 21748: Guidance for the Use of Repeatability, Reproducibility, and Trueness Estimates in Measurement Uncertainty Estimation*, ISO, Geneva, Switzerland, 2004.
11. W. W. Hauck, G. Deng, M. J. Glasgow, M. R. Liddell, P. Nithyanandan, and R. L. Williams. Variability of USP Lot P Prednisone RS Tablets. *Pharm. Technol*. **32**(6):90–94 (2008).
12. K. K. G. Lan, and D. L. DeMets. Discrete sequential boundaries for clinical trials. *Biometrika.* **70**:659–663 (1983).
13. P. C. O'Brien, and T. R. Fleming. A multiple testing procedure for clinical trials. *Biometrics*. **35**:549–556 (1979).
14. W. W. Hauck, A. J. DeStefano, W. E. Brown, *et al*. Description of the upcoming change in acceptance criteria for USP dissolution performance verification tests. *Pharm. Forum*. **34**(6):1630–1635 (2008).