Research Paper

The USP Performance Verification Test, Part II: Collaborative Study of USP's Lot P Prednisone Tablets

Maria Glasgow,[1] **Shawn Dressman,**[1] **William Brown,**[1] **Thomas Foster,**[2] **Stefan Schuber,**[1] **Ronald G. Manning,**[3] **Samir Z. Wahab,**[1] **Roger L. Williams,**[1,4] **and Walter W. Hauck**[1]

Purpose. Periodic performance verification testing (PVT) is used by laboratories to assess and demonstrate proficiency and for other purposes as well. For dissolution, the PVT is specified in the *US Pharmacopeia* General Chapter *Dissolution <711>* under the title *Apparatus Suitability Test*. For Apparatus 1 and 2, USP provides two reference standard tablets for this purpose. For each new lot of these reference standards, USP conducts a collaborative study.

Methods. For new USP Lot P Prednisone Tablets, 28 collaborating laboratories provided data. The study was conducted with three sets of tablets: Lot O open label, Lot O blinded, and Lot P blinded. The blinded Lot O data were used for apparatus suitability testing.

Results. Acceptance limits were determined after dropping data due to failure of apparatus suitability, identification of data as unusual on control charts, or protocol violations.

Conclusions. Results yielded acceptance criteria of (47, 82) for Apparatus 1 and (37, 70) for Apparatus 2. Results generally were similar for Lot P compared to results from Lot O except that the average percent dissolved for Lot P is greater than for Lot O with Apparatus 2.

KEY WORDS: acceptance limits; disintegration; dissolution; Performance Verification Test; United States Pharmacopeia.

INTRODUCTION

Dissolution is an important procedure in which attention to metrologic approaches is key to successful execution. Many of these approaches speak to ways that both define and reduce intralaboratory (repeatability) and interlaboratory (reproducibility) variance. A high-level review of these issues was the subject of the first article in this series [1]. That paper argued that the application of metrological principles to collaborative dissolution testing could help explain why intralaboratory (repeatability) results tended to show less variance and why interlaboratory (reproducibility) results tended to show greater variance. To address both types of variance, pharmaceutical manufacturers and the United States Pharmacopeial Convention (USP) have encouraged a Performance Verification Test (PVT), now termed *Apparatus Suitability Test* [2] in General Chapter *Dissolution <711>* in the *United States Pharmacopeia*

(*USP*). The *USP* PVT has used either commercially available or specially prepared tablets; the tablets currently used for the Apparatus 1 and 2 PVT are specially prepared. Although USP has never specified the frequency of the PVT, pharmaceutical and other manufacturers, as well as government control laboratories, have generally performed a PVT at least twice a year in their own laboratories, a periodicity that USP believes is appropriate.

In modern metrologic language, a PVT is a type of interlaboratory comparison with acceptance limits determined from a prior interlaboratory study conducted by USP. This type of study also has the character of both performance qualification (PQ) and proficiency testing, which is defined in the International Organization for Standardization (ISO) Guide 43-1, *Proficiency Testing by Interlaboratory Comparisons—Development and Operation of Proficiency Testing Schemes* [3]. As ISO 43-1 notes of proficiency testing,

> Participation in proficiency testing schemes provides laboratories with an objective means of assessing and demonstrating the reliability of the data they are producing ... 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 (p. v).

---

[1] United States Pharmacopeia, 12601 Twinbrook Parkway, Rockville, Maryland 20852, USA.

[2] USP 2005?–2010 Biopharmaceutics Expert Committee, University of Kentucky Medical Center College of Pharmacy, Lexington, Kentucky, USA.

[3] Department of Health and Human Services, Washington, DC 20201, USA.

[4] To whom correspondence should be addressed. (e-mail: rlw@usp.org)

For each new lot of tablets to be used in a *USP* PVT, USP conducts a multiple-laboratory collaborative study to determine acceptance criteria specific for the new lot. These studies generally include 25–30 laboratories representing industry, regulatory agencies, and pharmacopeias from multiple countries. For *USP* Apparatus 1 and 2, USP provides two types of reference standard tablets: prednisone and salicylic acid. This paper reports the collaborative study conducted for USP Lot P Prednisone Reference Standard Tablets, a new lot recently prepared. Characterization of Lot P Prednisone Reference Standard Tablets was discussed in Part I of the series (1). As the results of the present study show, repeatability and reproducibility results from collaborative testing of USP Reference Standard tablets display variance that can be excessive and thus influence results of the USP Performance test.

METHODS

Study Organization

In June 2005, USP invited 35 laboratories to participate in the USP Lot P Prednisone Tablets collaborative study. Twenty-eight laboratories from eight countries agreed to participate and completed the study.

Study Materials

USP provided each participating laboratory three separate sets of prednisone tablets, as follows: (1) Prednisone Lot O open label; (2) blinded Lot O; and (3) blinded Lot P.

Study Design

In prior collaborative studies of tablets for PVT, USP has focused on interlaboratory reproducibility as a means of determining acceptance criteria for a specified qualified tablet. In this study, a more complex study design allowed:

1. understanding of qualification results using blinded and open-label Prednisone Tablets Lot O;
2. as the main focus of the study, new acceptance criteria for the new Lot P Prednisone Tablets in Apparatus 1 and 2;
3. estimation of intralaboratory variance (intermediate precision) using a different analyst and equipment within a laboratory; and
4. comparisons between Lots O and P Prednisone Tablets.

Test conditions according to *USP* <711> were as follows: Apparatus 1 and Apparatus 2, six tablets each, 50 rpm, 500 mL deaerated purified water medium, 37°±0.5°, 30 min test time, with UV analysis at 242 nm.

Data

For the open label Lot O study, participating laboratories provided 54 experiments from 27 laboratories for Apparatus 1 and 56 independent experiments from 28 laboratories for Apparatus 2. One laboratory did not provide Apparatus 1 data. Another laboratory provided data on only one system, and another provided data on three systems for Apparatus 1 and two for Apparatus 2 (three experiments). For the collaborative study using blinded Lots O and P, for each lot a total of 54 experiments were reported from 27 laboratories for Apparatus 1 and 56 experiments from 28 laboratories for Apparatus 2. One laboratory did not provide data for Apparatus 1 for either lot.

Statistical Analyses

Each of the two Apparatus (1 and 2) was considered separately to establish acceptance criteria for Lot P Prednisone Tablets. For each apparatus, the statistical method was restricted maximum likelihood (REML) estimation of a nested, random-effects model. Specifically, the experiment was nested within laboratory, and laboratory and experiment were random effects. Analysis was done in SAS for Windows, Version 9.1 (SAS Inc., Cary, NC) using Proc Mixed. The default variance components covariance structure was used. This analysis estimated three variance components: interlaboratory, interexperiment (intralaboratory), and residual. These three components correspond approximately to reproducibility, intermediate precision, and repeatability (for further discussion, consult USP General Information Chapter *Validation of Compendial Procedures* <1225> (4) which, to the extent possible, is harmonized with the International Conference on Harmonization (ICH) Q2(R1) *Validation of Analytical Procedures: Text and Methodology*) (5).

The correspondence is not exact in two primary ways. First, the interexperiment component includes intermediate precision contributions only from analyst and equipment. Any other contributors to intermediate precision variability are included here in interlaboratory components. Second, repeatability should include multiple experiments by the same analyst on the same equipment. Any such variability over and above the residual variability is included here in the interexperiment component. The residual variability includes assay variability and any variability associated with the position of the vessel in the equipment and of the tablet within the vessel, as well as tablet-to-tablet variability. Preliminary analysis of the data (percent dissolved) in the original and in the natural log scales confirmed use of the log scale, as has been the case for prior USP collaborative studies. The choice was based on examination of the residuals for approximate symmetry. The natural log scale either improved the symmetry of the distribution of residuals or was little different. The (arithmetic) mean in the log scale was transformed back by antilog to the geometric mean in the original scale. All estimated variances, S^2, in the natural log scale are transformed back to coefficients of variation (CV) in the original, percent dissolved, scale as

$$CV(\%) = 100*\sqrt{\exp(S^2) - 1}$$

The acceptance limits are determined as $\exp\left(\overline{X} \pm z\sqrt{S_C^2 + S_E^2 + S_R^2}\right)$, where
$\overline{X}$ is the sample mean in the natural log scale,
z is a percentile of the standard normal distribution,
the three S^2 terms are the three variance component estimates, also in the natural log scale,
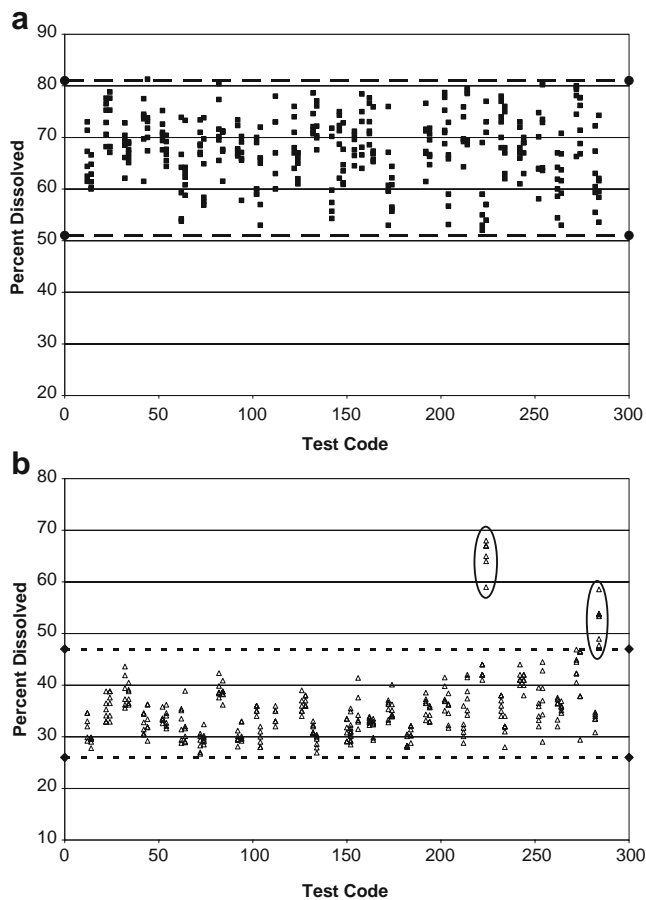
Fig. 1. Data for Lot O, open label. *Horizontal dashed lines* are the current Lot O acceptance criteria for Apparatus 1 (**a**) and Apparatus 2 (**b**). Two experiments (*circled*), one from each of two laboratories, fail for Apparatus 2. Another experiment that appears to fail due to a single value just outside the line for Apparatus 1 in the upper left passes using USP rounding rules. The *X*-axis is an arbitrary code identifying experiments. All experiments for a laboratory are grouped together with some horizontal separation between laboratories.

the *C*, *E*, and *R* subscripts denote Laboratory, Experiment, and Residual, and

the *exp* converts the acceptance limits from the natural log scale back to the percent dissolved scale.

Historically for USP collaborative studies, 95% limits ($z=1.96$ or 2.0) have been used. More recently, there has been recognition of the multiple testing associated with dissolution system performance checks. That is, although the acceptance limits are based on a single tablet, the performance test involves six tablets, all of which must pass. Use of the standard Bonferroni correction of 99% limits ($z=2.576$) is an approximate means of addressing this multiple testing.

Data Acceptance

Data were excluded for three reasons: (1) apparatus suitability failures, (2) unusual values as determined by Xbar and S control (Shewart) charts, and (3) protocol violations, in that order. Xbar control charts used ±6-sigma limits (6). Six-sigma was chosen instead of the more standard three-sigma

because the standard deviation used in the control chart limits is solely within-experiment, but the primary variability is between laboratories. The wider limits prevent too much data from being excluded. The S charts used two-sided 0.0027 probability limits, corresponding to ±3-sigma limits. For the S charts, limits corresponding to 3-sigma are used because within-experiment standard deviations are plotted. On the S charts only the upper limit was used because low variability is acceptable. The excluded data are described in the Results section.

Qualification Study

The collaborating laboratories were requested to perform and satisfy the requirements of the Apparatus Suitability Tests with Lot O (open label) before testing the blinded samples. Figure 1 shows these open-label Prednisone Lot O qualification study data. The horizontal lines are the current Lot O acceptance limits. In this and all subsequent figures, the *X*-axis is an arbitrary code that identifies experiments, and experiments from an individual laboratory are grouped
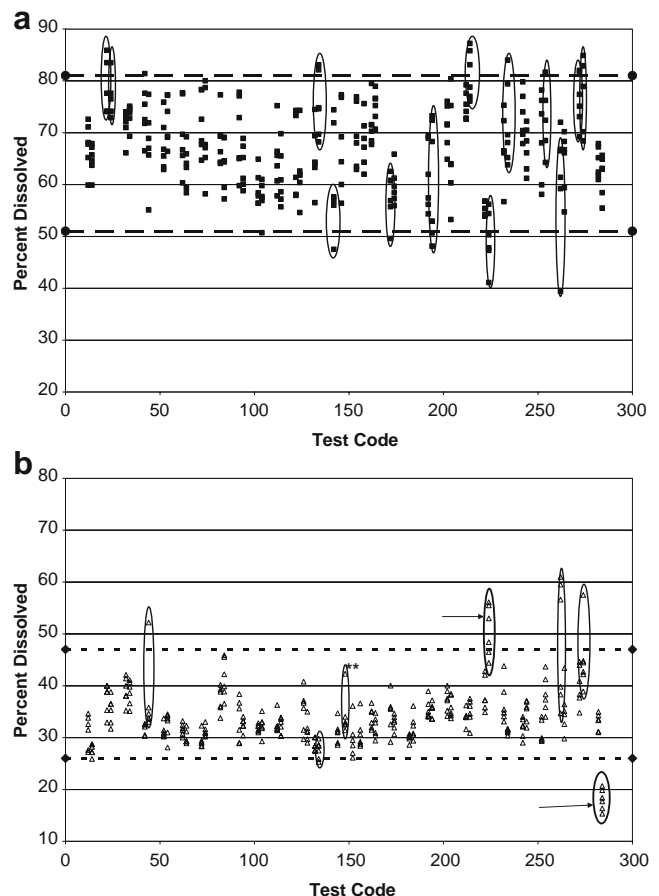

Fig. 2. Data for Lot O, blinded. There are 13 apparatus suitability failures for Apparatus 1 (**a**) and 7 for Apparatus 2 (**b**). Results that correspond to the two experiments that failed in the open label study, as shown in Fig. 1b, are *circled* (i.e., failed with blinded Lot O as well) and are marked by *arrows*. The one experiment for Lot P marked by ** here was excluded because that laboratory used different apparatus for Lot P than it used for the blinded Lot O suitability test.
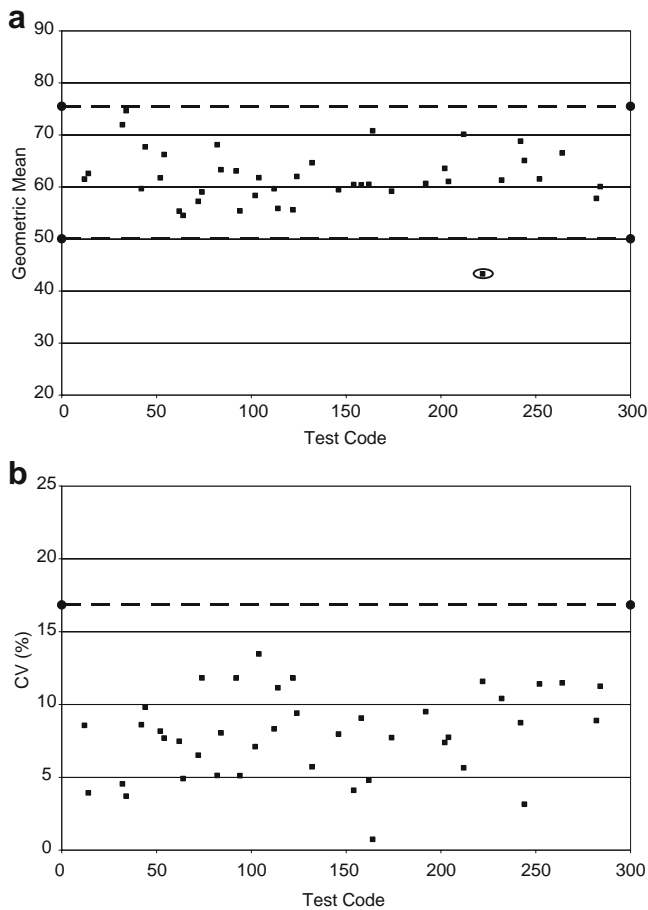
Fig. 3. Control charts for USP Lot P Prednisone Tablets, Apparatus 1. Control limits are shown as *dashed lines*. **a** is a six-sigma Xbar chart; **b** is a 0.0027 probability limit S chart. Only the upper limit is shown and used for the S chart. One unusual value on the Xbar chart is *circled*.

together. Laboratories failing suitability with open label Lot O were expected to make adjustments to their apparatus and/or technique and retest, and to continue this until they passed. Thus, it was expected that all data from this portion of the qualification study would be within the acceptance limits. Two laboratories failed the qualification study on Apparatus 2 but continued with the collaborative study.

Figure 2 shows the data for blinded Prednisone Lot O qualification study. Many more failures are evident (i.e., more data points are outside the limits) using the blinded data by comparison with the open label data of Fig. 1. For this reason, the blinded data were used for the apparatus suitability for inclusion in the determination of the Lot P limits. That is, in order for the experimental data for Lot P to be included in the determination of the acceptance limits, the same combination of apparatus and analyst had to satisfy the requirements of the Apparatus Suitability Tests *also* with blinded Lot O. Forty-one experiments for Apparatus 1 and 49 for Apparatus 2 satisfied this condition.

Control Charts

Figures 3 and 4 show the control charts for Lot P. One unusual value was identified on the Apparatus 1 Xbar chart

(Fig. 3a) and none on the S chart (Fig. 3b). For Apparatus 2, seven were identified on the Xbar chart (Fig. 4a) and one on the S chart (Fig. 4b). After these values were dropped, 40 experiments remained for Apparatus 1 and 41 for Apparatus 2. Although a control chart is not a time series, control charts are a useful means for examining variability in mean and standard deviation. Because control charts are not a time series, they do not facilitate examination of trends.

Protocol Violations

Laboratories were instructed to conduct two independent experiments for each lot; the second experiment was to use different equipment and a different analyst than the first. This step was not always followed. Six laboratories used the same equipment for both experiments for both Apparatus 1 and Apparatus 2. Of these six, three pairs were dropped because of study qualification and control chart considerations. For the three remaining pairs that had not followed the protocol, the second experiment was dropped from analyses. Thus, the final count for analyses was 38 experiments for Apparatus 1 and 40 for Apparatus 2.
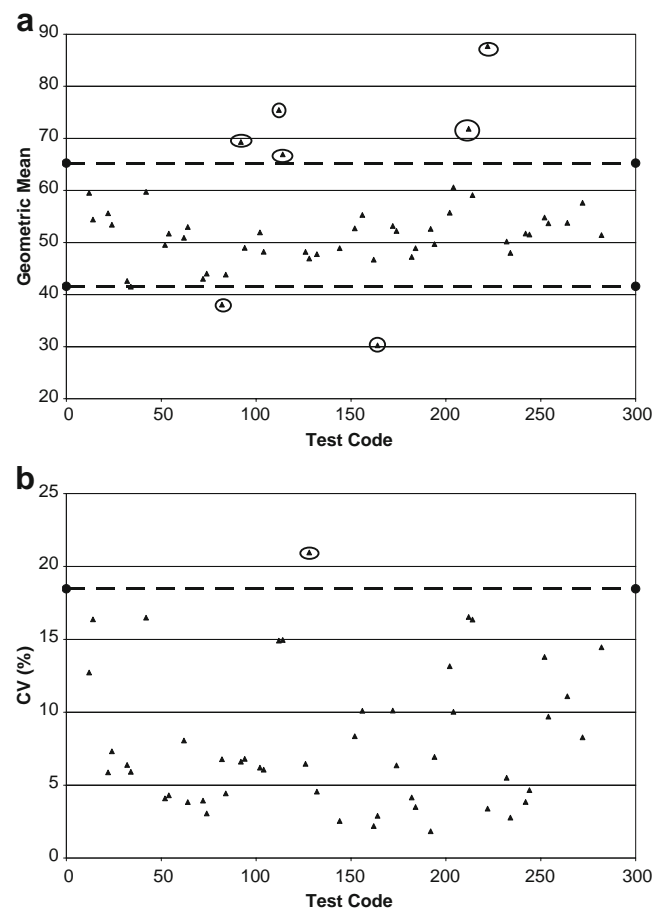

Fig. 4. Control charts for USP Lot P Prednisone Tablets, Apparatus 2 (labeling as for Fig. 3). Only the upper limit is shown and is used for the S chart. Seven unusual values on the Xbar chart and one on the S chart are *circled*.

Table I. Results from 2005 and 2003 Collaborative Studies for USP Prednisone Tablets

		Apparatus 1			Apparatus 2		
		Lot P, 2005	Lot O, 2005	Lot O, 2003	Lot P, 2005	Lot O, 2005	Lot O, 2003
Geometric mean		62.2	67.7	63.9	51.2	34.8	35.1
	Laboratory	5.2	6.7	2.5	8.8	6.6	8.1
CV (%)	Experiment	4.4	3.1	3.7	0.0	5.7	3.0
	Residual	8.1	8.6	7.9	8.5	7.6	7.7
Acceptance Limits: 99.0%		**(47, 82)**	(51, 91)	(51, 81)	**(37, 70)**	(26, 47)	(26, 47)

RESULTS AND DISCUSSION

This report yielded several observations.

Comparison of Lots O and P

Table I shows the results from the current collaborative study for Lot P. Results are also shown for the 2003 (open label) and 2005 (blinded) collaborative studies for Lot O. The results for Apparatus 2 for Lot P differ substantially from those for Lot O. Dissolution is faster with a (geometric) mean of 51% dissolved in 30 min for Lot P in comparison to 35% for Lot O. Variability is similar between the two studies, although the interlaboratory variability is somewhat greater for Lot P for Apparatus 2, leading to wider acceptance limits. Visual inspection of the vessels suggests an explanation: a dense and symmetric cone formation for Lot O in Apparatus 2 experiments was observed, in contrast to Lot P tablets where the cone was less dense and symmetric (1). USP may continue to explore this observation, although the impact of the cone formation on the use of new Lot P Prednisone Tablets is minimal. Any manufactured or specially prepared tablet must be re-qualified with changes in components or composition and/or method of manufacture/preparation, as was the case here (1).

Contributors to Variability

The "Laboratory," "Experiment," and "Residual" rows of Table I show the CVs corresponding to these three components of reproducibility in this study. The Laboratory values are the additional contribution to variability from differences between laboratories. This is a large contribution, particularly for Apparatus 2. The Experiment component is the contribution from analyst and equipment, two important components of intermediate precision. These components are relatively small, particularly for Apparatus 2. The Residual component includes all sources that contribute to variability within an experiment and is part of repeatability. This component is about 8% CV, consistent for the two lots and apparatus. From other studies (1) it is clear that variability associated with the tablet is less than 4–5%, so the balance of this variance component is due to position in the equipment, position of tablet within the vessel, and the assay procedure.

Acceptance Limits

Acceptance limits based on 99% are shown in Table I. The 99% limits shown in bold, namely (47, 82) for Apparatus 1 and (37, 70) for Apparatus 2, were proposed to and approved by USP's Biopharmaceutics and Reference Standards Expert Committees in March 2006. They are the limits that will apply to laboratories conducting a *USP* PVT using USP Lot P Prednisone Tablets, which have now entered commercial distribution. Figure 5 shows all Lot P data, including data not used in the analyses that determined the acceptance limits, and the acceptance limits for Lot P.

SUMMARY

USP is committed to conducting collaborative studies on its reference standards tablets for use in the *USP* PVT, using the highest level of laboratory, statistical, and metrological science. To further these ends, USP has many initiatives in
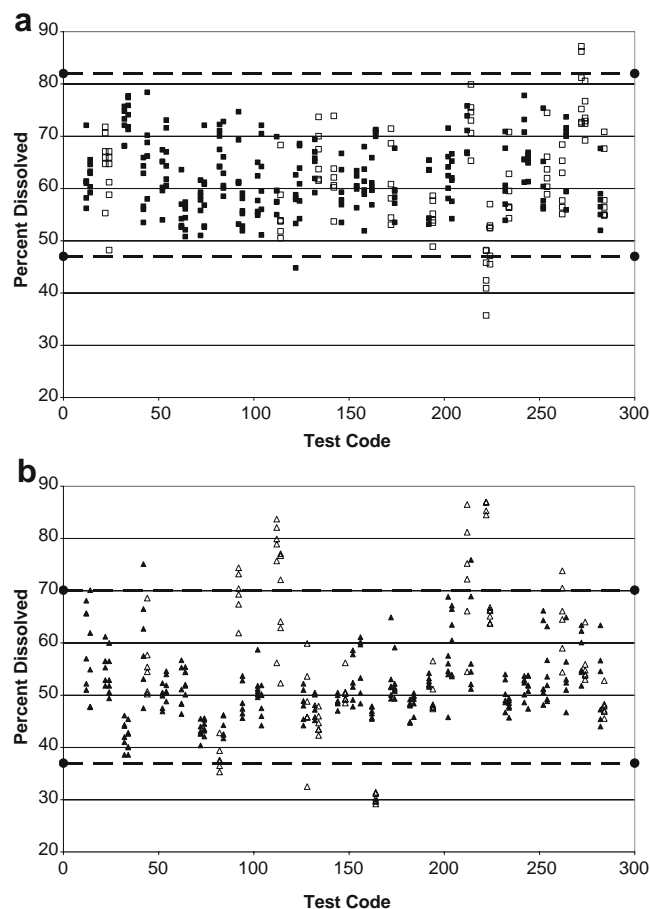

Fig. 5. Data for USP Lot P Prednisone Tablets and acceptance criteria. **a** Apparatus 1. **b** Apparatus 2. *Closed symbols* show data used in the determination of the acceptance limits. *Open symbols* are data not used in that determination.

progress or concluded (e.g., 7). USP is also looking into the processes and requirements for participating laboratories, given the substantial contribution of interlaboratory variability to the acceptance limits. USP is also considering changing the acceptance from a per-tablet basis, as reported here, to limits that follow ISO more closely. ISO Guide 5725-6 recommends limits for the laboratory average and for the within-laboratory variability (8). A positive consequence of such a change would be the elimination of the multiple-testing issue associated with testing six tablets to a per-tablet limit but under the constraint that all must pass for the system to be considered suitable (9). The general approach helps ensure the integrity of the dissolution procedure when applied to marketed solid oral dosage forms.

REFERENCES

1. G . Deng, A. J. Ashley, W. E. Brown, *et al*. The USP performance verification test, part I: USP Lot P Prednisone Tablets—quality attributes and experimental variables contributing to dissolution variance. DOI 10.1007/s11095-007-9498-7 (2008)

2. USP, *USP 30–NF 25, Dissolution <711>, Apparatus Suitability Test,* US Pharmacopeial Convention, Inc., Rockville, MD, 2006, p. 282.

3. ISO, *Guide 43-1. Proficiency Testing by Interlaboratory Comparisons—Development and Operation of Proficiency Testing Schemes,* 2nd ed., ISO, Geneva, Switzerland, 1997.

4. USP, *USP 30–NF 25, Validation of Compendial Procedures <1225>.* US Pharmacopeial Convention, Inc., Rockville, MD, 2006, pp. 680–683.

5. ICH, Q2 (R1) Validation of Analytical Procedures: Text and Methodology.www.ich.org/LOB/media/MEDIA417.pdf. Accessed June 28, 2007.

6. R. A. Johnson and D. W. Wichem, *Applied Multivariate Statistical Analysis*, fifth edition, Upper Saddle River, NJ: Prentice Hall, 2002.

7. W. W. Hauck, V. P. Shah, S. W. Shaw, and C. T. Ueda Reliability and reproducibility of vertical diffusion cells for determining release rates from semisolid dosage forms. *Pharm. Res*. **24**: 2018–2024 (2007).

8. ISO. Guide 5725. *Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 6: Use in Practice of Accuracy Values,* ISO, Geneva, Switzerland, 1994.

9. W.W. Hauck, R.G. Manning, T.L. Cecil, W. Brown, and R.L. Williams. Proposed changed to acceptance criteria for dissolution performance verification testing . *Pharm. Forum.* **33**:574–579 (2007).