Pharmaceutical Research, Vol. 22, No. 2, February 2005 (© 2005)
DOI: 10.1007/s11095-004-1184-4

Commentary

Oral Dosage Form Performance Tests: New Dissolution Approaches

Walter W. Hauck,[1] Thomas Foster,[2] Eric Sheinin,[3] Todd Cecil,[3] William Brown,[3] Margareth Marques,[3] and Roger L. Williams[3,4]

Received May 21, 2004; accepted October 26, 2004

The performance test is one of a series of tests that compose the specification in a United States Pharmacopeia (USP) dosage form monograph. For an orally administered, nonsolution dosage form, it is usually satisfied by either a dissolution or disintegration procedure. Dissolution acceptance criteria are usually set in private negotiations between an applicant and a regulatory agency. With information about this private agreement and other information provided in a sponsor's Request for Revision to USP, the USP's Council of Experts elaborates a public dosage form monograph. Based on the relationship between the regulatory decisions and the Request for Revision, the USP dissolution procedure links to a regulatory judgment about bioavailability and bioequivalence and, ultimately, to a judgment about safety and efficacy. The current dissolution procedure and acceptance criteria are perceived as having worked well over the years and are generally accepted. This article discusses new approaches that merit consideration. These approaches focus on a) explicit use of hypothesis testing, b) use of parametric tolerance intervals, c) improved ways to set dissolution acceptance criteria, and d) a more flexible protocol to assess conformity. Application of the proposed approaches may better assess, manage, and communicate both manufacturer and consumer risk for dissolution testing.

KEY WORDS: confidence intervals; content uniformity testing; dissolution testing; dosage forms; equivalence test; tolerance intervals.

INTRODUCTION

For orally administered nonsolution dosage forms, *in vitro* performance test procedures such as dissolution and disintegration are used to i) guide drug development and select formulations for further *in vivo* studies, ii) evaluate comparability between products before and after changes in formulation and/or manufacturing; iii) serve as a surrogate for *in vivo* bioequivalence studies, with suitable *in vitro/in vivo* correlations and/or use of the Biopharmaceutics Classification System approach, and iv) ensure batch-to-batch consistency for product performance (1–5). These procedures are frequently part of a private or public drug product specification, which is defined as a list of tests, references to analytical procedures to evaluate those tests, and appropriate acceptance criteria (6). Conformity testing to ingredient, product, and other specifications supports a conclusion that a drug product and its ingredients are acceptable for their intended use. These statements depart to some extent from the compendial approach termed "singlet testing" (7), which the authors believe needs revisiting.

A drug product specification submitted in a regulatory application frequently relies on public disintegration or dissolution procedures in the *United States Pharmacopeia–National Formulary* (*USP–NF*), with adaptation of specific elements to the particular product (8,9). After approval, the manufacturer may provide information to USP that allows the USP Council of Experts to consider the adapted procedure and acceptance criteria. With endorsement of the council and opportunity for public comment via the *Pharmacopeial Forum,* the private procedure and acceptance criteria can become the public USP performance test in the USP monograph for the specified dosage form. Private acceptance criteria for the dissolution procedure are generally agreed to between an applicant and a regulatory agency based on clinical trial, stability, and other development batches (2). The criteria can further be modified in postapproval filings based on subsequent manufacturing experience (10–11). The authors acknowledge that *in vitro* dissolution is primarily a quality control test, particularly in the absence of an *in vitro–in vivo* correlation. Nonetheless, the general approach allows the dissolution test in a USP dosage form monograph to link to the U.S. regulatory decision about an applicant's bioavailability and bioequivalence data, which in turn links to the regulatory decision about an applicant's safety and efficacy data (12). This is a crucial point, given that reliance on the dissolution procedure in a USP dosage form monograph by a manufacturer of the dosage form in another country does not, in and of itself, assure appropriate *in vivo* performance of that dosage form. For oral immediate and modified release dosage forms, *USP–NF* describes four types of apparatus, together

[1] Biostatistics Section, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA.
[2] University of Kentucky Medical Center, Colleges of Pharmacy and Medicine, Lexington, Kentucky 20536, USA.
[3] U.S. Pharmacopeia, Rockville, Maryland 20852, USA.
[4] To whom correspondence should be addressed. (e-mail: rlw@usp.org)

with means of assessing conformity to acceptance criteria (13).

Although the current dissolution procedure and acceptance criteria are perceived as having worked well and are generally accepted, their scientific basis has not been revisited recently. (Note: The authors acknowledge the harmonization of dissolution in the Pharmacopeial Discussion Group, which is composed of representatives of the European, Japanese, and United States Pharmacopeias. This harmonization has not concluded. The current article speaks to future considerations and is not intended to impact on the PDG harmonization effort.) For this reason, we reconsidered the dissolution procedure and suggest new approaches for further consideration. These approaches focus on a) explicit use of hypothesis testing, b) use of parametric tolerance intervals, c) improved ways to set dissolution acceptance criteria, and d) a more flexible protocol design to assess conformity.

HYPOTHESIS TESTING

The current USP dissolution procedure (9) relies on a three-stage sequential experimental design (S_1, S_2, S_3), with the possibility of stopping at each stage and meeting or failing to meet the criteria. In designing a sequential experiment, three elements should be considered: 1) sample size (N), 2) likelihood (alpha) of a false positive (type I) error (passing a test that should not be passed—consumer risk), and 3) likelihood (beta) of a false negative (type II) error (failing a test that should be passed—manufacturer risk). Typically, experimental design proceeds by specifying acceptable levels of alpha and beta and then determining the sample size, N. An alternative acceptable approach might specify alpha and N and then determine the power of the study (1 − beta). In contrast, the current USP dissolution procedure focuses on sample size, intended to be small, and beta, also intended to be small, without specifying the hypothesis to be tested and without explicit consideration of alpha—the degree of allowable consumer risk.

The application of statistical hypothesis testing to a dissolution procedure requires a statement of the null and alternative hypotheses. Stated explicitly, the null hypothesis is that the group of units, such as a batch, from which some are selected to be tested is not acceptable, and the alternative hypothesis is that it is acceptable. If the characteristics of the sample are sufficiently good, the null hypothesis is rejected in favor of the alternative (i.e., the test passes). Setting alpha at some specified level of consumer risk (e.g., a chance of making a type I error not greater than 1 in 20 times, or $p \leq 0.05$) allows a science-based approach for the dissolution procedure.

PARAMETRIC TOLERANCE INTERVALS

USP currently uses a mostly nonparametric approach for both content uniformity and dissolution testing. For content uniformity testing, Katori *et al.* proposed replacing the current USP approach (test by attributes) with a parametric approach (test by variables), using a tolerance interval (14). The Katori *et al.* proposal is under consideration in the Pharmacopeial Discussion Group. Moving from a nonparametric to a parametric approach depends in part on whether a specific distribution, such as the normal, can be assumed. The normal distribution can be a reasonable assumption for content uniformity testing but may not be for dissolution testing. Normal distribution of dissolution data may be unlikely due to the boundary of 100% dissolved—with values sometimes greater than 100% given variability in assay and content. The 100% boundary forces nonsymmetry and hence non-normality of distribution. Although data transformation may help, one solution to this issue would be to choose the time at which dissolution is measured to achieve values that are acceptably below 100%.

In contrast to use of a parametric tolerance interval approach for content uniformity testing, Katori *et al.* (15) proposed a parametric confidence interval approach for dissolution testing (Table I). Both confidence and tolerance intervals estimate the characteristics of a distribution. Confidence intervals assess the precision of estimates of single quantities (e.g., mean, variance). For example, a 95% confidence interval for a mean consists of all values within ±2 standard errors of the estimated mean; the more precise the estimate, the narrower the confidence interval. In contrast, tolerance intervals describe ranges of specified coverage for a distribution of values (16); for example, at least 80% of dissolution values for a product fall within a specified range with some specified level of confidence. A choice between the two approaches is

Table I. Dissolution Criterion

Stage	Number of units tested	Cumulative number of units tested	Acceptance criteria[a]		
			USP ⟨711⟩	JP proposal[b]	Tolerance interval proposal[c]
S_1	6	6		$\overline{X}_1 - 2.015S_1/\sqrt{6} \geq Q^*$	$\overline{X}_1 - 1.91S_1 \geq Q$
			All units $\geq Q + 5\%$	All units $\geq Q^* - 10\%$	
S_2	6	12	$\overline{X}_2 \geq Q$	$\overline{X}_2 - 1.796S_2/\sqrt{12} \geq Q^*$	—
			All units $\geq Q - 15\%$	All units $\geq Q^* - 10\%$	
S_3	12	24	$\overline{X}_3 \geq Q$	—	—
			All units $\geq Q - 25\%$		
			At least 22 units $\geq Q - 15\%$		

[a] $\overline{X}$ and S are the sample mean and standard deviation. Q and Q^* are the acceptance limits. Subscripts denote first or second stage; each mean and standard deviation is calculated using the cumulative number tested at that stage.

[b] The 2.015 and 1.796 in the JP proposal are 5th percentiles from t distributions with 5 and 11 degrees of freedom, respectively.

[c] This is an illustrative choice. Criterion is that at least 75% of the batch would be dissolved at least Q. The 1.91 is the approximate value for a one-sided normal tolerance interval for 75% coverage and 95% confidence (interpolated from Ref. 16, Table A12c).

based on what quantity (or quantities) best characterizes the distribution for the problem of interest. As discussed below, characterizing the distribution solely by the mean and its confidence interval is insufficient—they are essentially uninformative as to the width of the distribution. To consider also variability, standard deviation controls could be established. We argue that a tolerance interval approach is a good overall solution, based on the following discussion.

The confidence interval approach proposed by Katori *et al.* for dissolution testing is similar to that now used for *in vivo* bioequivalence studies. The general approach requires that an observed confidence interval, obtained from analysis of data comparing "test" and "reference," falls within prespecified equivalence limits (acceptance criteria or goalposts) at a certain level of confidence. For a dissolution procedure, the "test" data represents dissolution data for units of the dosage form under evaluation. The "reference" is a fixed value that is usually expressed as a percentage of the label claim—the amount of drug substance in the dosage form. The acceptance criteria (Q) could be one-sided (noninferiority) or two-sided (both noninferiority and nonsuperiority), depending on whether it is sufficient to control only for excessively slow (one-sided) or both excessively slow and excessively rapid dissolution (two-sided). The Katori *et al.* proposal specifies two stages and six units tested at each stage. It also includes a zero tolerance value (safety net); that is, no value for a single unit can be less than $Q - 10\%$.

The Katori *et al.* confidence interval approach for dissolution testing has some limitations. First, it is designed as if alpha (consumer risk) is ≤5%. However, the risk is actually closer to 8–9%, because it relies on the two specified stages with 5% consumer risk at each stage. This can be fixed with wider confidence intervals at each stage. More importantly, although both confidence intervals and tolerance intervals control a change in average dissolution, an approach consisting solely of a confidence interval for the average does not provide limits on variability. As an aggregate criterion, a tolerance interval approach directly controls both mean and variability (16). Figure 1 provides various combinations of means and standard deviations that correspond to 90% of a group of units tested. The control is direct, because a deviation in mean or an increase in variability will reduce the covered proportion. With a confidence interval approach, two similar means with widely differing distributions can fall within an equivalence interval, providing a sufficiently large sample is studied (Fig. 2). The Katori *et al.* confidence interval approach for dissolution testing exerts indirect control on variability by limiting the sample size, making it more difficult to pass when variability is increased. The zero tolerance component of both current USP practice and the Japanese Pharmacopeia (JP) proposal should also serve to limit variability. Tsong *et al.* note, however, that the current USP approach for dissolution does not control the proportion below the dissolution acceptance criteria (Q) very well (17).

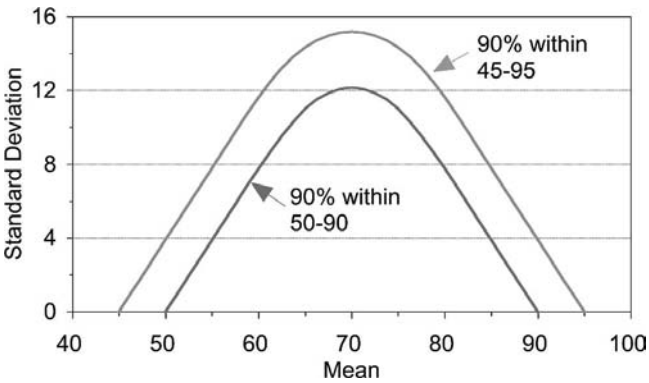


Fig. 1. For all combinations of means and standard deviations on or below each curve, at least 90% of the distribution falls within the tolerance limits. The upper curve is for tolerance limits of 75–125% of label claim. The lowest curve, with narrower specified tolerance limits of 80–120%, is more restrictive.

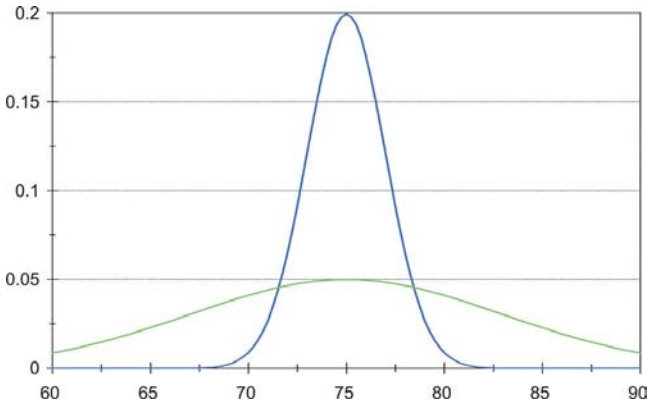


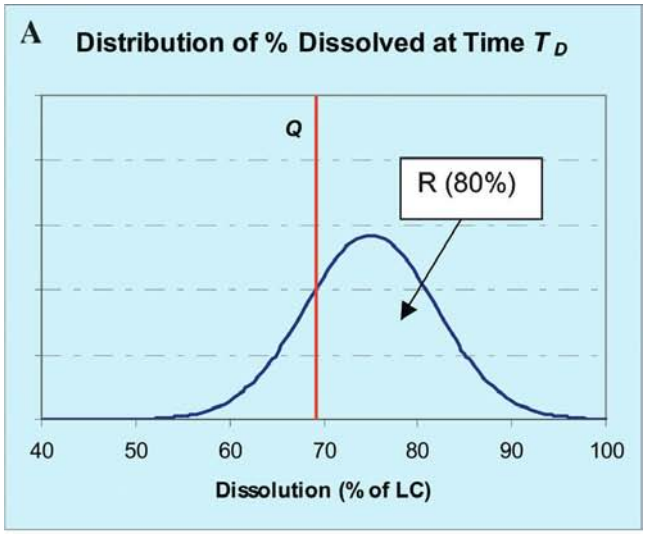
Fig. 2. The illustration demonstrates how different two distributions can be while having equal means.

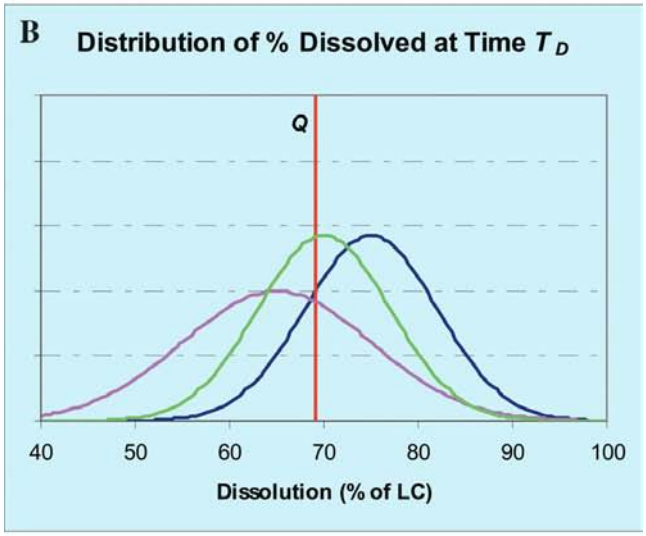
Based on these considerations, we propose using a parametric tolerance interval approach to analyze data from a dissolution procedure, just as it is being adopted for content uniformity testing. A parametric tolerance interval approach, by using actual values, ought to make better use of the data than approaches based on counts of values falling within specified intervals. A tolerance interval approach controls both mean and variance directly, in contrast to a confidence interval approach that focuses only on the mean. And, finally, a tolerance interval is more informative about the width of the distribution of dissolution values.

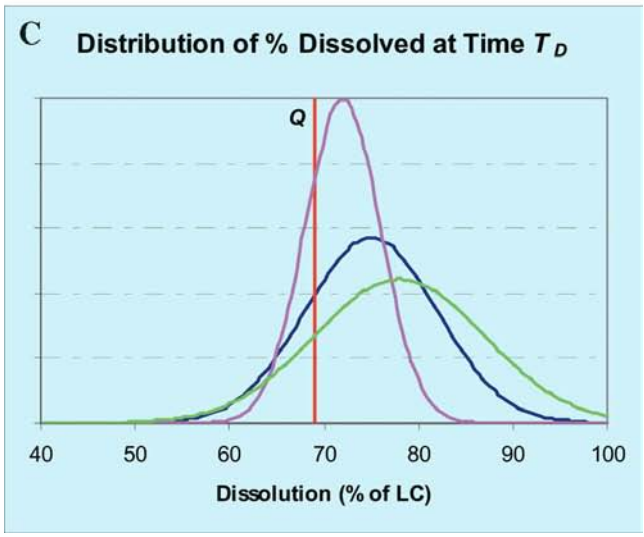
ACCEPTANCE CRITERIA USING A TOLERANCE INTERVAL APPROACH

Acceptance criteria for the dissolution procedure can either be a limit (the USP Q value) or a range (upper and lower percent dissolved at a specified time). Although acceptance criteria may be based on an *in vitro–in vivo* correlation (4,18), setting acceptance criteria frequently lacks a formal approach for most dosage forms (5). In this article, we propose a tolerance interval approach as a means of setting a dissolution limit or range, using *in vitro* dissolution results of preapproval batches—batches used in clinical trials, bioavailability and/or bioequivalence studies, stability testing, and other studies conducted during drug development. These preapproval batches directly or indirectly reflect the *in vivo* performance of batches used to document safety and efficacy. The application of a tolerance interval in this setting is distinct yet similar to the use of a tolerance interval approach to judge the acceptability of batches to be released into the marketplace, as discussed in the previous section. As with this approach, the proposed tolerance interval approach to setting accep-

tance criteria can be expressed as an equivalence test. In this setting, an equivalence test asks whether the dissolution of units in the postapproval batches is *sufficiently similar* (equivalent) to the dissolution of units in the preapproval batches. Using a tolerance interval approach, a percentile of the distribution of dissolution data from the preapproval batches is used to establish an acceptable minimum dissolution value. The approach represents a way of thinking about an acceptable minimum dissolution value (Q), using percentile as a means of defining acceptance, as follows (Figs. 3A–3C).







1. The value R defines a minimum proportion for the distribution of the dissolution results from preapproval batches at a specified time point (T_D). T_D is chosen to be in an acceptable region of the percent dissolved/time curve (e.g., <80%). To control for low dissolution values, R should be >50% or higher. Although R could be very large (e.g., 99%), statistical methods to deal with extreme tails require very large sample sizes. As an intermediate approach, we suggest R = 80% or 90%. Figure 3A shows an example with R = 80%.

2. For immediate-release dosage forms, Q is a limit below which only a small proportion of the dissolution values should fall. The choice of R determines Q. Q is that value of percent dissolved such that a proportion of the units (1 – R) in the baseline batches have dissolution values less than Q. The approach is also suitable if an upper limit is needed. The value for Q is initially a private standard and can become a public standard based on submission of suitable information in a Request for Revision to USP. In Fig. 3A, Q is found as 69.1%; 80% of the distribution exceeds 69.1%.

3. The value α represents the allowable level of consumer risk, typically taken to be 0.05, which is the probability of falsely claiming that the batch meets the set standard.

4. The value P represents the minimum proportion of individual dosage form units from a postapproval batch that should exceed Q. The choice of P requires some consideration. Choosing P greater than R requires that future batches be better (exhibit faster dissolution) than the preapproval batches. This is counter to the initial presumption that the premarket batches are acceptable. If P is set equal to R, the structure of the statistical hypothesis test dictates that the chance of passing is only α when the true proportion is R, what is, again, supposed to be an acceptable value. This leads to substantial producer risk or requires a manufacturer to accelerate dissolution relative to the premarket batches. Set-

Fig. 3. (A) The curve shows a possible baseline distribution. If we pick R = 80%, then Q = 69.1% of label claim (LC), the 20th percentile of this distribution. That is, at least 80% of the units in these baseline batches exhibit dissolution of at least 69.1% (Q) by time T_D. (B) The right-hand curve is the baseline distribution of Fig. 3A. The other two curves are two distributions that do not meet the standard of at least P = 75% (for illustrative purposes) of units greater than Q, the 20th percentile of the baseline distribution. The middle distribution is shifted to the left of baseline but has similar variability; its 25th percentile is 65.3% of label claim. The flatter distribution is shifted left and shows greater variability; its 25th percentile is 58.3% of label claim. (C) The middle curve is the baseline distribution of Fig. 3A. The other two curves are two distributions that do meet the standard of at least P = 75% (for illustrative purposes) of units greater than Q, the 20th percentile of the baseline distribution. The flatter distribution is more variable than the baseline, but the average percent dissolved is greater, so that the 25th percentile is 71.9% of label claim. The taller curve is shifted left to a lower average percent dissolved at T_D but shows less variability than the baseline distribution, so its 25th percentile is 69.3% of label claim, still greater than Q (69.1%).

ting P less than R, however, suggests a relaxation of a quality standard. We propose that P be less than but close to R, allowing a small difference (e.g., 5%) that is not discernible or likely to be related to changes in *in vivo* performance. Thus if R is set at 80%, P should be 75% (see examples in Figs. 3B and 3C).

Using a one-sided tolerance criterion, the statistical analysis of a sample from a postapproval batch determines the tolerance interval and compares it to the limit (acceptance criterion) Q. Specifically, the batch would pass if $\overline{X} - kS > Q^*$, where $\overline{X}$ is the average percent dissolution of the units in the postapproval batch to be released, S is the standard deviation of these units, and k is a constant that depends on P, α, sample size of each stage, and statistical design of the sampling. Table I provides a comparison of the three types of approaches: nonparametric (USP), parametric confidence interval (JP), and parametric tolerance interval (proposed in this article). The important aspects of the design will be the number of tiers (k will be different at each tier) and the number of units tested at each tier. Only a single-stage comparison is currently possible for all three approaches. For more than one stage, the k values need to account for both the multitier nature of the design and potential contributions to variability from experimentation in sets of six dosage form units. Methods for tolerance intervals allow for experimentation in sets of six (19–22), but these need to be extended to multitier designs, as has been done for content uniformity (23).

FLEXIBLE DESIGN OF THE DISSOLUTION TEST PROCEDURE

In a separate article, the possibility of more flexible protocol designs for content uniformity testing was discussed (24). A primary argument for flexibility is to give the manufacturer control over the risk of failure. The same approach could be applied for dissolution testing. With flexible study designs, the regulatory agency and/or compendium would agree on the allowable consumer risk and performance of an acceptable batch or unit (the alternative hypothesis in statistical language). Using a tolerance interval approach, this would be achieved by establishing i) the limit (Q), ii) the minimum proportion of the batch that should fall within the criteria (P), and iii) the degree of confidence needed to reach an accept/reject decision (1 – alpha). With a good understanding of the performance (mean and variability) of their product relative to these standards, a manufacturer determines the number of testing stages and units tested at each stage to satisfy three public standards: Q, P, and consumer risk (alpha). The specification for the dosage would then include the specified testing protocol and units tested, which might differ between manufacturers.

One advantage for the manufacturer of flexible designs is better control over producer risk of failing good batches or samples. This is a consequence of the manufacturer choosing the sample size. We consider it premature to develop details on sample sizes. We do note, however, that concern has been expressed that a tolerance interval approach, by dealing with the spread of the distribution, may require larger samples than a confidence interval approach. This is possible but very dependent on the choices of P and R.

CONCLUSIONS

For the dissolution procedure, this article suggests a) hypothesis testing, with clear delineation of consumer and producer risks, b) testing by variables as opposed to testing by attributes (parametric vs. nonparametric testing), c) a tolerance interval approach to set acceptance criteria, and d) more flexible study designs. With these approaches, both consumer and producer (manufacturer) risks are clearly assessed, managed, and communicated.

REFERENCES

1. M.-L. Chen. Lipid-based oral dosage forms—regulatory perspective. *Am. Pharm. Rev*. **5**:30–35 (2002).
2. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. *Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products–General Considerations*, Office of Training and Communications, Division of Communications Management, Drug Information Branch, HFD-210, Rockville, MD, 2003.
3. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. *Guidance for Industry: Waiver of In Vivo Bioavailability and Bioequivalence Studies for Immediate-Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System*, Office of Training and Communications, Division of Communications Management, Drug Information Branch, HFD-210, Rockville, MD, 2000.
4. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. *Guidance for Industry: Extended-Release Oral Dosage Forms: Development, Evaluation, and Application of In Vitro/In Vivo Correlations*, Office of Training and Communications, Division of Communications Management, Drug Information Branch, HFD-210, Rockville, MD, 1997.
5. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. *Guidance for Industry: Dissolution Testing of Immediate-Release Solid Oral Dosage Forms*, Office of Training and Communications, Division of Communications Management, Drug Information Branch, HFD-210, Rockville, MD, 1997.
6. International Conference on Harmonization Q6A. *Guideline: Specifications for New Drug Substances and Products: Chemical Substances,* 1999.
7. United States Pharmacopeia, General Notices, 8.
8. United States Pharmacopeia, General Chapter. *Disintegration* (701).
9. United States Pharmacopeia, General Chapter. *Dissolution* (711).
10. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for Industry: *SUPAC-IR: Immediate-Release Solid Oral Dosage Forms: Scale-Up and Post-Approval Changes: Chemistry, Manufacturing, and Controls, In Vitro Dissolution Testing, and In Vivo Bioequivalence Documentation*, Office of Training and Communications, Division of Communications Management, Drug Information Branch, HFD-210, Rockville, MD, 1995.
11. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for Industry: *SUPAC-MR: Modified-Release Solid Oral Dosage Forms Scale-Up and Post-Approval Changes: Chemistry, Manufacturing, and Controls; In Vitro Dissolution Testing and In Vivo Bioequivalence Documentation*, Division of Communications Management, Drug Information Branch, HFD-210, Rockville, MD, 1997.
12. R. L. Williams and T. Foster. The future of dissolution testing: USP, academic, FDA, and Industry perspectives—dissolution: a continuing perspective. *Dissolution Technol.* **11**:6–7 (2004).
13. United States Pharmacopeia General Chapter. *Drug Release* (724).
14. N. Katori, N. Aoyagi, and S. Kojima. A proposal for revision of the content uniformity test and weight variation test. *Japanese Pharm. Forum* **4**:81–92 (1995).
15. H. Katori, N. Kaniwa, N. Aoyagi, and S. Kojima. A new acceptance sampling plan for the official dissolution test. *Japanese Pharm. Forum* **7**:166–173 (1998).

16. G. J. Hahn and W. Q. Meeker. *Statistical Intervals: A Guide for Practitioners*, New York, Wiley, 1991.
17. Y. Tsong, T. Hammerstrom, K. Lin, and T. E. Ong. Dissolution test acceptance sampling plans. *J. Biopharm. Stats.* **5**:171–183 (1995).
18. United States Pharmacopeia General Chapter. *In Vitro and In Vivo Evaluation* (1088).
19. R. J. Beckman and G. L. Tietjen. Two-sided tolerance limits for balanced random-effects ANOVA models. *Technometrics* **31**: 185–197 (1989).
20. R. W. Mee. β-expectation and β-content tolerance limits for balanced one-way ANOVA random model. *Technometrics* **26**:251–254 (1984).
21. M. G. Vangel. New methods for one-sided tolerance limits for a one-way balanced random-effects ANOVA model. *Technometrics* **34**:176–185 (1992).
22. C. T. Liao and H. K. Iyer. A tolerance interval for the normal distribution with several variance components. *Statist. Sinica* **14**: 217–229 (2004).
23. W. W. Hauck and R. Shaikh. Sample sizes for batch acceptance from single and multi- stage designs using two-sided normal tolerance intervals with specified content. *J. Biopharm. Stats.* **11**: 335–346 (2001).
24. R. L. Williams, W. P. Adams, G. Poochikian, and W. W. Hauck. Content uniformity and dose uniformity: current approaches, statistical analyses, and presentation of an alternative approach, with special reference to oral inhalation and nasal products. *Pharm. Res.* **19**:359–366 (2002).