*Pharmaceutical Research, Vol. 19, No. 4, April 2002 (© 2002)*

*Review Article*

# Content Uniformity and Dose Uniformity: Current Approaches, Statistical Analyses, and Presentation of an Alternative Approach, with Special Reference to Oral Inhalation and Nasal Drug Products

**Roger L. Williams,[1] Wallace P. Adams,[2] Guirag Poochikian,[3] and Walter W. Hauck[4,5]**

*Received January 3, 2002; accepted January 5, 2002*

This article reviews current and proposed approaches to content uniformity testing. In addition, the article proposes an approach that allows regulatory agencies and compendia to clearly state allowable consumer risk. Further, the article suggests that producers be allowed to control producer risk through selection of numbers of units and testing tiers. The approach facilitates risk communication to practitioners and patients/consumers, which is impeded with current approaches, and reduces regulatory and compendial burden.

**KEY WORDS:** inhalation drug products; nasal drug products; content uniformity; dose uniformity; tolerance intervals.

## INTRODUCTION

Content uniformity is one test (attribute) in a series of tests in a therapeutic product specification that assesses quality of a batch. Testing for content uniformity helps assure that the strength of a therapeutic product remains within specified acceptance limits. Recent national and international regulatory and compendial efforts have focused on harmonizing content uniformity testing, with several different approaches under consideration. The differing approaches arise from differing motivations (batch release vs. marketplace testing of a single unit), testing (uniformity of content vs. uniformity of mass), products covered (oral vs. oral inhalation and nasal drug products), and statistical strategies. This article considers the various content uniformity approaches to: 1) promote optimal science-based assessment of the risk that content uniformity is designed to address; 2) establish clear producer and consumer strategies to manage this risk; and 3) facilitate communication of the results of content uniformity risk assessment and management to patients and consumers in a clear and comprehensible way.

## BACKGROUND

### Characterization, Specifications, and Standards

Characterization studies conducted during product development assess safety, efficacy, and quality measures for a therapeutic product. For therapeutic products approved through a regulatory process, safety and efficacy characterization studies are reflected in approved product labeling. Quality characterization studies are performed in relation to safety and efficacy studies and are at times associated with specified acceptance criteria. For example, product quality bioavailability and bioequivalence studies are one-time product performance characterization studies that, in the case of bioequivalence, may be assessed using specified criteria and pre-determined pass/fail bioequivalence limits (1,2). Quality characterization studies yield specifications, defined as a list of tests, references to analytical procedures to evaluate those tests, and appropriate acceptance criteria (3). Separate specifications are usually needed for the drug substance and the drug product and may be needed for intermediates used in manufacturing, raw materials, reagents, container/closure systems, and container labeling. Through a public process, private specifications agreed to between a single manufacturer and the Food and Drug Administration (FDA) may become incorporated in compendial monographs in the *USP-NF*. These standards are legally enforceable through the adulteration and misbranding provisions of the Food, Drug, and Cosmetic Act and the Public Health Service Act. Testing to a FDA-approved specification focuses on samples from a batch to allow batch release and to assure, in Agency terms, that its identity, strength, quality, and purity are maintained during shelf-life (4). Conclusions are directed to the batch. Testing to a USP standard tends to focus on the specimen tested, to assure, in USP terms, its identity, strength, quality, purity, packaging, and labeling (5). Conclusions are directed to the specimen tested. Both FDA and USP sets of terms refer generally to *quality*, as used in documents harmonized in the International Conference on Harmonisation. *Content Uniformity* (*see USP 24*, ⟨905⟩) is one test in the drug product specification that is applicable to many types of dosage forms. Standard content uniformity specifications for oral inhalation and nasal drug products are provided in compendial and regulatory documents, such as those listed in Table I.

[1] U.S. Pharmacopeia, Rockville, Maryland 20852.
[2] Office of Pharmaceutical Science, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Rockville, Maryland 20857.
[3] Office of New Drug Chemistry, Office of Pharmaceutical Science, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Rockville, Maryland 20857.
[4] Biostatistics Section, Thomas Jefferson University, Philadelphia, Pennsylvania 19107.
[5] To whom correspondence should be addressed. (e-mail: w_hauck@mail.jci.tju.edu)



0724-8741/02/0400-0359/0 © 2002 Plenum Publishing Corporation

### Tolerance Interval Testing for Content Uniformity

Current approaches to content uniformity testing are based on either parametric tolerance intervals or a nonparametric procedure that can be recognized as a nonparametric tolerance interval. Three decisions are needed to assess content uniformity using a tolerance interval approach. These are: 1) the acceptable tolerance limits (e.g., 85–115% of label claim); 2) the minimum proportion, $p$, of the batch that should fall within the limits (e.g., 90% of units in a batch); and 3) the degree of confidence needed to reach an accept/reject decision (e.g., 95%). For nonparametric approaches, conformance is determined based on the number of assay values that fall within a specified accept/reject limit, irrespective of the actual values. With parametric tolerance intervals, an accept decision is reached if the test data expressed in terms of the criterion yield an observed tolerance interval that falls entirely within the tolerance limits. Parametric tolerance intervals provide simultaneous direct control on the mean and standard deviation of the batch (Figs. 1 and 2). For parametric tolerance interval testing, the general form of the criterion is $\overline{Y} \pm kS$, where $S$ is the observed standard deviation, and $k$ is a tolerance interval constant that accounts for sample size as well as the population fraction $p$ (6). For this application, $\overline{Y}$ is the difference of a test mean, $\overline{X}$, and a reference mean. The reference mean in content uniformity could be fixed as either the label claim or rubric mean, $M$, expressed as a percent. The rubric mean can sometimes be greater or less than 100% of label claim, with corresponding changes in tolerance limits. After subtracting the reference mean the tolerance limits are similarly adjusted, e.g., 85%–115% of label claim becomes ±15%. When the tolerance limits are symmetric about zero, as ±15%, the decision to accept the batch can be made using the largest absolute value from the interval; that is, reducing the tolerance interval to a single value, $|\overline{Y}| + kS$. This absolute value form is employed in the remainder of this paper to correspond to common practice.

Usually parametric tolerance interval approaches assume normal distribution of the data, possibly following transformation, such as to log scale. Nonparametric approaches do not assume normal distribution of test data. If the normality assumption is correct (and normality seems reasonable for content uniformity testing), the parametric tolerance interval approach ought to make better use of the data than approaches based on counts of values falling within specified intervals (7). All current content uniformity tests, whether non-parametric or parametric, use a two tier approach with fixed numbers of units allowed for testing in each tier. If results are sufficiently positive in the first stage (tier) of testing, then the study stops. If insufficiently positive, the study may continue to the second stage. In the language of clinical trials, the initial tiers of multiple-tier testing are interim analyses. When two tiers are used, the calculation of the actual type I (false positive) error rate for the full two-tier design needs to take into account both tiers. For example, a 5% test at each tier will have a combined type I error rate exceeding 5% (8). For nonparametric tolerance intervals, the method of Simon (9) can be adopted for two-tier designs. Hauck and Shaikh (10) developed a method for parametric tolerance intervals and designs of one or more tiers.

## CONTENT UNIFORMITY TESTS

### Nonparametric Content Uniformity Testing

At least nine draft or final regulatory and compendial documents consider non-parametric approaches to content uniformity testing. These include: *USP* ⟨905⟩ and ⟨601⟩, two draft FDA guidances, two *EP*[6] documents, draft and final CPMP documents, and a draft Pharmacopeial Discussion Group (PDG)[7] document based on text developed in an ICH/PDG task force (Table I). Each document uses a two-tier approach. When accept/reject criteria for the first tier are not met, second tier testing may be conducted. The term "safety net" is used to indicate that portion of a criterion that prohibits *any* dose from falling outside some interval, e.g., the 75–125% of label claim in the FDA criterion. The safety net is intended to reduce the likelihood that a unit in a batch will deviate substantially from label claim.

*USP* ⟨905⟩ provides acceptance criteria of 85–115%, with a safety net of 75–125% at both tiers for most dosage forms, including inhalation solutions and powders in pre-metered dosage units such as ampoules, blister packages, and capsules. For DPIs and topical MDIs, ⟨905⟩ applies to single doses from multiple inhalers, i.e., canisters for MDIs, and allows the wider acceptance criteria of ⟨601⟩. If the average of the potency limits specified in the individual monograph is less than or equal to label claim, these acceptance limits are expressed as a percent of label claim. If the average of the potency limits exceeds 100%, the limits are expressed as a percentage of the smaller of the sample average and the average of the potency limits, but not less than label claim. The draft PDG document expands ⟨905⟩ to include MDIs, DPIs, and metered dose sprays (Table I), applies to single doses from multiple inhalers, and allows the wider acceptance criteria of ⟨601⟩. In *USP* ⟨601⟩, a section entitled *Dose Uniformity over the Entire Contents* describes a test that assesses a single inhaler at tier 1 and two additional inhalers at tier 2 for a total of 3. This would not be suitable to assess the quality of a batch. This type of testing is motivated by data showing high variability in delivered dose within a single canister (11).

The draft FDA guidances for MDIs, DPIs and nasal sprays propose a content uniformity test for both approval and batch release based on single doses from each of multiple inhalers, similar to the draft PDG document, rather than multiple doses from a single inhaler. Compared to *USP* ⟨601⟩ for MDIs and DPIs (Table I), the FDA guidance narrows the acceptance criteria to 80–120% from 75–125% of label claim and narrows the safety net to 75–125% from 65–135% of label claim. Compared to the PDG document, the FDA guidance also narrows the acceptance criteria but does not include the PDG and ⟨905⟩ adjustments for asymmetric potency limits. USP ⟨905⟩ for inhalations in pre-metered dosage units and FDA drafts for MDIs, DPIs, and nasal sprays depart from fully nonparametric approaches. This is achieved by con-

[6] Content uniformity specifications in the *British Pharmacopoeia (BP)* 2000 for MDIs and DPIs for oral inhalation, and for MDIs and sprays for nasal dosing, are identical to those of the *European Pharmacopoeia (EP)* Third Edition Supplement 1999 and EP Third Edition, 1997, respectively. Therefore, when the *EP* is referred to in this paper, it is understood to refer to the *BP* as well.

[7] The Pharmacopoeial Discussion Group is composed of representatives of the *European Pharmacopoeia,* the *Japanese Pharmacopoeia,* and the *United States Pharmacopeia.*

**Table I.** Comparison of Nonparametric Content Uniformity Criteria

| | *USP* <905> Inhalations in Premetered Dosage Units | *USP* <905> DPIs, Topical MDIs | *USP* <601>[a] MDIs, DPIs | FDA (draft)[b] MDIs, DPIs, Nasal Sprays |
|---|---|---|---|---|
| Unit for determination | Single Unit Container | Dose | Dose | Dose |
| # of units sampled for container | — | 1 | 10 | 1 |
| 1st Tier, # containers | 10 | 10 | 1 | 10 |
| 1st Tier, total # determinations | 10 | 10 | 10 | 10 |
| Accept after 1st tier if: | ≤ 1 outside 85–115% and 0 outside 75–125% of larger of rubric mean[c] and label claim | ≤ 1 outside 75–125% and 0 outside 65–135% of larger of rubric mean[c] and label claim | ≤ 1 outside 75–125% and 0 outside 65–135% of label claim | ≤ 1 outside 80–120% and 0 outside 75–125% of label claim |
| 2nd Tier; # *additional* containers tested | 20 | 20 | 2 | 20 |
| 2nd Tier; *total* # determinations both tiers | 30 | 30 | 30 | 30 |
| Accept after 2nd tier if: | ≤ 3 outside 85–115% and 0 outside 75–125% of larger of rubric mean[c] and label claim | ≤ 3 outside 75–125% and 0 outside 65–135% of larger of rubric mean[c] and label claim | ≤ 3 outside 75–125% and 0 outside 65–135% of label claim | ≤ 3 outside 80–120% and 0 outside 75–125% of label claim |
| Additional | ≤6.0% CV at Tier 1, ≤7.8% CV at Tier 2 | | | Sample mean within 85–115% of label claim at each tier |
| Reference(s) | 22 | 21 | 23 | 24, 25 |

**Table I.** Continued

| | *EP* and BP Oral inhalation MDIs, DPIs | *EP* and BP Nasal MDIs, Sprays (Suspension, emulsion) | CPMP (draft) Oral inhalation MDIs[d] | CPMP DPIs[e] | PDG (draft) MDIs, DPIs, Metered Dose Sprays |
|---|---|---|---|---|---|
| Unit for determination | Dose | Dose | Dose | Dose | Dose |
| # of units sampled per container | 10 | 1 | 10 | 10 | 1 |
| 1st Tier, # containers | 1 | 10 | 1 | 1 | 10 |
| 1st Tier, total # determinations | 10 | 10 | 10 | 10 | 10 |
| Accept after 1st tier if: | ≤ 1 outside 75–125% and 0 outside 65–135% of sample mean | ≤ 1 outside 75–125% and 0 outside 65–135% of sample mean | ≤ 1 outside 75–125% and 0 outside 65–135% of sample mean | ≤ 1 outside 75–125% and 0 outside 65–135% of sample mean | ≤ 1 outside 75–125% and 0 outside 65–135% of larger of rubric mean[c] and label claim |
| 2nd Tier; # *additional* containers tested | 2 | 20 | 2 | 2 | 20 |
| 2nd Tier; *total* # determinations both tiers | 30 | 30 | 30 | 30 | 30 |
| Accept after 2nd tier if: | ≤ 3 outside 75–125% and 0 outside 65–135% of sample mean | ≤ 3 outside 75–125% and 0 outside 65–135% of sample mean | ≤ 3 outside 75%–125% and 0 outside 65–135% of sample mean | ≤ 3 outside 75–125% and 0 outside 65–135% of sample mean | ≤ 3 outside 75–125% and 0 outside 65–135% of larger of rubric mean[c] and label claim |
| Additional | Above tests: For MDIs, unless otherwise justified and authorized. For DPIs, ranges may extend to 50–150% of sample mean if justified and authorized | Above tests: Unless otherwise justified and authorized | Unless otherwise justified and authorized. Sample mean within 85–115% of label claim if label claim is expressed as dose ex-actuator. Wider limits are accepted if label claim is expressed as metered dose | Sample mean within 80–120% of label claim per actuation should be possible | |
| Reference(s) | 26, 27 | 28, 29 | 30 | 31 | 32 |

[a] *Dose Uniformity over the Entire Contents.*

[b] *Dose Content Uniformity* and *Spray Content Uniformity* are listed in the table. Similar tests, entitled *Dose Content Uniformity Through Container Life* (for MDIs and device-metered DPIs) and *Spray Content Uniformity Through Container Life* (for nasal sprays), are also performed to assure content uniformity at beginning, middle, and end of label claim doses, or beginning and end of label claim doses, respectively. In addition, the acceptance criteria include a requirement that the sample means at each of beginning, middle and end, or beginning and end, respectively, are within 85–115% of label claim at each tier. This is to limit the magnitude of change in delivery over container life, which an overall mean criterion may not detect. For additional details, see the cited references.

[c] See text for the case where sample mean is between label claim and rubric mean.

[d] An additional test is also performed to evaluate dose ex-actuator including beginning and end of label claim doses.

[e] For device-metered units, dose delivery over the nominal content of the inhaler should also be presented.

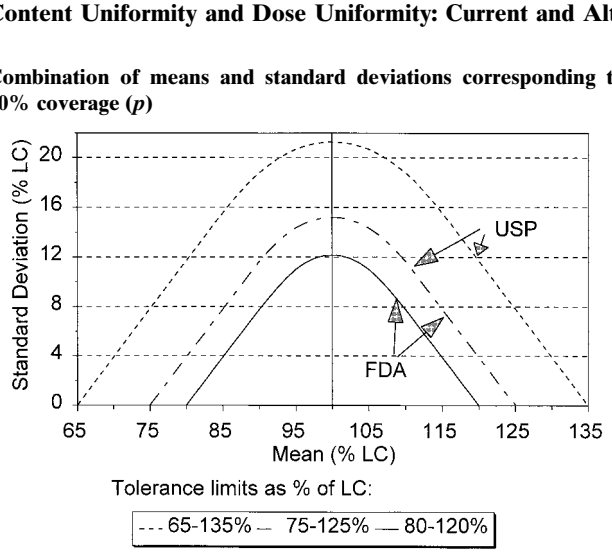


**Fig. 1.** For all combinations of means and standard deviations on or below each curve, at least 90% of the distribution falls within the tolerance limits. The upper curve is for wide tolerance limits, 65–135% of label claim, and thus allows more combinations of means and variances. The lowest curve, with narrow specified tolerance limits of 80–120%, is more restrictive.

straining %CV at each tier for USP ⟨905⟩ and by requiring at each tier that the sample mean remain within 85–115% of label claim for the two FDA documents.

The *EP* requirement for oral inhalation MDIs and DPIs is similar to *USP* ⟨601⟩ in requiring testing of multiple doses from a single inhaler. The *EP* requirement for nasal MDIs and sprays is similar to the draft PDG document and the FDA guidance in that it requires testing of a single dose from multiple inhalers. The draft CPMP MDI guidance uses the *EP* test for the oral inhalation MDIs. The CPMP DPI guidance is similar to *USP* ⟨601⟩ in requiring testing of multiple doses from a single inhaler. Also, like the FDA, the two CPMP guidances include an exception to a fully nonparametric approach by requiring, for MDIs, and recommending, for DPIs, that test means fall within a specified acceptance limit. Unlike

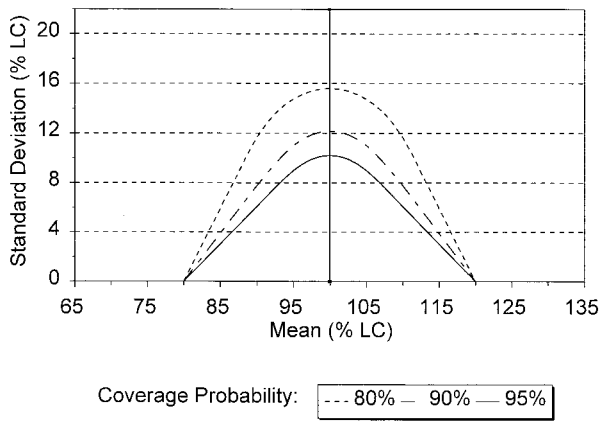


**Fig. 2.** For all combinations of means and standard deviations on or below each curve, at least a specified portion of the distribution falls within the tolerance limits of 80–120%. The lowest curve is a high coverage probability of 95% and thus allows fewer combinations of means and variances. The upper curve, with lower coverage probability is less restrictive.

the other documents summarized in Table I, the *EP*, *BP*, and CPMP requirements for individual doses are expressed as a percentage of sample mean rather than of label claim.

The allowance of a second tier of testing in these criteria for all nonparametric approaches summarized in Table I may do little to increase the probability of a batch passing. For criteria using 100% of label claim as the center of the acceptable tolerance limits, exact probabilities of a batch passing the two-tier component after the first and after the second tier are shown in Table II. These calculations do not take into account the additional criterion on the sample mean in the FDA draft. The probability of acceptance decreases as the batch mean deviates from label claim and as the standard deviation increases. Comparing the probabilities of accepting after 10 units tested to those for accepting after 30, those for after 30 are only marginally larger. The second tier testing using either of these criteria thus adds little to a sponsor's likelihood of passing. The probability calculations are sensible if one considers what it takes to not pass after 10 units tested but then pass after 30. There are two possibilities, both with the requirement of no units outside the outer limits (75–125% for the FDA): two outside the inner limits (80–120% for the FDA) of the first 10 and no more than one outside of the next 20; or three outside of the first 10 and none outside of the next 20.

Both the FDA and CPMP criteria, in total, are structured to limit both the variability and the deviation of the mean from label claim in a batch. They do this in different ways. In the CPMP criterion, the component applied at the end of each tier limits how many actuations may differ by more than 25% from the sample mean. This is solely a limit on variability, given that the sample mean may be any value. The CPMP uses another component of their criterion (Table I) to limit how far the sample mean may differ from label claim. In contrast, the FDA criterion, which constrains deviations from label claim at each tier, simultaneously places limits on both the variability and mean. The same is true of the *USP* criterion. While placing a narrower limit on the mean deviation from label claim may appear redundant, these approaches add some value in that they preclude unusual cases. For example, test data that are identically 120% of label claim would pass if the mean value were not constrained to 85–115% of label claim. The *USP* ⟨905⟩ criterion also places limits on the

**Table II.** Probability of Accepting Batch with Different Values of Mean and Standard Deviation (SD) between-Container of the Batch

| | | FDA Draft Criterion | | USP <905>[b] and PDG[c] Criteria | |
|---|---|---|---|---|---|
| Mean (% of LC[a]) | SD (% of LC) | Accept After 10 | Accept After 10 or 30 | Accept After 10 | Accept After 10 or 30 |
| 100 | 5 | >.999 | >.999 | >.999 | >.999 |
| | 10 | .824 | .847 | .979 | .984 |
| | 15 | .261 | .265 | .643 | .688 |
| 110 | 5 | .951 | .967 | .999 | .999 |
| | 10 | .349 | .359 | .793 | .852 |
| | 15 | .093 | .093 | .362 | .378 |

[a] LC (label claim).
[b] Criterion for DPI's and topical MDI's for the case where the rubric mean equals the label claim.
[c] For the case where the rubric mean equals the label claim.

%CV, which places an additional restriction on the variability. In contrast to the *USP*, FDA and CPMP criteria, the *EP* criterion limits variability through restrictions around the sample mean but not label claim. With this approach, test results could show substantial deviation from label claim.

### Parametric Content Uniformity Testing

In the last several years, several documents have developed parametric approaches to content uniformity testing. These are a final *Japanese Pharmacopoeia XIII* document, industry documents, and the draft PDG document (Table III). This approach to content uniformity testing was considered in a 1994 article (12) and subsequently discussed in the *Japanese Pharmacopoeial (JP) Forum* (13). The *JP Forum* proposal uses a standard tolerance interval criterion, $|M-\overline{X}| + kS$, where $M$ is label claim and $\overline{X}$ and $S$ are the mean and standard deviation based on Tier 1 and Tier 2 batch test data. For Tier 1, the tolerance interval constant, $k$, is 2.2; for tier 2, the $k$ is 1.9. The *JP Forum* proposed that the tolerance limit be ±15% of $M$. The *JP Forum* approach is termed a *test by variables* and the corresponding nonparametric approach is termed a *test by attributes*. The *JP Forum* approach became official in *JP XIII* (14). *USP* published the Japanese approach in a *Stimuli* article (15) and in *Pharmacopeial Previews* (16). Based on the Japanese work, the parametric approach to content uniformity was taken up in PDG and then in a joint ICH and PDG Task Force. In support of the Task Force, the Statistics Working Group of the Pharmaceutical Research and Manufacturers of America (PhRMA) proposed modifications to the Japanese approach that would align the producer and consumer risks of the approach more closely to those of the current *USP* method (17–19).

In the PhRMA approach, $M$ is replaced by $M'$, where $M'$ is defined as follows:

$$M' = \max\{0.965\ \text{LC}, \overline{X}\} \text{ if } \overline{X} \leq \text{LC}$$
$$M' = \min\{U, \overline{X}\} \text{ if } \overline{X} > \text{LC}$$

where $U = \max\{1.035\ \text{LC}, R\}$ and $R$ is defined as the average of the shelf limits specified in the potency definition for the drug product monograph, i.e., the "rubric mean.

On the lower bound, the use of $M'$ allows the value of $M$ to range from 96.5 to 100% of label claim depending on the value of $\overline{X}$. The approach creates a window within the tolerance limits that yields a value of 0 for the absolute value of $M-\overline{X}$ when $\overline{X}$ is within the window. On the upper bound, an additional window is created that expands based on the rubric mean. The PhRMA approach also increases the tolerance interval constant to 2.4 at Tier 1, which increases the likelihood of failure depending on the magnitude of sampling error (variance). The Tier 2 PhRMA approach is identical to that of the Japanese approach. The ICH/PDG Task Force agreed to a modification of the PhRMA approach, as follows. First, the window allowed in the PhRMA proposal was narrowed to 98.5 to 101.5%, thus creating a new $M''$:

$$M'' = \max\{0.985\ \text{LC}, \overline{X}\} \text{ if } \overline{X} \leq \text{LC}$$
$$M'' = \min\{U'', \overline{X}\} \text{ if } \overline{X} > \text{LC}$$

where $U'' = \max\{1.015\ \text{LC}, R''\}$. Second, $R$ was redefined as $R''$ to be the target test sample amount at the time of manufacture, which is at least 100%. With this redefinition, the upper bound window is reduced.

Based on the work of the ICH/PDG Task Force, PDG, with *USP* as the lead pharmacopeia, has published a draft Stage 4 proposal in the May-June 2001 *Pharmacopeial Forum*. This document proposes applying the parametric approach to inhalations (powders or solutions) in pre-metered dosage units (e.g., ampoules, capsules, and blister packages), but not to MDIs, DPIs, or metered dose sprays. If successful, this document will become harmonized in the three PDG pharmacopeias.

## DISCUSSION

The risk assessment strategy underlying content uniformity testing is the assumption that some pre-specified limits

**Table III.** Comparison of Parametric Content Uniformity Criteria[a]

| | JP Applicable dosage forms | PhRMA Inhalations | PDG (draft) Inhalation Solutions, Powders in Premetered Dosage Units; Inhalations in Single-Unit Containers |
|---|---|---|---|
| Unit for determination | Not specified | Dose | Single Unit Container |
| # of units sampled per container | Not specified | Not specified | —[b] |
| 1st Tier, # containers | Not specified | Not specified | 10 |
| 1st Tier, total # determinations | 10 | 10 | 10 |
| Accepted after 1st tier if: | $\lvert M - \overline{X}\rvert + 2.2S$ contained with ±15% | $\lvert M' - \overline{X}\rvert + 2.4S$ contained within ±15% and 0 outside 75%–125% of $M'$ | $\lvert M'' - X\rvert + 2.4S$ contained within ±15% |
| 2nd Tier; # *additional* containers | Not specified | Not specified | 20 |
| 2nd Tier; *total* # determinations both tiers | 30 | 30 | 30 |
| Accepted after 2nd tier if: | $\lvert M - \overline{X}\rvert + 1.9S$ contained within ±15% and 0 outside 75%–125% of label claim | $\lvert M' - \overline{X}\rvert + 1.9S$ contained within ±15% and 0 outside 75%–125% of $M'$ | $\lvert M'' - \overline{X}\rvert + 2.0S$ contained within ±15% and 0 outside 75%–125% of $M'$ |
| Reference(s) | 14 | 18, 19 | 32 |

[a] See text for definitions of $M$, $M'$, and $M''$ for JP, PhRMA, and PDG criteria, respectively. Sample mean, $\overline{X}$, and the sample standard deviation, $S$, are expressed as % of label claim.

[b] The unit sampled reflects the contents of the entire container.

exist where safety and efficacy outcomes may change if content uniformity fails. To assess this risk, the PDG parametric tolerance interval approach relies on statistical hypothesis testing to determine whether the results of a content uniformity test, expressed as an observed tolerance interval, fall within the tolerance limit at a certain level of confidence. Content uniformity testing has the character of equivalence testing. In bioequivalence testing, assay results for both test and reference products are compared. In parametric content uniformity testing, test assay results, $\overline{X}$, are compared to a fixed value $M$.

Current and proposed approaches to content uniformity testing listed in Table I obscure risk communication. They are also limiting to pharmaceutical manufacturers. First, they specify the number of units that may be tested at each testing tier. Second, they limit the number of testing tiers. Third, they are discordant between agencies and pharmacopeias. A more appropriate approach perhaps would be for regulatory agencies and pharmacopeias to agree on what is an acceptable batch or unit (the alternative hypothesis in statistical language) and allowable consumer risk. Using a tolerance interval approach, this would be accomplished by setting: 1) the tolerance limits; 2) the minimum proportion, $p$, of the batch that should fall within the limits; and 3) the degree of confidence needed to reach an accept/reject decision. According to this approach, regulatory and/or compendial documents might, for example, indicate: at least 90% of the batch should fall within ±15% of label claim at a consumer risk not greater than 5%. The manufacturer then becomes responsible for determining the number of units tested and number of testing tiers to provide assurance that the specified level of consumer risk is achieved, testing in such a way so as control producer risk. As in other sample size calculations, the power (here, the probability of the batch passing) of the content uniformity test outcome would depend on the proportion of the batch the manufacturer would be willing to assume, based on manufacturing experience, actually falls within the tolerance limits. For example, the sponsor could seek a producer risk of no more than 10% if at least 95% of the batch actually falls within the tolerance limit. The number of samples to be tested at different tiers would depend on that assumption and the desired level of producer risk. A producer with a product performing well within a specified tolerance limit will require a smaller sample than one whose batch performance, either in terms of mean and/or variance, is closer to the tolerance limits. Some examples for a nonparametric and a parametric approach are shown in Table IV. The sample size of either approach depends on how good a batch the sponsor is willing to assume. The greater the proportion of the batch that is assumed within the target interval, the smaller the sample size required. For example, if the target coverage probability is set at 85% and the sponsor believes that at least 95% of the batch actually falls within the acceptable tolerance limits, the two-tier nonparametric design would use 56 units at the first tier and 41 at the second if needed (97 total) to control the producer risk at no more than 5%. In contrast, if the sponsor believed 98% of the batch falls within the tolerance limits, the sample size would fall to 24 at the first tier and 18 at the second (42 total). Comparing the total sample sizes from the nonparametric and parametric approaches shows that the

**Table IV.** Some Designs with 5% Consumer Risk[a]

| Target Coverage Probability (%) | Actual % Inside Acceptable Limits | Producer Risk (%) | Nonparametric Approach[b] | | | | Parametric Approach[c] | |
|---|---|---|---|---|---|---|---|---|
| | | | First Tier | | Second Tier | | | |
| | | | $N_1$ | Max. Outside to Accept | $N_1 + N_2$ | Max. Outside to accept | $N_1$ | $N_1 + N_2$ |
| 80 | 95 | 10 | 23 | 1 | 51 | 4 | 14 | 42 |
| | | 5 | 24 | 1 | 55 | 5 | 18 | 54 |
| 85 | 95 | 10 | 45 | 2 | 80 | 6 | 31 | 93 |
| | | 5 | 56 | 3 | 97 | 8 | 40 | 120 |
| 90 | 95 | 5 and 10 | | | >100 | | | |
| 80 | 98 | 10 | 15 | 0 | 25 | 1 | 6 | 18 |
| | | 5 | 15 | 0 | 34 | 2 | 8 | 24 |
| 85 | 98 | 10 | 20 | 0 | 48 | 2 | 10 | 30 |
| | | 5 | 24 | 0 | 42 | 2 | 12 | 36 |
| 90 | 98 | 10 | 33 | 0 | 84 | 3 | 21 | 63 |
| | | 5 | 50 | 1 | 99 | 4 | 27 | 81 |
| 95 | 98 | 5 and 10 | | | >100 | | | |
| 60 | 91 | 10 | 10 | 1 | 22 | 3 | | |

[a] For these calculations, the consumer risk was set at 5% (the probability of falsely *passing* a batch if the actual percent within the tolerance limits is as given in the first column) and the producer risk at 10% or 5% (the probability of falsely *failing* the batch if the actual percent within the tolerance limits is as given in the second column).

[b] Sample sizes for the nonparametric approach are based on the method of Simon (9) for two-tier designs with minimum expected sample size adopted to use with nonparametric tolerance intervals. A "> 100" indicates that a design with these properties would have a total sample size greater than 100. The ">" rows are included to show that the sample size would need to be very large for these cases. Sample sizes for the nonparametric approach do not depend on the actual values of the acceptable tolerance limits.

[c] Sample sizes for the parametric approach use the method of Hauck and Shaikh. In these designs, the first tier is one-third of the total sample size. Specification for the parametric approach requires identifying the tolerance limits, and the mean and standard deviation of the batch at which to determine the producer risk. Here, the tolerance limits were fixed at (80-120%), the mean was fixed at 102%, and the standard deviation chosen to have the desired percentage of the batch within the tolerance limits.

parametric approach will normally require the smaller sample size. The exact comparison depends on the particular nonparametric and parametric approaches used and what the sponsor assumes regarding the batch. The last row of Table IV is included to provide an example of a combination of target coverage probability, 60%, and assumed batch property, 91%, that would lead to a two-tier design similar to that currently specified by the FDA. The approach is applicable in other settings where content uniformity is of interest, e.g., solid oral-dosage forms and blend uniformity testing (20).

Allowing the producer to choose a sample size according to their desired risk level conflicts with the safety nets that are part of the current criteria. Because the safety nets allow no units outside some limits, the larger the sample, the more likely a batch would fail by chance.

A further way to reduce producer risk might be drawn from experience with bioequivalence testing. The parametric tolerance interval approach for content uniformity testing compares test results to a fixed mean with a criterion that does not depend on the nature of the product; i.e., a one size fits all approach. In contrast, certain FDA criteria for bioequivalence comparisons (21) allow scaling of the bioequivalence limits to reference variability. What would take the place of the reference in content uniformity testing? Perhaps the clinical trial batch has a role in this regard. Tolerance limits at the time of marketing might be widened or narrowed based on variability of this batch, providing it meets minimum dose content uniformity standards. Tolerance limits might also be set depending on whether the active ingredient is or is not a narrow therapeutic range drug. Another approach might be to set the tolerance limits based on a better understanding of population and individual dose/response curves for efficacy and toxicity.

## ACKNOWLEDGMENTS

This work was supported in part by a contract from the US Food and Drug Administration to Jefferson Medical College, Philadelphia, PA. This article represents the personal opinions of the authors and does not necessarily represent the views or policies of the U.S. Pharmacopeia or the U.S. Food and Drug Administration.

## REFERENCES

1. R. L. Williams, W. Adams, M.-L. Chen, D. Hare, A. Hussain, L. Lesko, R. Patnaik, V. Shah, and FDA Biopharmaceutics Coordinating Committee. Where are we now and where do we go next in terms of the scientific basis for regulation on bioavailability and bioequivalence? *Eur. J. Drug Metab. Pharmacokinet.* **25**:7–12 (2000).
2. U.S. Department of Health and Human Services (DHHS). Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER). October 2000, *Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products – General Considerations*. (Internet) http://www.fda.gov/cder/guidance/index.htm.
3. International Conference on Harmonisation. Draft Guidance on Specifications: Test procedures and acceptance criteria for new drug substances and new drug products: Chemical substances. *Fed. Regist.* **62**:62890–62910 (1997).
4. Section 505(d)(3) of the Federal Food, Drug, and Cosmetic Act, as amended February 1998 (21 U.S.C. 355(d)(3)).
5. General Notices and Requirements, *U.S. Pharmacopeia*, 24th rev.; United States Pharmacopeial Convention, pp. 1–14.
6. A. Weissberg and G. H. Beatty. Tables for tolerance-limit factors for normal distributions. *Technometrics* **2**:483-500 (1960); erratum. *Technometrics* **3**:576–577 (1961).
7. J. D. Esinhart and V. M. Chinchilli. Sample size considerations for assessing individual bioequivalence based on the method of tolerance intervals. *Int. J. Clin. Pharmacol. Ther* **32**:26–32 (1994).
8. K. McPherson. Statistics: The problem of examining accumulating data more than once. *N. Engl. J. Med.* **290**:501–502 (1974).
9. R. Simon. Optimal two-stage designs for Phase II clinical trials. *Control. Clin. Trials* **10**:1–10 (1989).
10. 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. Stat.* In press.
11. T. D. Cyr, S. J. Graham, K. Y. R. Li, and E. G. Lovering. Low first-spray drug content in albuterol metered-dose inhalers. *Pharm. Res.* **8**:658–660 (1991).
12. N. Katori. Statistical considerations for weight variation and content uniformity tests. *Eisei Shikenjo Hokoku* **112**:200–201 (1994).
13. 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).
14. General Tests. 10, Content Uniformity Test, *The Japanese Pharmacopoeia,* XIII (English version). The Society of Japanese Pharmacopoeia, Tokyo, 1 April 1996, p. 25.
15. N. Katori, N. Aoyagi, and S. Kojima. A proposal for revision of the content uniformity test and weight variation test. Stimuli to the revision process. *Pharm. Forum* **23**:5325–5333 (1997).
16. Pharmacopoeial Previews. ⟨905⟩ Uniformity of dosage units. *Pharm. Forum* **23**:4683–4686 (1997).
17. Pharmaceutical Research and Manufacturers of America. (PhRMA) Statistics Working Group. Content uniformity-Evaluation of the USP Pharmacopeial Preview. *Pharm. Forum* **24**:7029–7044 (1998).
18. PhRMA Statistics Working Group. Content uniformity-Alternative to the USP Pharmacopeial Preview. *Pharm. Forum* **25**:7939–7948 (1999).
19. PhRMA Statistics Working Group. Recommendations for a globally harmonized uniformity of dosage units test. *Pharm. Forum* **25**:8609–8624 (1999).
20. J. K. Prescott and T. P. Garcia. A solid dosage and blend content uniformity troubleshooting diagram. *Pharm. Technol.* **25**:68–88 (2001).
21. DHHS. FDA, CDER. January 2001, *Guidance for Industry: Statistical Approaches to Establishing Bioequivalence*. (Internet) http://www.fda.gov/cder/guidance/index.htm.
22. ⟨905⟩, Uniformity of dosage units, Content uniformity, *U.S. Pharmacopeia*, 24th rev., Suppl. 1, United States Pharmacopeial Convention, Rockville, Maryland, 2000, pp. 2702–2704.
23. ⟨601⟩, Aerosols, Metered-dose inhalers, and dry powder inhalers, Dose uniformity over the entire contents, *U.S. Pharmacopeia,* 24th rev., Suppl. 1, United States Pharmacopeial Convention, Rockville, Maryland, 2000, pp. 2674–2688.
24. DHHS. FDA, CDER. October 1998, *Guidance for Industry: Metered Dose Inhaler (MDI) and Dry Powder Inhaler (DPI) Drug Products, Chemistry, Manufacturing, and Controls Documentation.* (Internet) http://www.fda.gov/cder/guidance/index.htm.
25. DHHS. FDA, CDER. May 1999, *Guidance for Industry: Nasal Spray and Inhalation Solution, Suspension, and Spray Drug Products, Chemistry, Manufacturing, and Controls Documentation.* (Internet) http://www.fda.gov/cder/guidance/index.htm.
26. Preparations for Inhalation/Inhalanda, *European Pharmacopoeia*, 3rd ed., Suppl. 1999, Council of Europe, Strasbourg, pp. 984–989.
27. Preparations for Inhalation. *British Pharmacopoeia* 2000, Vol. II, The Stationery Office, London, pp. 1668–1673.
28. Nasal Preparations/Nasalia, *European Pharmacopoeia*, 3rd ed., 1997, Council of Europe, Strasbourg, pp. 1763-1765.
29. Nasal Preparations. *British Pharmacopoeia* 2000, Vol. II, The Stationery Office, London, pp. 1674–1676.
30. Committee for Proprietary Medicinal Products (CPMP), European Agency for the Evaluation of Medicinal Products, London, 16 November 2000, *Note for Guidance on Requirements for Pharmaceutical Documentation for Pressurised Metered Dose Inhalation Products*, CPMP/QWP/2845/00 (Draft Guideline).
31. CPMP, European Agency for the Evaluation of Medicinal Products, London, December 1998, *Note for Guidance on Dry Powder Inhalers* (Adopted Guideline), CPMP/QWP/158/96.
32. ⟨905⟩, Uniformity of dosage units, Harmonization. *Pharm. Forum* **27**:2615–2619 (2001).