Product Quality Research Institute Evaluation of Cascade Impactor Profiles of Pharmaceutical Aerosols: Part 2 - Evaluation of a Method for Determining Equivalence

Submitted: May 3, 2006; Accepted: October 13, 2006; Published: January 19, 2007

David Christopher,[1] Wallace P. Adams,[2] Douglas S. Lee,[3] Beth Morgan,[4] Ziqing Pan,[5] Gur Jai Pal Singh,[2,6] Yi Tsong,[7] and Svetlana Lyapustina[8]

[1]Statistics, Schering-Plough Research Institute, Kenilworth, NJ
[2]Office of Generic Drugs, Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA), Rockville, MD
[3]Nonclinical Statistics, Pfizer Global R&D, Groton, CT
[4]Manufacturing and Supply, GlaxoSmithKline, Zebulon, NC
[5]Statistics, Schering-Plough Research Institute, Kenilworth, NJ
[6]Current address: Watson Laboratories Inc, Corona, CA
[7]Quantitative Methods Research Staff, CDER FDA, Rockville, MD
[8]Drinker, Biddle and Reath LLP, Washington, DC

ABSTRACT

The purpose of this article is to present the thought process, methods, and interim results of a PQRI Working Group, which was charged with evaluating the chi-square ratio test as a potential method for determining in vitro equivalence of aerodynamic particle size distribution (APSD) profiles obtained from cascade impactor measurements. Because this test was designed with the intention of being used as a tool in regulatory review of drug applications, the capability of the test to detect differences in APSD profiles correctly and consistently was evaluated in a systematic way across a designed space of possible profiles. To establish a "base line," properties of the test in the simplest case of pairs of identical profiles were studied. Next, the test's performance was studied with pairs of profiles, where some difference was simulated in a systematic way on a single deposition site using realistic product profiles. The results obtained in these studies, which are presented in detail here, suggest that the chi-square ratio test in itself is not sufficient to determine equivalence of particle size distributions. This article, therefore, introduces the proposal to combine the chi-square ratio test with a test for impactor-sized mass based on Population Bioequivalence and describes methods for evaluating discrimination capabilities of the combined test. The approaches and results described in this article elucidate some of the capabilities and limitations of the original chi-square ratio test and provide rationale for development of additional tests capable of comparing APSD profiles of pharmaceutical aerosols.

Corresponding Author: Svetlana Lyapustina, Drinker, Biddle and Reath LLP, 1301 K Street NW, Suite 900 East Tower, Washington DC 20005-3317. Tel: (202) 230-5179; Fax: (202) 230-5379; E-mail: svetlana.lyapustina@dbr.com

KEYWORDS: Chi-square ratio, bioequivalence, cascade impactor, particle size distribution.

INTRODUCTION

The potential importance of the aerodynamic particle size distribution (APSD) to the performance, safety, and efficacy of orally inhaled and nasal drug products (OINDP) is generally recognized; and therefore a comparison of APSD profiles obtained with test and a reference products was recommended as one of the important in vitro tests in the 1999 Food and Drug Administration (FDA) Guidance[1] regarding bioequivalence studies. A statistical test based on the chi-square ratios was also recommended. An earlier study[2] on this subject provided a background and description of that test and outlined the process for investigating properties of the test in a collaborative process involving scientists from the FDA, industry, academia, and the United States Pharmacopeia (USP), through the Product Quality Research Institute (PQRI). In essence, the Working Group has studied the ability of the chi-square ratio test to "make correct decisions" (ie, correctly declare either equivalency or lack of demonstrated equivalency) when applied to a wide variety of APSD profiles that might be encountered during review of actual drug applications.

The interim findings of the PQRI Working Group are summarized in this report, as required by the approved Work Plan.[3] The designation "interim" is being used to distinguish this report from the final submission to the FDA, which will mark the official completion of the group's investigations and the subsequent "sun-setting" of the Working Group per the PQRI process.

This article describes the PQRI studies on this topic in the order in which they occurred. The article maintains this ordering because each step builds upon the prior work. At each

step, an appropriate method had to be developed based on the results of the previous step and taking into account the ultimate objectives for the test, which are explained in detail in the background report.[2] The statistical work began with (1) translating the chi-square ratio test's description in the published guidance into an executable, programmable algorithm, and (2) the development of simulation methods capable of modeling any number of realistic APSD profiles based on real product data. Using these tools, initial studies of the chi-square ratio test focused on pairs of identical-profiles, as described in the "Methodology" section. Knowledge gained from that study was used during the next step, which focused on pairs of profiles differing in a specified, systematic way on a single deposition site. After that study, an important concept of "target" profiles was developed, which underpins the methodology of assessing future tests' performance. The "Results and Discussion" section presents the results obtained in this phase of work and describes the approaches proposed for the future work.

METHODOLOGY

Identical-Profiles Studies

In order to understand the properties of the chi-square ratio test and its ability to make correct decisions consistently under a wide variety of possible situations, the Working Group undertook a systematic investigation of the test's properties. To start with, the general description of the test was translated into an executable algorithm.[4-6] In parallel, a method was created for simulating realistic profiles (based on real products' profiles) such that the interstage correlations were preserved.[7] These statistical tools were used to set the stage for the main work assigned to the Working Group—a study of the test's response to a variety of possible situations that might be encountered in real life (including a variety of APSD profile types and a variety of patterns of changes in those profiles). The basic real-life profiles that seemed to be representative of metered-dose inhalers (MDIs) and dry-powder inhalers (DPIs) were obtained from an industry database containing several real world MDI and DPI profiles. These data provided information about mean deposition and variance of deposition on each cascade impactor (CI) site plus interrelationships among the sites. Simulated profiles were then generated using Monte-Carlo methods.

The chi-square ratio test was developed to assess equivalence between profiles, as opposed to detecting differences. As such, the test has the best ability (ie, statistical power) to correctly declare equivalence when the profiles being compared are identical. Therefore, the first case to be examined was where the Test and Reference profiles were *identical,* meaning that the same set of data was used as both Test and Reference (ie, same data set, different labels).

Here, as in the rest of the article, the term "profile" refers to a set of data from 30 CI runs—1 each from 30 product containers, as directed by the test described in the draft guidance. Overall, a profile is characterized by the average deposition and variability on each CI site and accessory.

For the chi-square ratio test, all data for each CI run are normalized by dividing the absolute deposition on each site by the total amount captured on all sites, also following the draft guidance.

When the identical-profiles study was initiated, work was still being done to develop statistical models to produce simulated data that preserved intersite correlations. Therefore, the simulated profiles for the identical-profiles study were generated with no intersite correlation structure (ie, the distribution for any deposition site was independent of all other deposition sites). Furthermore, it was assumed that deposition on each site is normally distributed, which is a reasonable assumption for the purpose of the identical-profiles study because real data suggest that depositions on each site typically follow an approximately normal distribution. Occasional outlier results may occur but they are not likely to dramatically affect the general behavior of the test.

The chi-square ratio is inherently invariant under additive translations, and this realization made it possible to systematize the possible scenarios to study. The above-mentioned invariance means that the value of the chi-square ratio is independent of the order of the sites in a profile, and consequently it is sufficient to study *rank-ordered* profiles (rank-ordered by the expected normalized deposition on sites) rather than the actual profiles, where depositions are displayed in the "natural" order of the sites as they occur in a cascade impactor. An example of a profile used in the identical-profiles study is presented in Figure 1, where deposition sites (eg, throat or induction port, stages 0-7, filter) are presented in the "natural" order. Figure 2 shows the same profile, but with the deposition sites rank-ordered ("re-numbered") to reflect the descending order of the expected mean recoveries on each site. The standard deviations (SDs) do not necessarily change proportionally to the means, as can be seen from the SD values plotted as bottom panels on Figures 1 and 2. The relative standard deviations (RSDs) or the coefficients of variation (CVs)—both of which refer to the ratio SD/mean expressed as percentage—could in fact increase, decrease, or stay constant across the rank-ordered profile or some portions of it.

Another finding that allowed investigation of the potentially limitless number of possible profiles was the demonstration[5,6] that rank-ordered profiles could be modeled using beta-distributions, and that the chi-square ratio statistic's response depended on essentially 3 variables describing rank-ordered profiles; namely, their skewness, the SD of the first site, and the change of CV across a rank-ordered profile (ie, CV

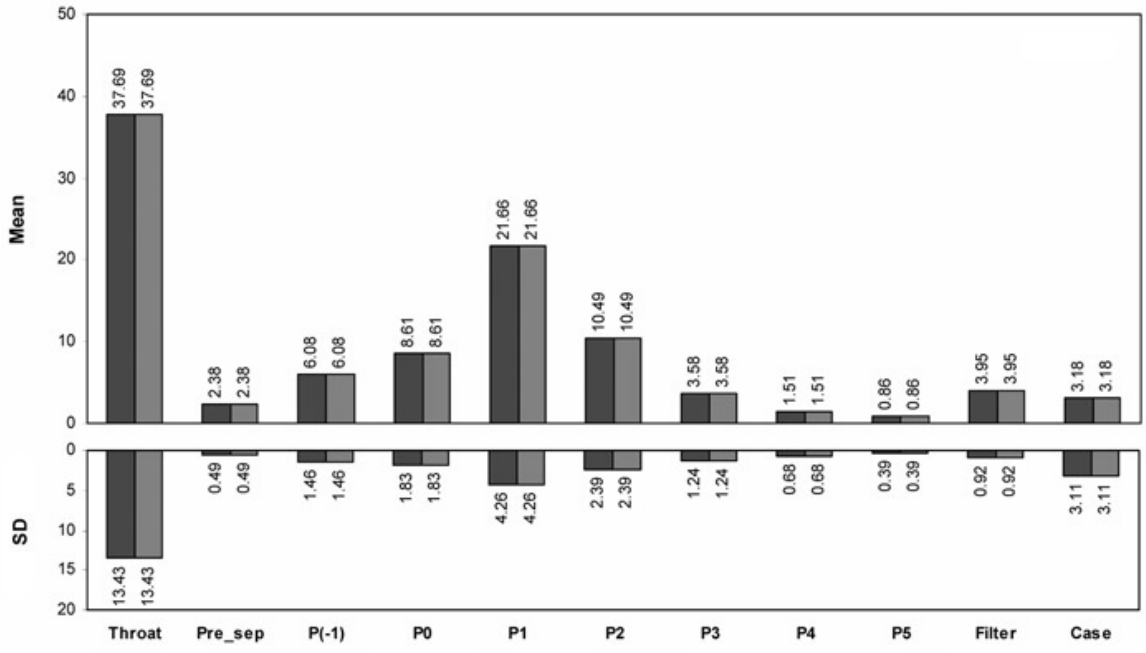

Figure 1. An example of a pair of profiles used in the identical-profiles study (ie, where the Test [gray] and Reference [black] profiles are identical. The profile is presented in the "natural" order that the deposition sites occur in a cascade impactor. The top panel shows mean deposition and the bottom panel shows the standard deviation (SD).

"slope"). The studied design space is illustrated in Figure 3, where the central cube identifies these variables and their ranges. From left to right across the design cube, the shape of the profile ranges from uniform (constant mean depositions on all sites) to maximum skew (large changes in mean deposition across the profile). From top to bottom of the design cube, the amount of variability, in terms of standard deviation (SD), for the first rank-ordered deposition site ranges from 1 to 10. From front to back of the design cube, the change over the entire profile in the coefficient of variation (CV, or SD relative to the mean deposition) for a specific site, ranges from 0 to 15 (ie, 0 change indicates constant CV across the profile; 15 indicates large increase in CV across the profile). The points at the 8 vertices of the design cube are illustrated

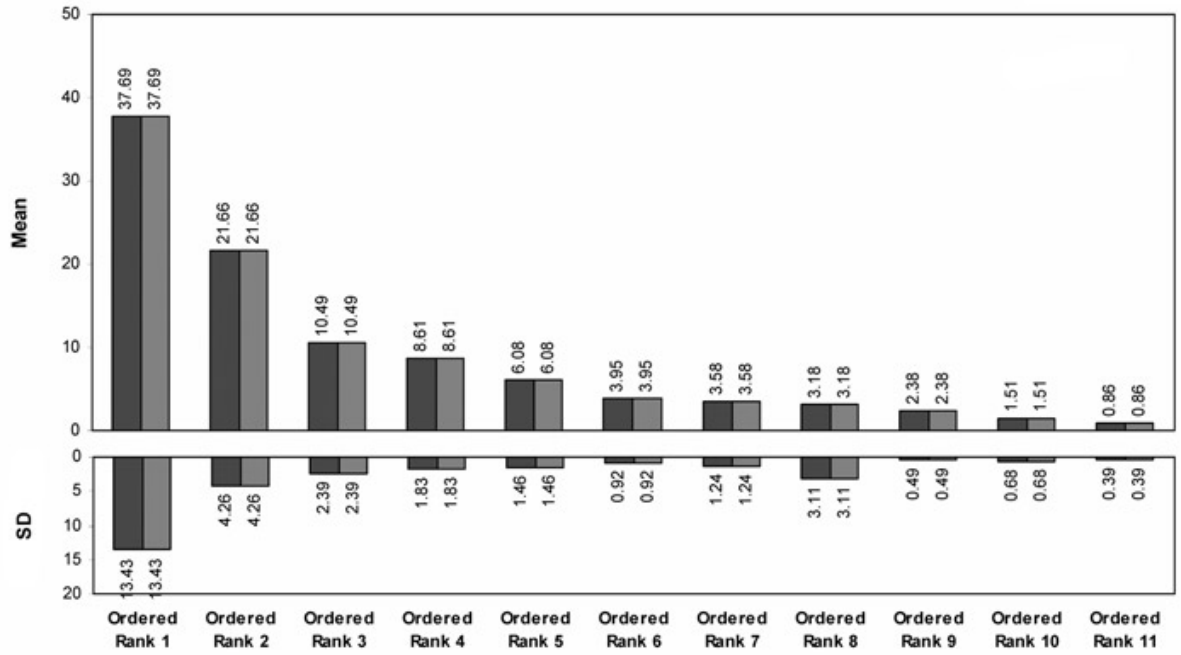

Figure 2. The same profile as in Figure 1, shown in rank-order. The "natural" deposition sites are reordered in the order of the highest-to-lowest mean deposition, which is shown on the top panel. The bottom panel shows the standard deviation (SD) on each site.

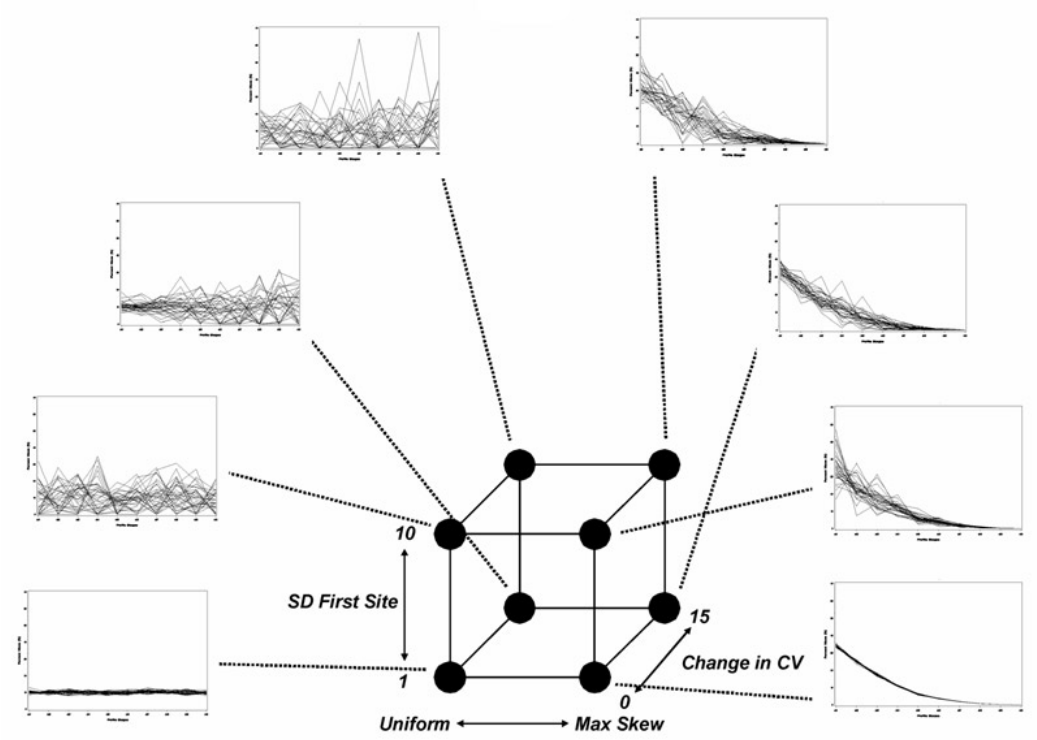

Figure 3. Experimental design approach to examining chi-square ratio stability for identical-profiles data. CV indicates coefficient of variation; SD, standard deviation. Max Skew indicates maximum skewed.

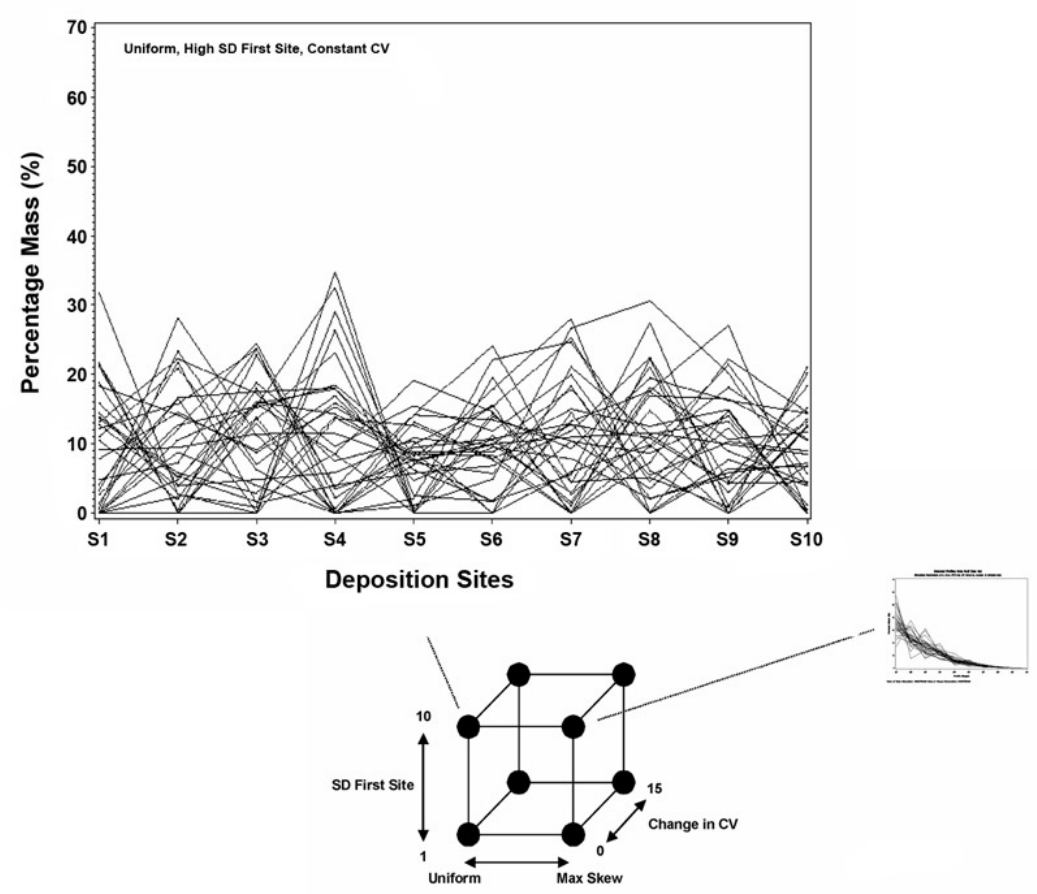

Figure 4. Simulated profile for a selected point, where the profile has a uniform shape, the standard deviation (SD) of the first rank-ordered site is at the highest level in the design space,[8] and the coefficient of variation (CV) is constant across the entire profile. Max Skew indicates maximum skewed.

by thumbnail plots of representative profiles. Each of the thumbnail plots is a "spaghetti" plot displaying results of 30 individual CI tests, thereby visualizing both the expected mean deposition and variability, and representing a product profile with the skewness, first-site SD, and the pattern of CV change as identified by the corresponding coordinates on the design cube. The axes on the thumbnail spaghetti plots are "Percentage Mass" and "Deposition Site." A close-up view of 2 of the spaghetti plots is presented in Figures 4 and 5. Figure 4 zooms in on a "uniform" profile, in which all deposition sites have the same mean deposition. Figure 5 zooms in on a highly skewed profile, which is generally representative of real-world profiles for MDIs and DPIs, which can be seen by the comparing this profile to the profile in Figure 2.

In the identical-profile study, for a given profile from the design space, both Test and Reference data were simulated from the same profile (a set of mean depositions and SDs), and the proposed chi-square ratio test was performed. This was repeated many times and finally the distribution of the obtained chi-square ratio test quantities was studied. This identical-profiles study showed what difference can be expected when "a product is compared to itself" and thus—among other things—gives some information on what levels are unsuitable for the critical value (ie, those levels that consistently result in "inequivalence" even though the profiles are identical). The identical-profiles study of all these profiles further showed that in general, the stability of the chi-square ratio statistic (ie, its consistency over a range of profile shapes and variability patterns) increases as the number of stages increases (profiles with 4, 7, 11, and 13 sites were examined).[6,9] The stability of the test statistic increases (ie, the variability of the test statistic decreases) for more linear rank-ordered profiles. The chi-square ratio statistic is less stable for profiles that are more common to MDIs and DPIs, which are less linear (ie, more skewed). The findings from the identical-profiles study provided a clearer path for subsequent work.

Changes on a Single Impactor Site—Concept Study

As the next level of complexity, the case was studied where deposition on a single deposition site differed between Test and Reference profiles. This effort was accomplished systematically through simulations, by changing the deposition on a single site by a specified amount, as described in Table 1. A visual explanation of the studied patterns of changes is available through the Minutes of the Working Group's discussions, which are posted publicly on the Internet[10] and by comparing the resultant Test to the original Reference profile via the chi-square ratio test. In these simulations, the interstage correlation was included in the model.

For this "single-site-change" concept study, a pulmonary chlorofluorocarbon metered-dose inhaler (CFC MDI) and a device-metered DPI product were selected as the basis for modeling the simulated profiles. The Reference-to-Test modifications consisted in increasing deposition on a single site for the Reference profile (one site at a time, each site in turn) by –10%, +20%, +40%, or +60% and renormalizing the modified profile. Mean chi-square ratios were calculated for a range of different scenarios. The mean chi-square ratio generally increased from the initial value (when Test and Reference are drawn from the same distribution); however, the amount of the increase depended on the mean deposition of the site that was changed.

The results are illustrated in Figures 6 and 7. In Figure 6, different lines correspond to the various changes in deposition on a single site, one site a time. Each of these modified profiles was compared with the original profile and a chi-square ratio was calculated. The mean chi-square ratios obtained from this process are shown along the y-axis in Figure 6.

Figure 7 demonstrates that the chi-square ratio depends not only on the amount of change (percentage) but also on the particular CI site where the change is observed. Here, the induction port, with its highest deposition, gives rise to the largest changes in the chi-square ratio, whereas sites with lower depositions (eg, lines H; compare with Figure 6) are less capable of influencing the chi-square ratio.

These results showed that for the studied product profiles, the chi-square ratio statistic was more sensitive to changes in deposition on sites with high deposition than on sites with low deposition. In this example, changes in the induction port (which had the highest mean deposition) caused much larger increase in the chi-square ratio than for any subsequent site, by multiple-fold. Contrary to the intuitive assumption, the chi-square ratio can decrease as the difference in deposition between the Test and Reference profiles on a particular stage or accessory component increases. The underlying reason for this behavior of the test is not clear. In summary, these results showed that not only how much of a change, but where the change occurs, can affect the chi-square ratio statistic. For the studied profiles, the method was more sensitive to stages with higher deposition than to stages with lower deposition. The conclusions of this study are important not from the perspective of particular profiles, but from the perspective of how the chi-square ratio test behaves given the specified differences in mean and variability between Test and Reference.

"Target" Profiles

Through the studies described above, the Working Group also realized that in order to compare the chi-square ratio test with any alternative test, it would be necessary to have some "targets" (ie, profiles that would be a priori pronounced

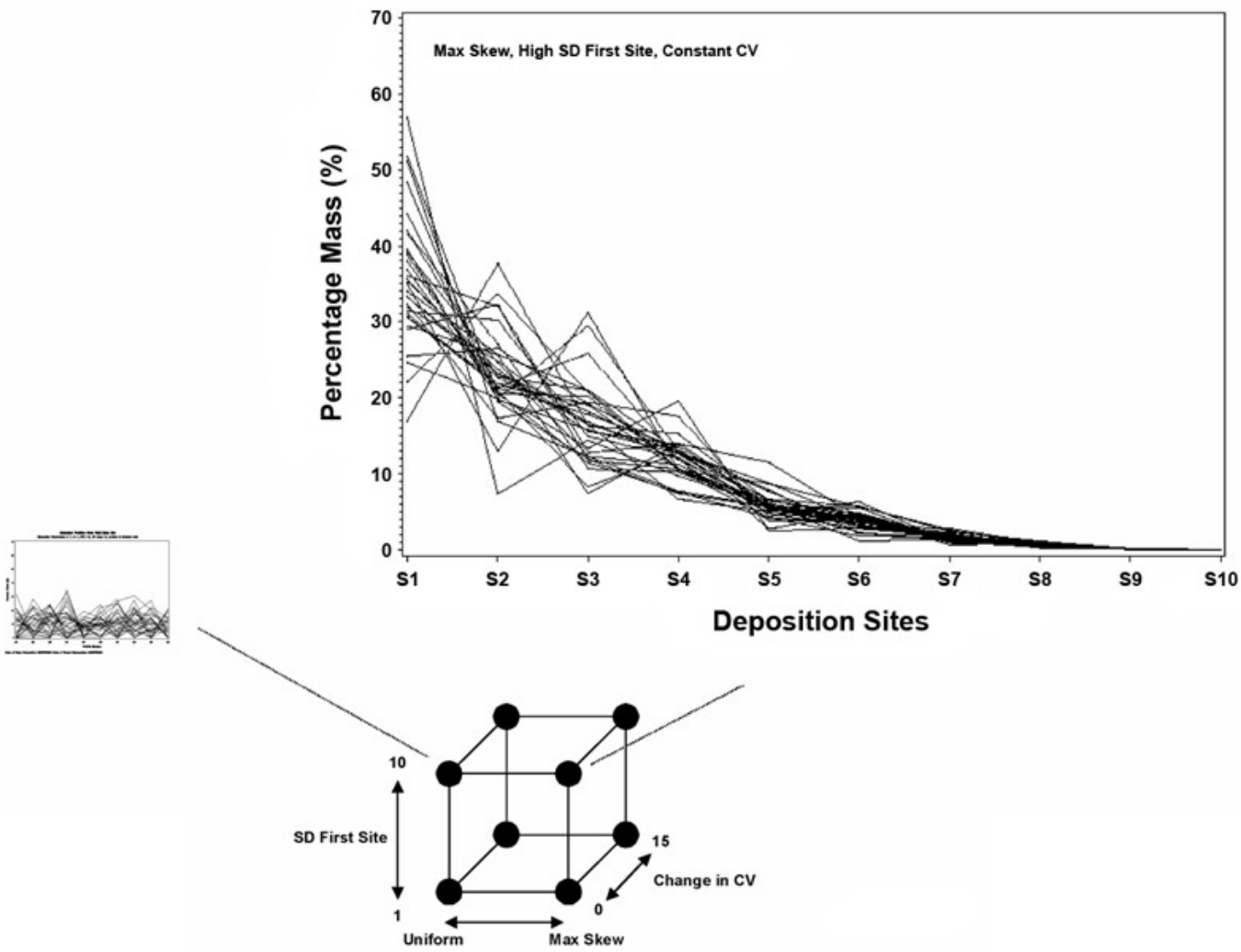

Figure 5. Simulated profile for a selected point where the profile has a maximum skewed (Max Skew) shape; a large change in mean deposition from the first rank-ordered site to the last, the standard deviation (SD) of the first rank-ordered site is at the highest level in the design space,[8] and the coefficient of variation (CV) is constant across the entire profile. Max Skew indicates maximum skewed.

"equivalent" (not "equal," but "sufficiently similar"), as well as profiles that a priori are pronounced as "failing to show equivalence." With a given set of such target profiles, the Working Group would be able to estimate how consistently the chi-square ratio (or an alternative) test makes correct decisions.

A thorough consideration was given to the selection of target profiles, which included the following points. Just as the definition of "equivalence" cannot be achieved by statistics alone, some nonstatistical input is generally needed at least initially in the selection of the target profiles. Neither can equivalence be determined in a general way based on clinical reasoning alone, because clinical performance is drug-specific, while this test aims to provide an unambiguous, objective tool for determining in vitro equivalence. Therefore, when the Agency was developing the chi-square ratio method, FDA pharmacologists' input was used to suggest equivalent differences, or differences that fail to establish equivalence, and the statisticians helped to translate that judgment into a numerical critical value. In order to find a consistent basis for selecting sufficiently similar or dissimilar profiles ("target profiles"), and preferring to build on the Agency's previous work, the Working Group followed the pathway by which the Agency arrived at the critical value quoted for albuterol MDI. In this study, the Working Group used the systematic set of scenarios of interest that had already been developed per the protocol in Table 1. For each of these scenarios, the performance of the chi-square ratio test was revealed via simulations defined as follows:

- Simulations incorporate interstage correlation observed from absolute recovery data.
- The variation in the simulated data bracketed the variation observed from the baseline data.
- For each scenario or design point, the chi-square ratio algorithm was applied 1000 times. This means that 30 Test and

Table 1. Scenarios Studied Initially, Involving Combinations of Changes in the Mean Profile of the Test Population and Changes in the Levels of Test and Reference Variability*

Types of Changes	Levels Studied in Initial Scenarios
Changes in the Test population mean profile	The following levels of change in the Test population mean were studied: 1. No change 2. 10% increase for 1 site with high recovery 3. 10% increase for 1 site with medium recovery 4. 10% increase for 1 site with low recovery 5. 100% increase for 1 site with high recovery 6. 100% increase for 1 site with medium recovery 7. 100% increase for 1 site with low recovery Definitions of high, medium, and low recovery sites are dependent on product type and are guided by analysis of industry data and additional sources. Recoveries on the other stages and sites are decreased in proportion to the percentage deposited at that site.
Changes in the levels of Test and Reference population variability	A "Mid" level of variation is the estimated RSD associated with a particular site from the observed data profiles. The "Low" level represents half the estimated RSD from each site, while the "High" level represents a tripling of the estimated RSD. The estimated RSD associated with a particular site reflects the total variability from the observed data. The following combinations were studied: 1. Reference variation Low, Test variation Low 2. Reference variation High, Test variation Low 3. Reference variation Low, Test variation High 4. Reference variation High, Test variation High 5. Reference variation Mid, Test variation Mid
Changes in both mean and variability, which resulted in the combined studied scenarios	Based on the above design in which the 5 levels of Test and Reference population variability are paired with the 7 levels of changes in the mean profiles, a total of 35 scenarios were considered. Three additional combinations were added to investigate midpoint changes: 50% increase in the Test population mean with mid levels of Test and Reference variability across low, medium, and high recovery sites. The 3 additions bring the total number of scenarios to 38. The spaghetti plots and corresponding statistics for these 38 scenarios are available on the PQRI Web site.[8]

*RSD indicates relative standard deviation.

30 Reference profiles were generated 1000 times, and each time the profiles were generated, the data were simulated as if the units were tested from one batch (ie, the total variation is the within batch variability).

- For each simulation, the 30 Test and 30 Reference profiles yielded a population of unique triplets of profiles (Test, Ref1, Ref2). From this population, 500 triplets were sampled, and the mean ratio of chi-squares obtained. The selection of 500 triplets and calculation of the mean ratio was repeated 300 times, yielding, for each simulation, a distribution of 300 means. From this distribution of 300 means, 4 summary statistics were reported: the 50th percentile, the 90th percentile, the 95th percentile, and the mean. This entire process was repeated 1000 times, resulting in a distribution of 1000 90th percentiles, 1000 95th percentiles, and so forth.
- For each of the 4 summary statistics listed above, the entire distribution across the 1000 simulations was reported.
- The assessment of stability of the chi-square ratio was primarily focused on understanding how changes across the design space influence the location and shape of the distributions resulting from 1000 applications of the chi-square ratio algorithm.
- Second, the assessment explored options for determining critical values. Selection of a critical value should incorporate understanding both the location and shape of the distributions resulting from the simulations. This increased understanding will (1) highlight additional areas within this design space (or more complicated design spaces) that need further investigation; and (2) eliminate certain regions of the design space from further investigation.
- The information gathered from this initial protocol will help determine what designs should be considered next when investigating changes across multiple impactor deposition sites.

Finally, the performance of the chi-square ratio test under each scenario was compared with the predetermined assignment of equivalent profiles, which was provided by FDA scientists. In other words, based on an independent FDA estimate

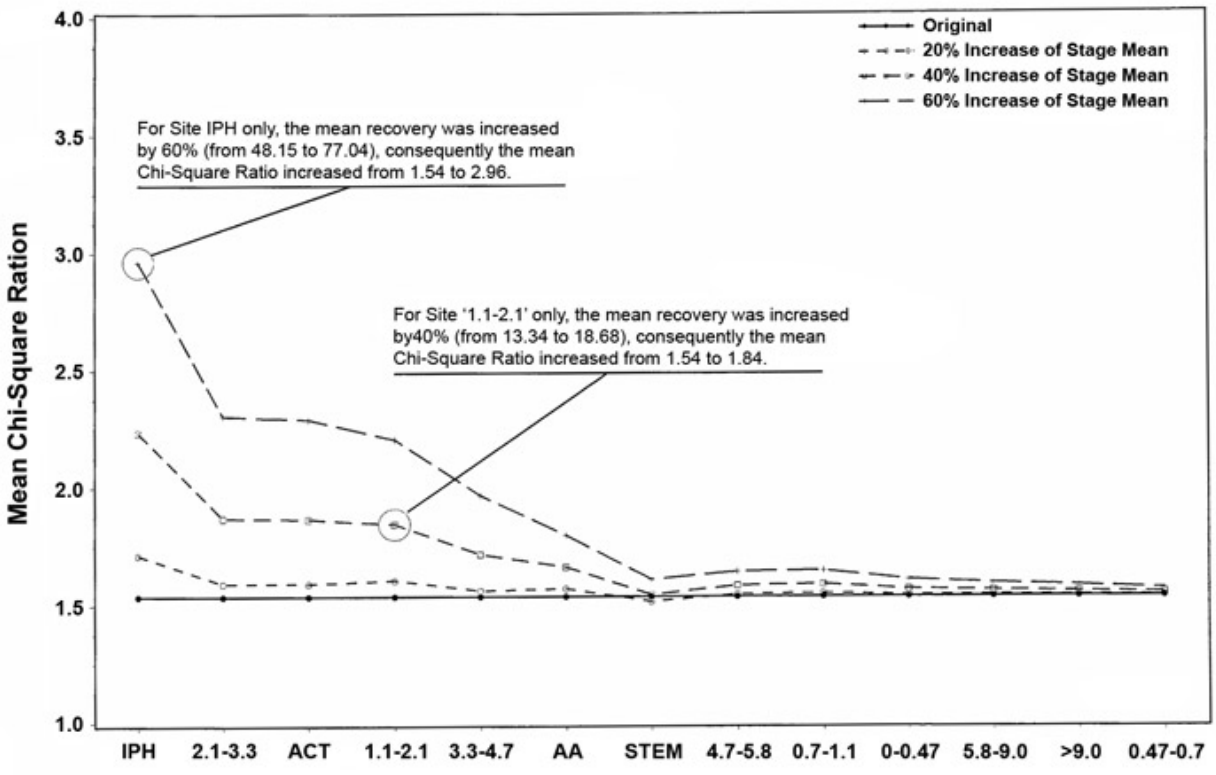

Figure 6. Mean chi-square ratios for a concept study, where Test and Reference differed on a single cascade impactor (CI) site represented on the x-axis (rank-ordered from highest to lowest mean deposition). IPH indicates induction port; ACT, actuator; AA, additional accessories. CI stages are identified by their nominal size cutoffs.

of equivalence for the purpose of this exercise, the Working Group has been able to see how often the chi-square ratio test can arrive at the same answers regarding equivalence as those predetermined by FDA scientists. In the end, the probability, or frequency, of accurate decisions made by the test should help decide whether the test is fit for the intended purpose.

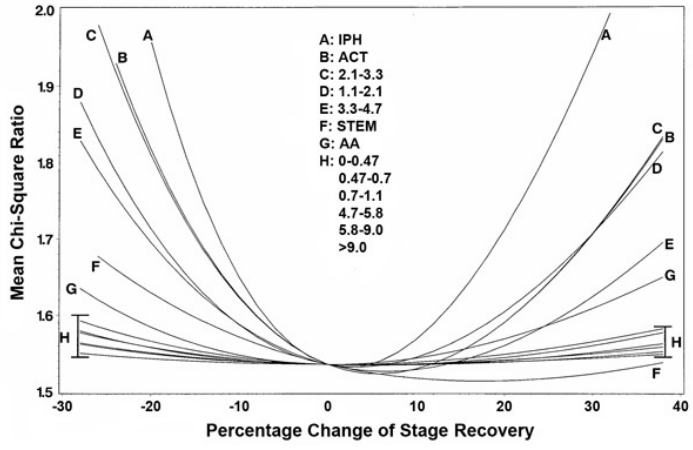

Figure 7. Mean chi-square ratio versus percentage change of site recovery for the "single-site change" concept study. Each curve corresponds to the change on a single site (same notation as in Figure 6). The amount of change is shown on the x-axis. The resultant mean chi-square ratio is shown on the y-axis.

Results and Discussion

Outcome of Protocol Investigations

By the end of 2004, the Working Group completed its investigations as described above. The obtained set of realistic simulated scenarios and detailed information about their basic statistics is publicly available on the PQRI Web site.[8] A chi-square ratio statistic was calculated 300,000 times for each of these scenarios as explained in the previous section. From these calculations, a distribution of the various percentiles of the chi-square ratio metric was obtained; and the central tendency and spread of these percentiles were visually presented as box-and-whisker plots. An examination of these plots showed that it would be difficult to select a critical value of the chi-square ratio that could consistently separate equivalent profiles from those failing to show equivalence. This difficulty can be observed in the overlap of the 95th percentile's distributions for some scenarios.[11-14]

The conducted systematic analysis also revealed that at least in some situations, the chi-square ratio statistic has differing sensitivity depending on the size fractions where changes occur. As illustrated by the curves in Figure 7, the greatest sensitivity (discriminating power) has being observed for changes in that portion of the APSD profile that has no relevance for "respirable fraction" of the inhaled drug (related to efficacy), although it might have relevance

for the swallowed portion (related to safety). The overpowering effect of the sites with high deposition (such the CI throat or preseparator) leads to the limited ability of the test to detect differences between profiles that were deemed significant in an independent assessment by an internal FDA team involving chemists, statisticians, pharmacologists, and application reviewers.[15]

Based on these results and observations, in early 2005 the conclusion was reached that the chi-square ratio test by itself is not sufficiently discriminating, as in some scenarios the test declared "equivalent" profiles that in the FDA's judgment should be regarded as "failing to show equivalence." To enhance the discrimination of the chi-square ratio test, the FDA participants on the Working Group proposed adding an impactor-sized mass (ISM) test as part of the in vitro profile comparison test. Impactor-sized mass (ISM) was defined by the Working Group as the sum of the drug mass on all stages of the cascade impactor plus the terminal filter, but excluding the initial stage because of its lack of a specified upper cut-off limit. The form, properties, and criteria for this test are being explored by the full Working Group. In this ongoing new work, the Working Group is also looking at more complex and realistic scenarios than those involving only a change in deposition on a single site.

During the Working Group's discussion, a question has been raised whether a different statistical test, not based on the chi-square ratio, could prove more suitable. To date no such test has been proposed within or to the Working Group. For an alternative test to be considered a viable replacement of the currently contemplated tests, it would have to demonstrate superior consistency in making correct decisions and would have to meet the same design objectives; namely, taking into account deposition on each of the sites individually and rewarding the Test product for lower variability than the Reference product. Any such test, when and if proposed, should be evaluated using the tools and models that have been used to evaluate the chi-square ratio test, as reported here.

Ongoing and Future Work

Investigations of Realistic Scenarios

From early 2005, the Working Group has focused on 2 main objectives: (1) developing a new protocol for generating scenarios that would be more realistic than the simpler-case scenarios described above, which were based on the changes introduced on a single site, and (2) investigating the proposed additional test (ISM) in combination with the chi-square ratio test using the new scenarios. As with the first protocol, the Working Group, with input from the FDA participants, will determine which of the new scenarios represent "sufficiently similar profiles." Based on this a priori determination, the chi-square ratio test supplemented with an ISM test will be evaluated using the already developed statistical tools and techniques. A report about results of these investigations will be prepared and publicized in the near future.

Testing Design

Cascade impactor results are known to be highly dependent on the particular impactor type, model, and configuration, the physical condition of the particular unit on which the testing is performed, operator skills, environmental conditions, and sample preparation—to name just a few main factors.[16-18] To prevent such non–product-related variability from influencing the comparison, crossover designs should be followed for APSD profile comparisons similar to the crossover designs used in clinical studies to address subject-to-subject variability. In the current context, the crossover design would prescribe how the Test and Reference products and impactors should be arranged in cohorts and tested contemporaneously (ie, each Test/Reference pair is to be tested by the same analyst, on the same impactor, on the same day), in order to address variations that are irrelevant to the product. These and other considerations related to testing design will be addressed by the Working Group and will become part of the final recommendations.

Critical Value

A critical value for defining equivalence should be based on both clinical and quality considerations. However, no generally predictive correlation has been shown between cascade impactor profiles and clinical response.[19] Statistics alone cannot determine the selection of target profiles, just as statistics alone cannot define what is meant by "equivalence" in a given context. Therefore, the critical value for equivalence will be defined in terms of product quality with input from the FDA. Given an outside definition of equivalence (eg, based on clinical or quality considerations), statistical tools could be developed that test for such equivalence with desired levels of sensitivity, discrimination, and confidence. Using the conceptual and statistical tools developed by the Working Group as described above, and given an indication of "sufficiently similar" target profiles, the Working Group will have an objective way for setting critical values or at least will provide some recommendations regarding the critical values.

Stage Width Adjustment

During the Working Group's discussions, the issue was considered of whether the cascade impactor data on stages need to be normalized to unit size interval in order to account

for different stage widths.[20] In light of the FDA focus on component-to-component comparisons rather than on continuous density models, the Working Group did not pursue this line of investigation. Two members of the Working Group subsequently studied this issue independently from the Working Group.[21] Future investigations of an optimal profile comparisons test may want to consider the stage width adjustment in order to improve statistical properties and generalizability of the test. This work, however, is outside the scope of the current PQRI Working Group.

CONCLUSIONS

The PQRI process has been an effective way to investigate and build broad consensus on such a complicated regulatory and scientific issue as a statistical method for comparing aerodynamic particle size distribution profiles of pharmaceutical aerosol products. The properties of the test in the draft guidance have been characterized, features important to industry and regulators have been clarified, and the development of an improved approach has been made possible.

The studies reported here have shown that the chi-square ratio statistic is more sensitive to changes on cascade impactor sites with higher deposition and less sensitive to changes on sites with lower deposition. As a result, this test may not always detect differences in APSD profiles that could be significant from the patient's perspective. To augment the discriminating power of the chi-square ratio test, an additional test for impactor-sized mass as evaluated by Population Bioequivalence methods was proposed and is being investigated.

ACKNOWLEDGMENTS

This article represents professional opinions of the authors and does not necessarily represent the views or policies of the FDA or any other institution with which the authors are affiliated. The Working Group would like to thank the PQRI and all its member organizations for encouraging and supporting this work. The authors also thank the AAPS PharmSciTech Editor-in-Chief Pat DeLuca and the anonymous referees for helpful comments and suggestions made during review of the manuscript. The following PQRI APSD Profile Comparisons Working Group contributed to this work. The PQRI APSD Profile Comparisons Working Group is chaired by David Christopher (Schering-Plough Research Institute). Members of the Working Group (in alphabetical order) are Wallace P. Adams (FDA), Anton Amann (ACN Pharma), Craig M. Bertha (FDA), Peter R. Byron (Virginia Commonwealth University), William H. Doub (FDA), Craig Dunbar (Alkermes), Walter W. Hauck (Thomas Jefferson University, current address: USP), Douglas S. Lee (Pfizer), Richard Lostritto (FDA), Svetlana Lyapustina (Drinker, Biddle and Reath LLP), Jolyon P. Mitchell (Trudell Medical Group), Beth Morgan (GlaxoSmithKline), Steve Nichols (Aventis), Ziqing Pan (Schering-Plough Research Institute), Gur Jai Pal Singh (FDA), Terrence Tougas (Boehringer Ingelheim), Yi Tsong (FDA), Ronald K. Wolff (Nektar), Bruce Wyka (Schering-Plough Research Institute), and the Working Group's liaisons to the PQRI Drug Product Technical Committee Michael Golden (GlaxoSmithKline) and Guirag Poochikian (FDA).

REFERENCES

1. FDA/CDER. Draft Guidance for Industry: Bioavailability and Bioequivalence Studies for Nasal Aerosols and Nasal Sprays for Local Action. *Food and Drug Administration Web site.* 1999. Available at: http://www.fda.gov/cder/guidance/2070DFT.pdf. Accessed: April 17, 2006.

2. Adams WP, Christopher D, Lee DS, et al. PQRI evaluation of cascade impactor profiles of pharmaceutical aerosols: Part 1- Background for a statistical method. *AAPS PharmSciTech.* In press.

3. PQRI PSD Profile Comparisons Working Group. Work plan for the investigation of an optimized chi-square method for comparing particle size distribution profiles obtained by cascade impactors with specific reference to equivalence testing of orally inhaled and pressurized nasal drug products. 2002. Available at: http://www.pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/workplan02.pdf. Accessed: August 18, 2006.

4. Tsong Y. Statistical comparison of particle size distribution profiles. 2004. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/Addl/DC01-475116-v2-Yi_Tsong_Statistical_Archive_PQRI_Profile_Comparisons.DOC. Accessed: April 20, 2006.

5. Lee D. Searching for the holy grail of a single PSD profile comparator. 2004. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/Addl/DC01-504108-v1-Identical_Profiles_and_Stabilty_Slides.PPT. Accessed: April 20, 2006.

6. Lee D. Searching for the holy grail of a single PSD profile comparator: technical challenges with the chi-squared ratio approach. In: Dalby RN, Byron PR, Peart J, Suman JD, Farr SJ, eds. *Respiratory Drug Delivery IX: Biological, Pharmaceutical, Clinical, and Regulatory Issues Relating to Optimized Drug Delivery by Aerosol.* vol 1. River Grove, IL: Davis Healthcare International Publishing; 2004:161–169.

7. Pan Z, Christopher JD, Lyapustina S, Chou E. Statistical techniques used in simulation of cascade impactor particle size distribution profiles. In: Dalby RN, Byron PR, Peart J, Suman JD, Farr SJ, eds. *Respiratory Drug Delivery IX: Biological, Pharmaceutical, Clinical, and Regulatory Issues Relating to Optimized Drug Delivery by Aerosol.* vol 3. River Grove, IL: Davis Healthcare International Publishing; 2004:669–672.

8. PQRI Profile Comparisons Working Group. Systematic Study of 38 Scenarios. 2005. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/Addl/38%20Systematic%20Scenarios.pdf. Accessed: August 17, 2006.

9. PQRI PSD Profile Comparison Working Group. Draft simulation results on stability of the chi-square ratio statistic under the null case of identical sets of reference and test profiles. 2002. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/Addl/DC01-504131-v1-PQRI_Profile_Comparisons_Slides_for_DPTC.PPT. Accessed: August 17, 2006.

10. PQRI PSD Profile Comparisons Working Group. Minutes of the teleconference on 7 July 2004:7-8. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/070704min.pdf. Accessed: August 17, 2006.

11. PQRI PSD Profile Comparisons Working Group. Minutes of the teleconference on 27 October 2004:3-7. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/102704min.pdf. Accessed: August 17, 2006.

12. PQRI PSD Profile Comparisons Working Group. Minutes of the teleconference on 22 September 2004:4-14. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/092204min.pdf. Accessed: August 17, 2006.

13. Christopher D, Ziqing P, Lyapustina S. Aerodynamic particle size distribution profile comparison: considerations for assessing statistical properties of profile comparison tests. *Am Pharm Rev.* 2005;8:68–72.

14. PQRI PSD Profile Comparisons Working Group. Minutes of the teleconference December 14, 2005:7. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/121405min.pdf. Accessed: August 18, 2006.

15. PQRI PSD Profile Comparisons Working Group. Minutes of the teleconference February 16, 2005:5-6. Available at: http://pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/021605min.pdf. Accessed: August 17, 2006.

16. Hauck W. Statistical analysis, equivalence and inequivalence: an overview of data reporting. Presentation at AAPS/FDA/USP Workshop: Particle Size Analysis; April 30 to May 2, 2003; Washington, DC.

17. Christopher D, Curry P, Doub B, et al. Considerations for the development and practice of cascade impaction testing, including a mass balance failure investigation tree. *J Aerosol Med.* 2003;16:235–247.

18. Mitchell JP. Regarding the development and practice of cascade impaction testing, including a mass balance failure investigation tree. *J Aerosol Med.* 2003;16:433.

19. Newman SP, Wilding IR, Hirst PH. Human lung deposition data: the bridge between in vitro and clinical evaluations for inhaled drug products. *Int J Pharm.* 2000;208:49–60.

20. Hinds WC. *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles*. New York, NY: John Wiley & Sons; 1999:75–110.

21. Dunbar C, Mitchell J. Analysis of cascade impactor mass distributions. *J Aerosol Med.* 2005;18:439–451.