Testing the equivalence of aerodynamic particle size distributions

This article compares different methods for equivalence testing of aerodynamic particle size distributions, with a focus on the modified chi-square ratio statistic

Benjamin Weber, MStat, PhD
Boehringer Ingelheim

Objective methods for evaluating the equivalence of aerodynamic particle size distributions (APSD) remains a key goal for European and US guidelines aimed at determining whether a generic inhalable drug product is equivalent to an innovator drug product.[1-5] Multi-stage cascade impactors, such as the Andersen Cascade Impactor or the Next Generation Impactor (NGI), are typically used for determining the APSD of inhaled drug products. Each impactor stage captures particles in a defined range of aerodynamic particle sizes. One impactor test yields multiple data points because there are multiple impactor stages. Two inhalable drug products with different APSDs will likely differ from each other on multiple stages thus creating a multivariate data analysis problem. This problem is further complicated by the fact that an impactor setup contains non-size measurement components (e.g., induction port or pre-separator) that are not well defined as particle size fractionators as are the impactor stages themselves. Figures 1A and 1B illustrate the setup of an NGI, including a categorization of stages into size measurement and non-size measurement components.

Two topics concerning APSD equivalence testing have been controversially discussed in the last years. First, to which extent, if at all, can APSD data be used for drawing conclusions about the efficacy and safety of inhaled drug products? Second, which statistical methodology should be applied to APSD data when formally testing for equivalence between a test and a reference product? While the former is beyond the scope of this article, the latter is complicated by the fact that some of the available methods are complex and not very tangible. A stepwise cascade impactor equivalence test has recently been proposed for equivalence testing in APSD that is based on the so-called population bioequivalence[A] (PBE) test (see article sidebar) and the modified chi-square ratio statistic (mCSRS)[6-10] (Figure 2). Both the PBE test and the mCSRS certainly qualify as being non-intuitive methods. Therefore, it is the focus of this article to foster understanding about the proposed stepwise cascade impactor equivalence test, including the PBE test and the mCSRS. Moreover, the PBE test is compared with the average bioequivalence (ABE) test (see article sidebar) and the outcome of the stepwise cascade impactor equivalence test is compared with performing statistical tests on individual stages. Based on a practice data set (available at www.bweber.net/downloads/) of NGI profiles obtained from a reference (REF) and a test (TEST) product (Figure 3), different APSD equivalence testing methods (see above) are demonstrated and critical aspects of the computation and interpretation of the results are explained. Additionally, a user-friendly web application that enables the users to apply the stepwise cascade impactor equivalence test to their own NGI data is provided (available at https://bweber.shinyapps.io/MmCSRS_NGI/).

The data set

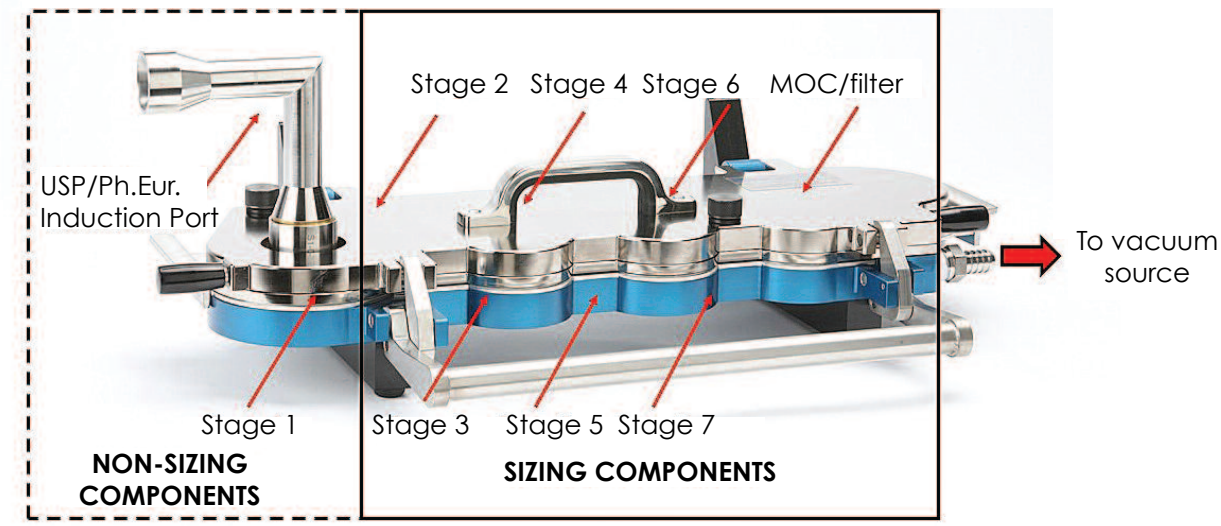
The data set consists of a total of 60 NGI profiles from a hypothetical REF product (n = 30) and a hypothetical TEST product (n = 30) (Figure 3). A publicly-available source of impactor profiles was used as the basis for constructing the data set via Monte-Carlo simulations.[11-12] Whether TEST and REF products from which profiles were simulated were, in fact, the same or different is revealed in this article's conclusions.

[A] It should be noted that the term "population bioequivalence" refers to a statistical methodology[10] and involves here only *in vitro* data, not "biological" or *in vivo* data.

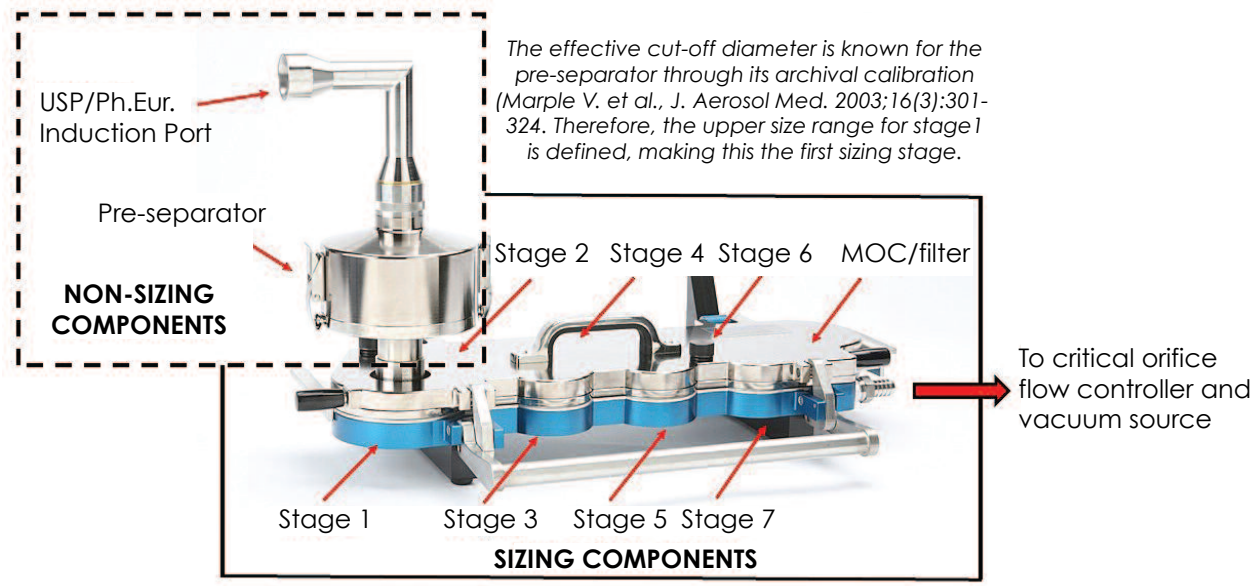
Figure 1(A,B)

(A) NGI Sampling Train for Testing Pressurized Metered Dose Inhalers

The effective cut-off diameter is not defined for the induction port, therefore, the upper size range for stage 1 is also undefined, making stage 2 the first sizing stage.



(B) NGI Sampling Train for Testing Dry Powder Inhalers



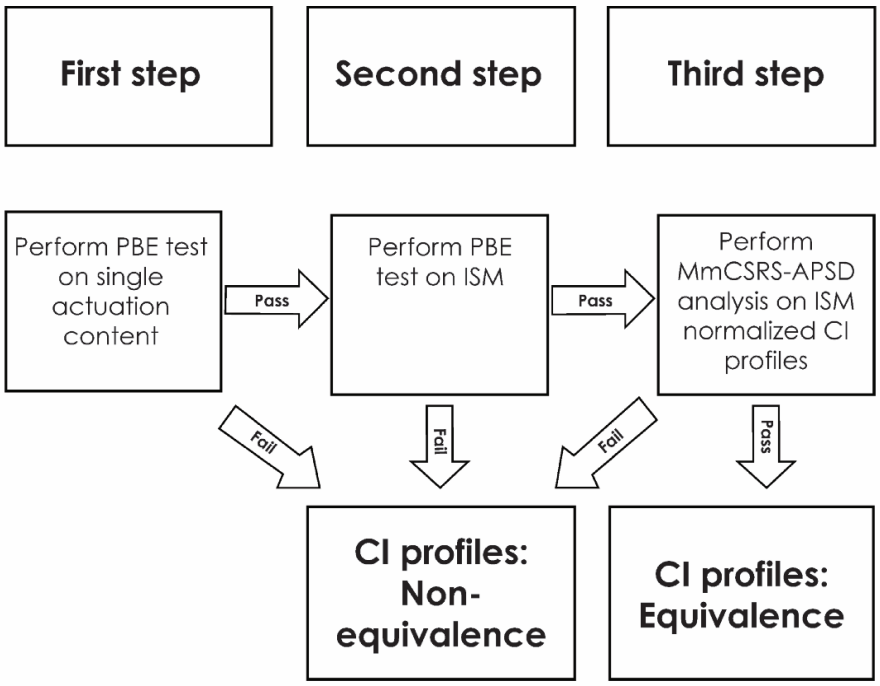
The NGI in setups for testing pressurized metered dose inhalers (pMDIs) and dry powder inhalers (DPIs), including categorization of NGI stages into size measurement and non-size measurement components. Figures courtesy of Jolyon Mitchell Inhaler Consulting Services, Inc.; NGI photos courtesy of MSP Corporation.

Visual analysis

Visual analysis together with simple summary statistics of the data are key elements of cascade impactor-derived data analysis and should always be conducted before performing any statistical tests.

Figure 3 and Table 1 provide a visualization and summary statistics of the data, respectively. TEST and REF products seem to differ on the non-size measurement stages induction port and pre-separator and on the size measurement stages S2, S3, S4 and S5. The differences on the size measurement stages, however, become more apparent on a different y-axis scale (Figure 3; middle panel). While visual analysis does not reveal big differences on stages S7 and MOC (micro-orifice collector), numerical comparison shows differences, in particular, when expressed as a TEST-to-REF ratio (1.11 and 0.72, respectively). Interestingly, expressing the difference on the pre-separator as a TEST-to-REF ratio yields 0.9, pointing out an important aspect when performing equivalence tests on individual stages based on

Figure 2



Flow chart of the proposed stepwise cascade impactor equivalence test. ISM: impactor sized mass. This flow chart was previously published in Weber B, Lee SL, Delvadia R, Lionberger R, Li BV, Tsong Y, et al. Application of the modified chi-square ratio statistic in a stepwise procedure for cascade impactor equivalence testing. The AAPS Journal. 2015;17(2):370-9.

ratios. Namely, small absolute differences on stages with a low deposition can result in relatively large difference in ratios compared with stages with a high deposition. It may thus be worthwhile to consider alternative approaches for analyzing equivalence on low deposition sites. The total mass on all ten stages and the impactor sized mass (ISM[B]) were 328 (48.9) and 93.9 (6.65) mcg, respectively, (mean (standard deviation)) for the TEST product and 331 (38.9) and 87.2 (5.39) mcg, respectively, for the REF product.

Stepwise cascade impactor equivalence test

The first and second steps of the stepwise cascade impactor equivalence test are performing PBE tests on single actuation content (here replaced by total mass on all components) and calculation of ISM. Performing a PBE test for total mass and ISM results in a fail (criterion > 0) and a pass (criterion < 0), respectively. For the former, the test product gets punished for its larger variability (see article sidebar). Performing an ABE test, as an alternative to the PBE test, yields 90% confidence intervals of 0.927-1.05 and 1.05 -1.11 for total and ISM TEST-to-REF ratios, respectively.

The modified chi-square ratio statistic

The computational form of mCSRS is summarized below.

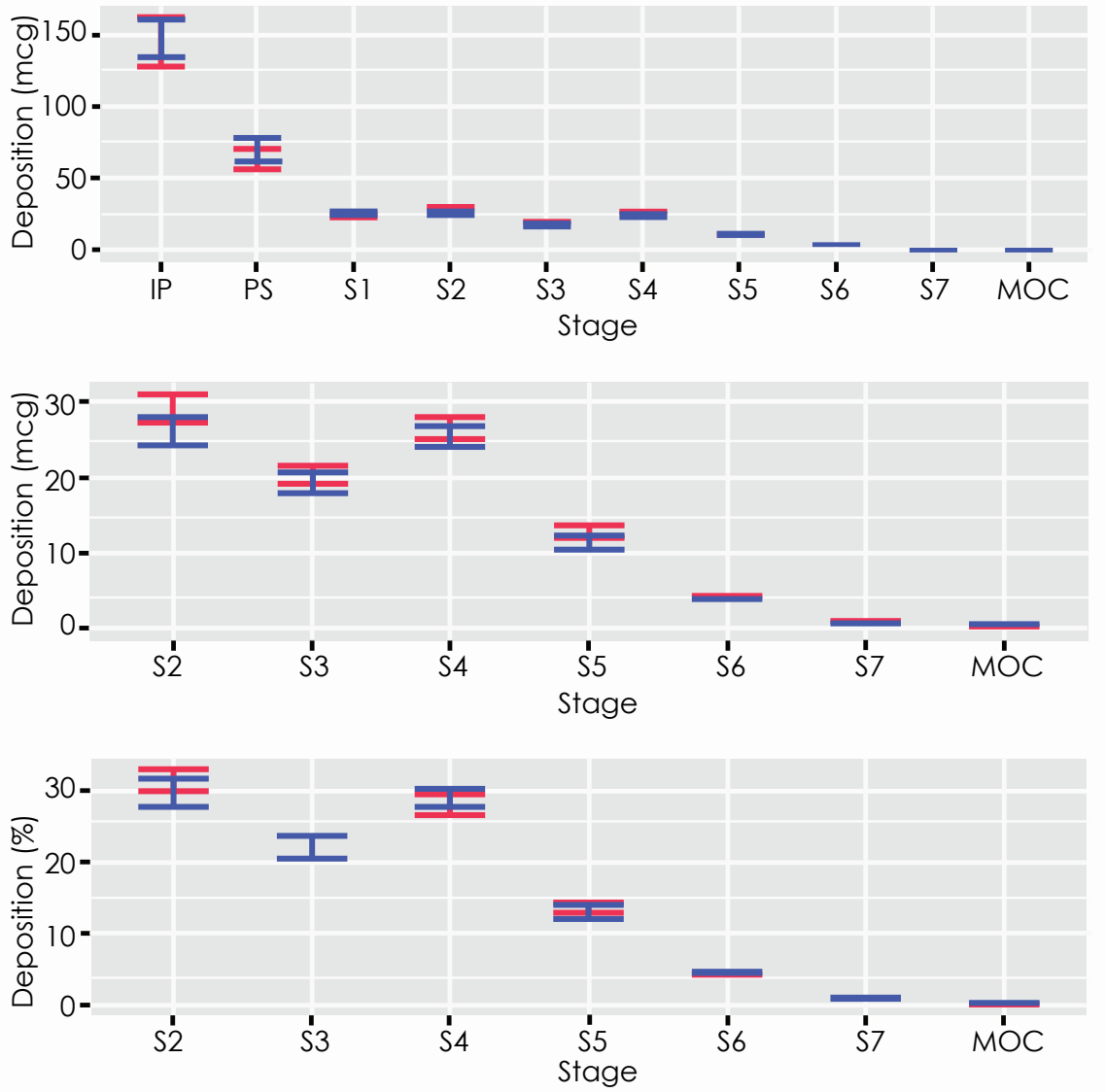
$$mCSRS_{jk} = \frac{\sum_{i=1}^{p} \frac{(T_{ij} - \bar{R}_i)^2}{\bar{R}_i}}{\sum_{i=1}^{p} \frac{(R_{ik} - \bar{R}_i)^2}{\bar{R}_i}}$$

where p represents the number of deposition stages, T_{ij} and R_{ik} represent the normalized deposition (i.e., by dividing the absolute deposition on the i^{th} stage by the total deposition on all stages under consideration) on the i^{th} stage of the j^{th} profile ($j = 1, \ldots, n_T$) of the TEST sample and on the i^{th} stage of the k^{th} profile ($k = 1, \ldots, n_R$) of the REF sample, respectively. n_T and n_R represent the number of samples (here 30 and 30) that were obtained from the TEST and REF product, respectively, and $\bar{R}_i$ represents the sample mean on the i^{th} stage of all REF profiles.

The mCSRS and its predecessor the "original" chi-square ratio statistic[13-15] were designed with the goal of providing a single metric that summarizes potential differences on multiple stages of a TEST and a REF profile. Moreover, both were designed to reward or penalize TEST products for being more or less variable than REF products, respectively. In this context, rewarding and penalizing can be understood as increasing and decreasing the likelihood of demonstrating equivalence for two products with a defined difference in their means, respectively. It should be noted that the PBE test has the same property (see article sidebar). It should thus not be a surprise that

[B] ISM is defined as the sum of the drug mass on all NGI stages plus the MOC, but excluding the initial stage (S1) because of its "lack of a specified upper cut-off size limit." It should be noted that, in fact, S1 had a specified upper cut-off size here since a pre-separator was present in the NGI setup (Figure 1). Nonetheless, it was excluded to be consistent with the definition of ISM in previous publications.[6-9]

Figure 3



Reference (n = 30, blue) and test (n = 30, red) product NGI profiles.
Upper panel: 95% confidence intervals for mean deposition (all deposition stages).
Middle panel: 95% confidence intervals for mean deposition (ISM deposition stages).
Lower panel: 95% confidence intervals for mean deposition (ISM deposition stages after normalization).
ISM: impactor sized mass. NGI: Next Generation Impactor.
IP: induction port. PS: pre-separator. MOC: micro-orifice collector.

equivalence testing based on the mCSRS requires scaling on the variability of the reference product similar to the PBE test. Scaling on the variability can be understood as defining the acceptance criterion as a function of the reference variability.

The mCSRS can, in theory, be applied to any number of stages. It is recommended, however, that it is only applied to stages that are comprised in the definition of the ISM to avoid testing for potential differences on non-size measurement-related components. Before the mCSRS is applied, all TEST and REF profiles need to be normalized from absolute deposition (unit: mcg) into percent deposition (relative to those stages on which the mCSRS is applied). Table 2 gives an example for this procedure, based on randomly-selected REF and TEST samples from the pool of 60 profiles. For instance, the deposition of 25.1 mcg on stage S2 of the REF profile is divided by the total mass on all ISM stages of 88.7 mcg and a normalized deposition of 28.3% is obtained. After the normalization, the following stepwise procedure for APSD equivalence testing based on mCSRS can be applied.

First, the mean (normalized) REF profile needs to be obtained by simply calculating the average of all stages to which the mCSRS is applied (here: S2 – MOC). Then, a single mCSRS is calculated for a randomly-selected pair of one (normalized) TEST and one (normalized) REF profiles from the pool of all TEST and REF profiles. In the numerator of the mCSRS, the distance from the randomly-selected TEST profile to the mean REF profile is calculated. This is accomplished by, first, obtaining the distance for each of the stages separately (for instance, on S4 (30.2 – 29)**2/29 = 0.0497; Table 2) and, then, summing up

Table 1

Summary statistics of all deposition stages from a sample of 30 TEST and 30 REF Next Generation Impactor profiles.

Stage	TEST Mean (mcg)	REF Mean(mcg)	TEST sd(mcg)	REF sd(mcg)	TEST CV (%)	REF CV (%)
IP	145	148	44.8	35.1	31	23.8
PS	63.5	70.2	17.4	19.9	27.5	28.4
S1	25.7	26.3	3.61	4.15	14	15.8
S2	29.2	26	4.59	4.24	15.7	16.3
S3	20.3	19.3	2.84	3.38	14	17.5
S4	26.5	25.3	3.9	3.69	14.7	14.6
S5	12.6	11.4	2.03	1.88	16.1	16.4
S6	4.15	4.11	0.564	0.498	13.6	12.1
S7	0.814	0.735	0.511	0.562	62.8	76.4
MOC	0.253	0.351	0.275	0.34	109	96.8

IP: induction port PS: pre-separator MOC: micro-orifice collector
sd: standard deviation CV: coefficient of variation

the seven differences to a single number (1.038;Table 2). In the denominator of the mCSRS, the distance from the randomly-selected REF profile to the mean REF profile is similarly calculated and yields 2.42 (Table 2). The mCSRS yields 1.038/2.42 = 0.429 for this pair of TEST and REF profiles and thus, the difference from the randomly-selected REF profile to the REF mean profile is larger than that from the randomly-selected TEST profile to REF mean profile. This procedure is now repeated for all 900 different pairs of one TEST and one REF profile that can be obtained for the sample of 30 TEST and 30 REF profiles. Consequently, a distribution of 900 mCSRS is obtained from which the median is selected to represent the entire data set (here: 0.75). The median was selected as summary statistic over other measures (e.g., the mean) since it was shown to be independent of the shape and the number of stages to which the mCSRS is applied. The interpretation of the MmCSRS (median of the distribution of mCSRSs) of 0.75 is as follows. Since the MmCSRS is one when TEST and REF products are the same and can only be smaller than one when the TEST product is less variable (rewarding property of mCSRS), the TEST product is likely to be less variable than the REF product. On the other hand, values greater than one need to be interpreted more carefully, as this could indicate either a larger difference in means or a larger variability of the TEST product (penalizing; see below). The reference variance scaling becomes important to be able to distinguish these cases.

However, it is first necessary to calculate a confidence interval for the MmCSRS for equivalence testing. In contrast to ABE or PBE, where the data are assumed to be log-normally distributed and therefore parametric confidence intervals can be computed, the probability distribution of the MmCSRS is unknown. Thus, a non-parametric confidence interval is calculated by bootstrapping. A simplified description is that a new data set of 30 TEST and 30 REF products is obtained by re-sampling with replacement (meaning that a single profile could be included multiple times in the new data set) for which the MmCSRS is then calculated. This procedure is replicated 2,000 times and a distribution is obtained. The lower and upper 5% of the distribution of 2,000 MmCSRS are then excluded from further analysis and the so-called 90% bootstrapping percentile confidence interval is 0.521-0.948.

Next, an acceptance limit (or critical value) for the mCSRS-based APSD equivalence needs to be derived, incorporating the following criteria: regulatory-defined acceptance limit and variability of the REF product. A table of regulatory constants for difference acceptance limits is available elsewhere.[9] The variability of the REF product is calculated with the following formula that was determined as optimal for estimating the variability of the REF product:

Table 2

Numerical example for calculating the mCSRS for a pair of one TEST profile and one REF CI profile on the seven stages that comprise the definition of the impactor sized mass (ISM).

	Unit	S2	S3	S4	S5	S6	S7	MOC	ISM	Sum
REF	mcg	25.1	18.7	28.4	10.9	4.32	0.214	1.05	88.7	
TEST	mcg	29.5	22.4	29.6	10.7	5.14	0.174	0.388	97.9	
REF ISM Normalized	%	28.3	21.1	32	12.3	4.87	0.241	1.18	100	
TEST ISM Normalized	%	30.1	22.9	30.2	10.9	5.25	0.178	0.396	99.9	
REF Mean Normalized	%	29.8	22.1	29	13.1	4.71	0.843	0.403	100	
mCSRS Numerator		0.00302	0.029	0.0497	0.369	0.0619	0.525	0.00012		1.038
mCSRS Denominator		0.0755	0.0452	0.31	0.0489	0.00544	0.43	1.5		2.42

IP: induction port PS: pre-separator MOC: micro-orifice collector

$$\sqrt{\frac{\sum_{i=1}^{p} \bar{R}_i \bullet CV_i^2}{\sum_{i=1}^{p} \bar{R}_i}}$$

where CV_i represents the coefficient of variation (in %) of the i^{th} deposition stage of the REF sample (after normalization on ISM); R_i and p were defined above. The so-called reference variance scaling metric is 17.3 here.

For acceptance limits of 10%, 15%, 20%, 25% and 30%, the critical values are 1.41, 1.95, 2.74, 3.77, and 5.05, respectively. The equivalence test is then performed by comparing the upper bound of the confidence interval (here 0.948) with the respective critical values. Since the critical values are all larger than 0.948, both products would be concluded to be equivalent in their ISM profiles. However, since the stepwise procedure for APSD equivalence testing includes performing a PBE test on single actuation content (or total mass as a surrogate) and ISM in addition to the MmCSRS analysis, the overall conclusion would be non-equivalent due to the fail for PBE on total mass. However, using the ABE test, equivalence in total mass and ISM would be concluded.

Conclusions

In this article, the proposed stepwise cascade impactor equivalence test (Figure 2) was applied to a practice data set consisting of TEST and REF NGI profiles for illustrating some important computational aspects of this procedure. The practice data set was simulated from two products with a maximum mean difference on single stages of 5% and the TEST product being 20% less variable than the REF product. With the exception of failing the PBE test on total mass, the proposed stepwise cascade impactor equivalence test would conclude equivalence in APSD of the two products.

"Average bioequivalence" test on single stages

For the ABE approach, 90% confidence intervals for the TEST-to-REF ratio are calculated while assuming equal variances. The lower and upper limits of the confidence intervals are then compared with regulatory acceptance limits (e.g., +/- 10%). The data are log-transformed prior to the analysis to allow calculation of parametric confidence intervals (point estimate +/- 1.645* standard error). Table 3 shows that for most of the stages, the equivalence conclusion depends on the applied acceptance limit and number of samples (increasing the sample size would lead to narrower confidence intervals). For S7 and MOC, however, a pass for any reasonable acceptance limits and number of samples is unlikely given the wide confidence intervals and the point estimates of 1.12 and 0.72, respectively, (see below) thus clearly demonstrating a potential drawback of performing the ABE approach on single stages.

"Population bioequivalence" test on single stages

First, the results of a PBE test are not as tangible as those of the ABE test. Without going into detail, the PBE test considers differences in the means and the variability between two products. While it can theoretically be applied for any setup, the US Food and Drug Administration (FDA) budesonide draft guidance refers to a difference of 10% in means and a reference vari-

ability of 0.1 for scaling. It should be noted that this scaling is similar to that applied for the mCSRS (see main article). A rather complex mathematical operation is applied for calculating a criterion that is used for decision making (a criterion smaller than zero implies equivalence). The results of applying the PBE test on the single stages are shown in Table 3. Interestingly, on all stages other than the induction port, equivalence would be concluded thus demonstrating a key feature of the PBE test. Namely, the TEST product gets rewarded for being less variable than the REF product on stages S7 and MOC and it gets punished for being more variable on the induction port (see standard deviations in Table 1). More details on this rewarding/penalizing property are given in this article in the section on the mCSRS.

It should be mentioned that performing ABE and PBE tests on single stages usually requires a multiplicity correction[c] for performing multiple statistical tests that was not included here.

[c] A multiplicity correction ensures that the overall experiment error rate is maintained (e.g., at most a 5% risk of concluding equivalence when the products are, indeed, not equivalent).

Table 3

Results of applying average "bioequivalence" (ABE) and population "bioequivalence" (PBE) test on each of the ten deposition stages.

Stage	ABE 90% CI	PBE Criterion
IP	0.836-1.1	0.0137
PS	0.796-1.05	-0.141
S1	0.916-1.04	-0.0229
S2	1.05-1.2	-0.00617
S3	0.988-1.13	-0.0376
S4	0.983-1.12	-0.0128
S5	1.03-1.19	-0.00943
S6	0.953-1.07	-0.00802
S7	0.776-3.28	-5.35
MOC	0.19-2.19	-7.44

IP: induction port
PS: pre-separator
MOC: micro-orifice collector
CI: confidence interval

References

1. Lee SL, Adams WP, Li BV, Conner DP, Chowdhury BA, Yu LX. *In vitro* considerations to support bioequivalence of locally acting drugs in dry powder inhalers for lung diseases. The AAPS Journal. 2009;11(3):414-23.

2. Guideline on the requirements for clinical documentation for orally inhaled products (OIP) including the requirements for demonstration of therapeutic equivalence between two inhaled products for use in the treatment of Asthma and Chronic Obstructive Pulmonary Disease (COPD) in adults and for use in the treatment of asthma in children and adolescents: Committee for Medicinal Products for Human Use (CHMP); 2009 [Nov 5, 2014]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003504.pdf.

3. Draft Guidance on Fluticasone Propionate; Salmeterol Xinafoate: Food and Drug Administration; 2013 [November 5, 2014]. Available from: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm367643.pdf. %20Last%20accessed%20Jul%202014.

4. Draft Guidance on Budesonide Food and Drug Administration; [November 27, 2014]. Available from:http://www.fda.gov/downloads/Drugs/.../Guidances/UCM319977.pdf.

5. Garcia-Arieta A. A European perspective on orally inhaled products: *In vitro* requirements for a biowaiver. Journal of Aerosol Medicine and Pulmonary Drug Delivery. 2014;27(6):419-29.

6. Weber B, Adams W, Lionberger R, Li B, Tsong Y, Hochhaus G, et al. Evaluation of statistical methods for determining equivalence of aerodynamic particle size distribution. In: Dalby RN, Byron PR, Peart J, Suman JD, Farr SJ, Young PM, editors. Respiratory Drug Delivery 2012. Richmond, VA: Virginia Commonwealth University; 2012. p. 803-8.

7. Weber B, Hochhaus G, Adams W, Lionberger R, Li B, Tsong Y, et al. A stability analysis of a modified version of the chi-square ratio statistic: Implications for equivalence testing of aerodynamic particle size distribution. The AAPS Journal. 2013;15(1):1-9.

8. Weber B, Lee SL, Lionberger R, Li BV, Tsong Y, Hochhaus G. A sensitivity analysis of the modified chi-square ratio statistic for equivalence testing of aerodynamic particle size distribution. The AAPS Journal. 2013;15(2):465-76.

9. Weber B, Lee SL, Delvadia R, Lionberger R, Li BV, Tsong Y, et al. Application of the modified chi-square ratio statistic in a stepwise procedure for cascade impactor equivalence testing. The AAPS Journal. 2015; 17(2): 370-9.

10. Statistical Approaches to Establishing Bioequivalence: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER); [08 June 2014]. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070244.pdf.

11. Realistic Scenarios 1-33 of 55: PQRI APSD Profile Comparisons Working Group; [08 June 2014]. Available from: http://www.pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/Addl/Realistic%20Scenarios%201-33%20of%2055.pdf.

12. Realistic Scenarios 34-55 of 55: PQRI APSD Profile Comparisons Working Group; [08 June 2014]. Available from: http://www.pqri.org/commworking/minutes/pdfs/dptc/psdpcwg/Addl/Realistic%20Scenarios%2034-55%20of%2055.pdf.

13. Adams WP, Christopher D, Lee DS, Morgan B, Pan Z, Singh GJ, et al. Product Quality Research Institute evaluation of cascade impactor profiles of pharmaceutical aerosols, Part 1: Background for a statistical method. AAPS PharmSciTech. 2007;8(1):E32-E7.

14. Christopher D, Adams W, Amann A, Bertha C, Byron PR, Doub W, et al. Product Quality Research Institute evaluation of cascade impactor profiles of pharmaceutical aerosols, Part 3: Final report on a statistical procedure for determining equivalence. AAPS PharmSciTech. 2007;8(4):65-74.

15. Christopher D, Adams WP, Lee DS, Morgan B, Pan Z, Singh GJ, et al. Product Quality Research Institute evaluation of cascade impactor profiles of pharmaceutical aerosols: Part 2 - Evaluation of a method for determining equivalence. AAPS PharmSciTech. 2007;8(1):E39-E48.

Benjamin Weber, MStat, PhD is Senior Pharmacometrician, Translational Medicine and Clinical Pharmacology, Boehringer Ingelheim, Biberach an der Riss, Germany, Tel: +49 7351 54143520, benjamin.weber@boehringer-ingelheim.com, Website: www.bweber.net.