JMP for Biosimilars: Tools for Analytical Similarity

W. Heath Rushing, Principal Consultant and Co-founder, Adsurgo LLC
Andrew T. Karl, Senior Management Consultant, Adsurgo LLC
Richard Burdick, Statistical Consultant, Elion Labs

1. Introduction

Generic pharmaceutical drugs have been available in the U.S. market since 1984. The Center for Drug Evaluation and Research (CDER) defines a generic as a "drug product that is comparable to a brand/reference listed drug product in dosage form, strength, route of administration, quality and performance characteristics, and intended use." Generics can be licensed once a brand *chemical* drug (referred to as a *small molecule* drug) is off patent; however, they must contain the same active product ingredients and dosage form. The Hatch-Waxman Act of 1984 provided a licensure pathway for generics; pre-clinical and clinical trials do not have to be repeated for drugs approved after 1962. The generic needs to be shown to have comparable quality and effectiveness as the brand drug. The common approach is to show statistical equivalence for quality and performance characteristics such as potency, purity, and stability (http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/SmallBusinessAssistance/ucm127615.pdf).

Biological products are drugs that are developed from living organisms: these drugs were patented by companies starting in the 1982. These products contain proteins or other complex structures that are at risk of degradation throughout the manufacturing process. The additional structural complexity in biologic products creates additional challenges when defining targets for potency, purity, and stability in these products (Feroz *et al.* 2015). Until recently, there has been no similar pathway for biological generics.

The Biologics Price Competition and Innovation Act (BPCI) of 2009 created an abbreviated licensure pathway for biological products shown to be similar or "biosimilar" to a pre-existing FDA-licensed product. The BPCI Act was an amendment [section 351(k)] to the Public Health Service (PHS) Act. Section 351(k) of the PHS Act defines biosimilar as "highly similar to the reference product notwithstanding minor differences between the biological product and the reference product in terms of safety, purity, and potency of the product [FDA (2015) Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, Guidance for Industry, United States FDA, Silver Spring, Maryland, USA]."

Although there are currently 46 biosimilars approved in markets worldwide (http://www.biosimilarz.com/?page_id=242), there is only one biosimilar approved by the U.S. FDA. Sandoz, Inc. (a Novartis company) received approval for Zarxio, a biosimilar to Amgen's Neupogen, on Mar 6, 2015. Amgen originally filed the biologics licensing agreement (BLA) for Neupogen on Feb 20, 1991. Zarxio received approval for all five indications associated with Neupogen (http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/SmallBusinessAssistance/ucm127615.pdf).

It is estimated there are at least 280 biosimilars in the pipeline and clinical trials are growing at a rate of 20% per year (http://www.biopharma-reporter.com/Markets-Regulations/Where-are-tomorrow-s-biosimilar-hotspots). This paper

outlines and demonstrates the current regulatory thoughts in this emerging area. Section 2 describes the current FDA thinking on the subject and explains their recommended approach to establishing criteria for analytical similarity. Section 3 demonstrates the recommended approach, the sample size and variance adjusted margin method, using a JMP script. Section 4 outlines future considerations for the current FDA approach.

2. Current FDA Thinking/Approach

In 2012, the FDA issued three (draft) guidance documents on the subject of biosimilars:

- FDA (2012a) Scientific considerations in demonstrating biosimilarity to a reference product. FDA, Silver Spring, Maryland, USA.
- FDA (2012b) Quality considerations in demonstrating biosimilarity to a reference protein product. FDA, Silver Spring, Maryland, USA.
- FDA (2012c) Biosimilars: Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009. FDA, Silver Spring, Maryland, USA.

In Scientific considerations in demonstrating biosimilarity to a reference product, the FDA introduces the concept of *analytical similarity*. To demonstrate biosimilarity, the FDA recommends applicants use a stepwise approach for comparison of the biosimilar to the reference product, referred to as the *totality of evidence* approach. This approach includes comparison of the biosimilar to reference products with respect to: structure, function, animal toxicity, human pharmacokinetics (PK) and pharmacodynamics (PD), clinical immunogenicity, and clinical safety and effectiveness. It starts with extensive evaluation of critical-to-quality attributes (CQAs) associated with the structural and functional characterization of both the biosimilar and reference product. This extensive evaluation serves as the foundation of a biosimilar development program and provides a means of establishing *analytical similarity*. Analytical similarity requires the biosimilar to be "highly similar to the reference product notwithstanding minor differences in clinical inactive components (Tsong *et al.,* 2015)." The term analytical similarity is used due to the heavy reliance on analytical methods for comparison of CQAs between the reference product and biologic (Tsong *et al.,* 2015).

Analytical similarity begins by identifying the CQAs. Each CQA is then classified into three tiers of criticality: Tier 1, 2, and 3. Analytical similarity is measured for each of the CQAs using varying degrees of statistical rigor, depending on the level of criticality assigned to the attribute (Chow, 2014).

Tier 1 contains the most critical attributes and makes use of a statistical equivalence test in assessing analytical similarity. Tier 2 contains mildly to moderately critical attributes, using a quality-range approach to verify that a sufficient number of samples from the proposed biosimilar fall within a fixed confidence interval of the established attribute mean for the reference product. Tier 3 holds the least critical outcomes, employing graphical methods to verify similar performance across products (Tsong *et al.,* 2015).

The focus of this paper is on the set of the most critical attributes contained in Tier 1. Due to their criticality, these CQAs are subject to more intense scrutiny than attributes in other tiers. A statistical equivalence test is used comparisons of Tier 1 CQAs associated with the proposed biologic to the reference product. Essentially, a confidence interval is formed around the difference between the attribute means across products and compared to an equivalence criterion represented by the range between two constants δ_1 and δ_2. Typically, $\delta_2>0$, $\delta_1 = -\delta_2$ and δ_2 is called the equivalence margin. If the entire confidence interval falls between δ_1 and δ_2, the attribute is deemed to be similar across products. Figure 1 shows attributes that are statistically equivalent in green; those attributes that are not statistically equivalent in red.

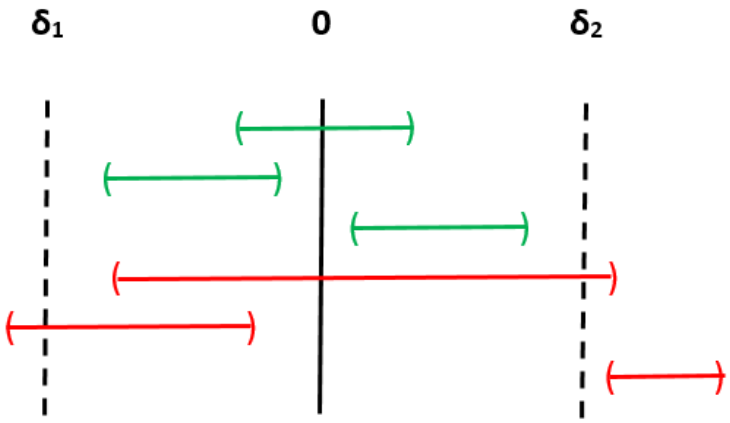


Figure 1. Graphical representation of a statistical equivalence test.

There are several choices available in the implementation of the equivalence test, including the method for the selection of the equivalence margin, the sample sizes of the proposed and referenced product, the test size (confidence level for the confidence interval), and the assumptions made when estimating the attribute variance from the samples. The most impactful of these is the selection of the equivalence margin.

One existing method is to simply fix a margin (e.g. 0.9 to 1.1) about the proportion of the sampled means, regardless of the variability observed in the sample. However, the size of the confidence interval about the difference in means depends on both the sample size and the estimated variance. As a result, the power of such an equivalence test decreases with smaller sample sizes, leading sponsors to propose overly wide margins in order to increase a product's likelihood of passing when only a small number of samples are used (Tsong *et al.*, 2015).

Alternatively, the Limentani approach uses a margin that is tied to the estimated variability *and* to the sample sizes: the margin increases as the sample size decreases in order to keep the power of the test constant across sample size. However, this approach tends to reward the use of smaller sample sizes, yielding a margin that can be up to five standard deviations wide (Tsong *et al.*, 2015).

At this time, the proposed FDA method is to use the *sample size and variance adjusted margin*. This approach gives the ability to control power for small sample sizes while still rewarding larger sample sizes. The equivalence margin is formed as a fixed multiple (1.5 is used in practice) of the sample reference product standard deviation. (Note that this fixed multiple is not tied to the sample size.) The

sample size and variance adjusted margin method for developing acceptance criteria has three basic steps:

- Step 1: Determine the variability of reference product using the standard deviation (s).
- Step 2: Calculate statistical equivalence acceptance criteria (EAC) as ($\pm c * s$), where c = 1.5.
- Step 3: Compute a 90% confidence interval on the difference in means between the two products. Compare against acceptance criteria (margin).

However, there are several possible problematic factors to consider in the implementation of this method. How many reference lots are available? Should reference lots come from different markets and different dosage forms? How many lots of each product should be used to estimate the standard deviation? How many lots should be used to estimate the difference across products and associated 90% confidence interval? Should the variances of the attribute in each population be assumed equal? **Note: From results of power calculations for varying sample sizes (for both reference and biosimilar lots) and test sizes (confidence levels), the FDA recommends a value for *c* of 1.5 because it provides a reasonable acceptable power when the difference in means is equal to *s*/8.**

The implementation of the method depends on the number of reference versus the number of biosimilar lots used in the comparison:

- The total number of reference lots are less than or equal to the total number of biosimilar lots.
- The total number of reference lots are greater than the total number of biosimilar lots.

If the total number of reference lots are less than or equal to the total number of biosimilar lots:

1. Let N_r denote the total number of reference lots and N_b denote the number of biosimilar lots where $N_r \leq N_b$.
2. All reference lots (N_r) and all biosimilar lots (N_b) will be used to calculate the difference between mean attribute measurements between the two products, and the associated 90% confidence interval.
3. All reference lots (N_r) are used to estimate the standard deviation (s) of the reference material. The test assumes that the attribute variance is constant across products.
4. Calculate the acceptance criterion as $\pm 1.5 * s$.
5. Calculate a 90% confidence interval for a difference in means using the data from step 2.
6. Compare the 90% confidence interval to the acceptance criteria. The null hypothesis (products differ with respect to the attribute) is rejected if the endpoints of the confidence interval lie within the acceptance margin.

(Dong, 2015)

In its current implementation, it is expected that the total number of reference lots measured will be greater than the number of biosimilar lots. This is primarily due to the greater availability of reference product samples: the reference product is available for purchase and has an established large-scale manufacturing process, whereas the biosimilar product will likely be produced at a higher cost in a laboratory-scale setting. Using the total number of reference lots and biosimilar lots:

1. Let N_r denote the total number of reference lots, and N_b denote the number of biosimilar lots where $N_b < N_r$. The difference between the lot sizes will be represented by $m_r = N_r - N_b$.
2. A random subset (of size N_b) of the total number of reference lots (N_r) and all N_b biosimilar lots will be used to calculate the difference between mean attribute measurements between the two products, and the associated 90% confidence interval.
3. Use the remaining number of reference lots (m_r) to estimate the standard deviation (s) of the reference material. The test assumes that the attribute variance is constant across products.

4. Calculate the acceptance criterion as ± 1.5 * *s*.
5. Calculate a 90% confidence interval for a difference in means using the data from step 2.
6. Compare the 90% confidence interval to the acceptance criteria. The null hypothesis (products differ with respect to the attribute) is rejected if the endpoints of the confidence interval lie within the acceptance margin.

(Chow, 2014)

The power of this test can be calculated for a given number of reference and biosimilar lots, and test size (α) (note that a test of size α will construct a (1-2* α) confidence interval). This power calculation uses the FDA-recommended acceptable difference (*s*/8) and value for *c* (*c* = 1.5). Assuming the variances of the biosimilar and reference products are equal, and letting $\Im_d(\cdot\,|p)$ represent the CDF of the non-central *t* distribution with non-centrality parameter p, the power of the sample size and variance adjusted margin equivalence test is the maximum of 0 and

$$\Im_{N_b+N_r-2}\left(\frac{1.5}{\sqrt{\frac{1}{N_b}+\frac{1}{N_r}}}-t_{N_b+N_r-2}(\alpha)\middle|\frac{1}{8\sqrt{\frac{1}{N_b}+\frac{1}{N_r}}}\right)$$
$$-\Im_{N_b+N_r-2}\left(-\frac{1.5}{\sqrt{\frac{1}{N_b}+\frac{1}{N_r}}}+t_{N_b+N_r-2}(\alpha)\middle|\frac{1}{8\sqrt{\frac{1}{N_b}+\frac{1}{N_r}}}\right)$$

when all N_r reference lots are compared to the N_b biosimilar lots.

When only a subset of N_b of the N_r reference lots is used for the comparison of means, the power is the maximum of 0 and

$$\Im_{2N_b-2}\left(\frac{1.5}{\sqrt{\frac{2}{N_b}}}-t_{2N_b-2}(\alpha)\middle|\frac{1}{8\sqrt{\frac{2}{N_b}}}\right)$$

3. Demonstration of JMP

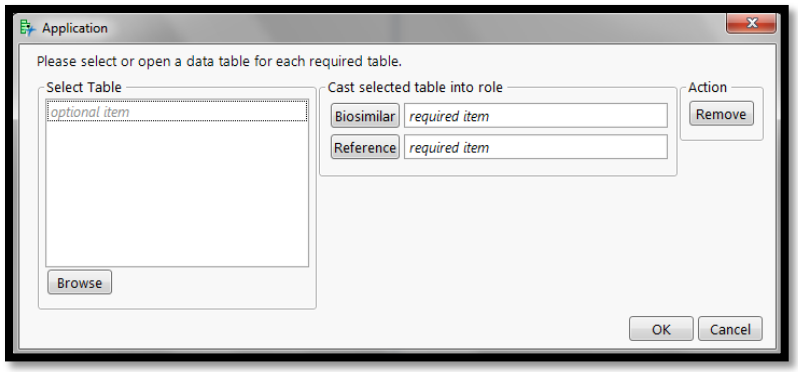
This demonstration evaluates multiple scenarios:

- Bioequivalence should be established when the number of reference lots and biosimilar lots are equal.
- Bioequivalence should not be established when the number of reference lots and biosimilar lots are equal.
- Bioequivalence should be established when the number of reference lots is greater than the number of biosimilar lots.
- Bioequivalence should not be established when the number of reference lots is greater than the number of biosimilar lots.

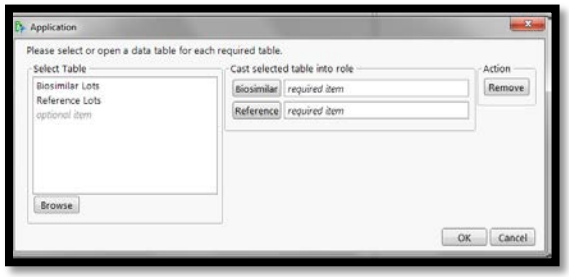
In addition, we evaluated the difference in power between the two implementations: when all reference lots are used to estimate the standard deviation for the EAC versus when a subset (random sample) is used to estimate the standard deviation for the EAC. It is shown the power is less for the second scenario even when the sample size is greater.

In the first scenario, 10 reference lots are generated from a distribution that is Normal(100,10) while 10 biosimilar lots are generated from a distribution that is Normal(101,10). Bioequivalence should be established.

1. Open the script.

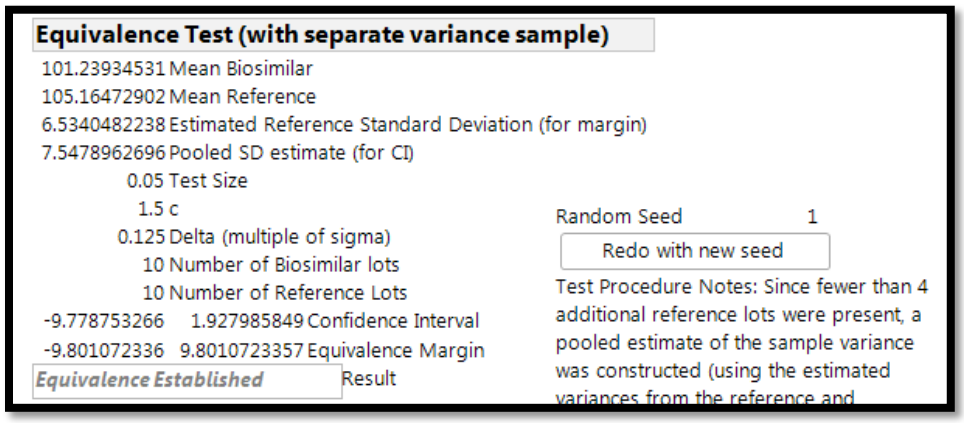


2. Select the data set containing the potency measurements from both the reference and biosimilar lots.



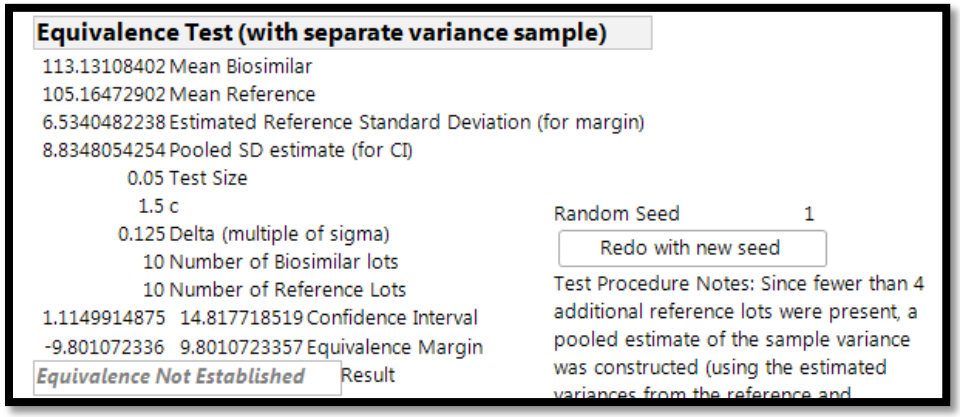
3. Select the data table containing a column of quality attribute measurements from the biosimilar and reference lots.

4. Select OK.



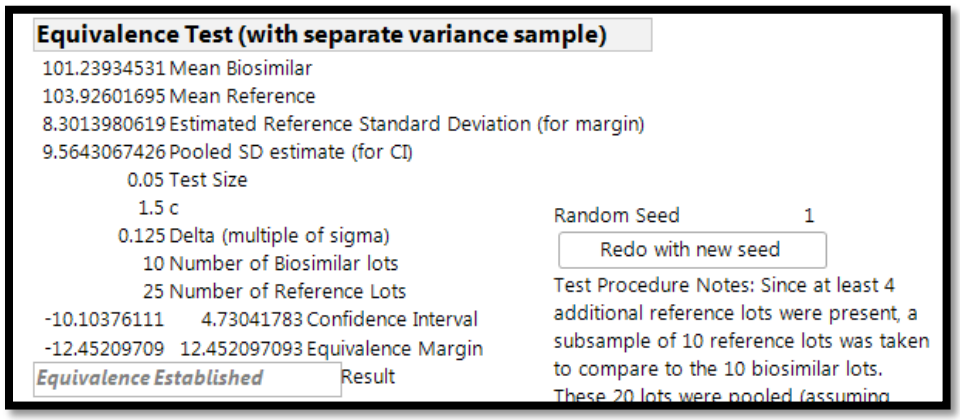
The 90% confidence interval for the difference in means is (-9.78, 1.93) while the equivalence margin is ±9.80. Therefore, **statistical equivalence is established**.

In the second scenario, 10 reference lots are generated from a distribution that is Normal(100,10) while 10 biosimilar lots are generated from a distribution that is Normal(110,10). Bioequivalence should not be established.



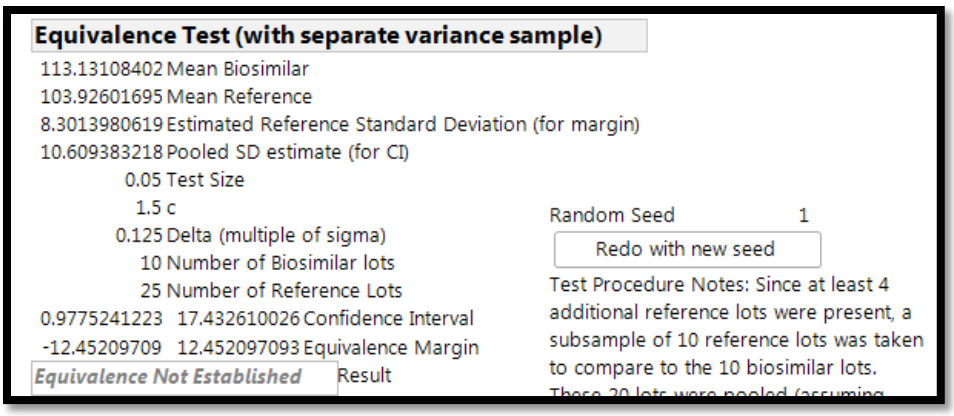
The 90% confidence interval for the difference in means is (1.11, 14.82) while the equivalence margin is ±9.80. Therefore, **statistical equivalence is not established**.

In the third scenario, 25 reference lots are generated from a distribution that is Normal(100,10) while 10 biosimilar lots are generated from a distribution that is Normal(101,10). Bioequivalence should be established.



The 90% confidence interval for the difference in means is (-10.10, 4.73) while the equivalence margin is ±12.45. Therefore, **statistical equivalence is established**.

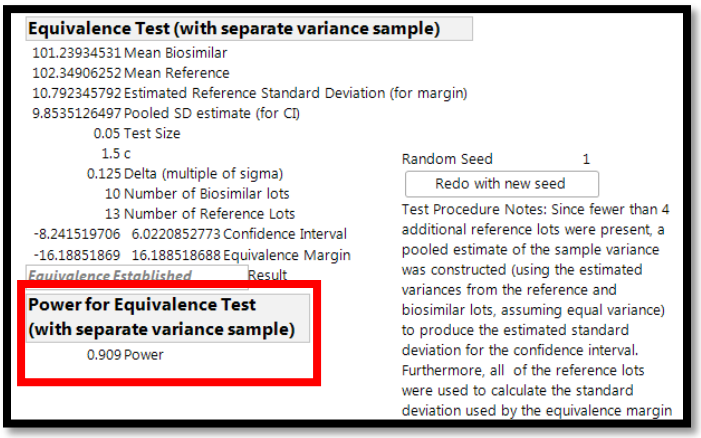
In the third scenario, 25 reference lots are generated from a distribution that is Normal(100,10) while 10 biosimilar lots are generated from a distribution that is Normal(110,10). Bioequivalence should not be established.



The 90% confidence interval for the difference in means is (0.98, 17.43) while the equivalence margin is ±12.45. Therefore, **statistical equivalence is not established**.

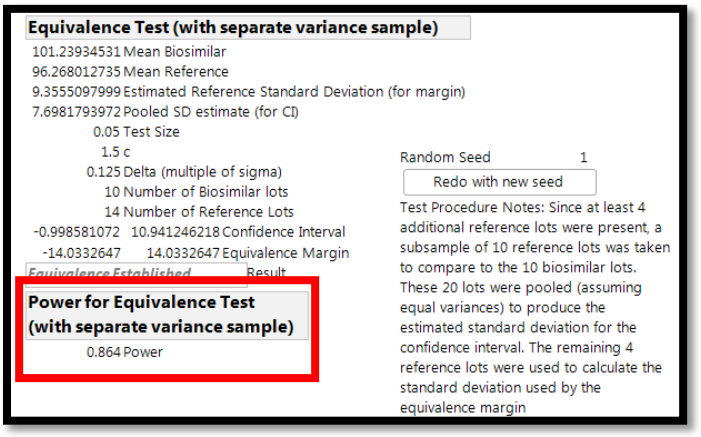
Additionally, the power of the two different methods for estimating the standard deviation were evaluated: using all reference lots for the estimate of the standard deviation (for the EAC) versus when a subset (random sample) of reference lot are used.

In the first scenario, 13 reference lots are generated from a distribution that is Normal(100,10) while 10 biosimilar lots are generated from a distribution that is Normal(101,10). Although the current FDA-recommended approach calls for a subset of size 3 to be used to estimate the standard deviation for the EAC, all reference lots are used. This was done for a comparison of power between the two methods.



Note the power of this method is 0.909.

In the second scenario, 14 reference lots (increased sample size) are generated from a distribution that is Normal(100,10) while 10 biosimilar lots are generated from a distribution that is Normal(101,10). This demonstration uses the current FDA-recommended approach: take subset of size 4 to estimate the standard deviation for the EAC.



Note, although we use an increased number of reference lots, the power of this method (0.864) is lower than if we used all reference lots for estimation of the standard deviation.

This identifies one of the multiple issues with the current FDA-recommended approach.

4. Future considerations

The FDA's current approach has been outlined and demonstrated using a JMP script. However, this method has only been evaluated on one biosimilar to date. Both the FDA and industry experts recognize there are several of issues with the current approach that should be addressed.

At a the recent DIA/FDA Statistics Forum, Yi Tsong and the OB CMC Analytical Biosimilar Method Development Team presented multiple future considerations for analytical biosimilarity. They would like consistent comparisons across submissions for the same reference product and adjustments to power for both small sample sizes and large variance. They also discussed:

- Development of a more appropriate test (than a *t*-test) since the approach uses an estimate of the standard deviation (*s*) to define the equivalence margin.
- Development of multiple approaches to be based on number of reference lots (some procedures better for small samples, others are better for large samples).
- An adjustment to the acceptance criteria (margin) based upon a constant shift.
- How to determine margin when equivalence uses ratio (versus difference).

In addition, two industry experts from Amgen – Rick Burdick and Jose Ramirez – discussed future considerations for the methods used for comparison for all three tiers of (quality attribute) criticality. For Tier 1 criticality attributes, they discussed the use of all reference lots for 90% confidence interval if acceptance criteria set using scientific knowledge. Additionally, they proposed developing the equivalence margin based on effect size (the difference in product means divided by the reference product standard deviation) directly, and then use a confidence interval on the effect size. Lastly, they explained how correlated reference lots will require the value of *c* to be increased to maintain the desired power (Burdick and Ramirez, 2015).

The currently proposed method has many future development opportunities in the emerging area of biosimilar analysis. It is possible that specific methods may be developed or adjusted based upon the particular application of the test for analytical similarity. The details of the testing procedure will likely be fine-tuned in future revisions of the FDA guidance documents based on findings from the ongoing research discussed in this section. JMP offers an extremely flexible basis for quickly updating and maintaining a GUI front-end for practitioners through its Application Builder and JMP scripting language (JSL).

5. References

1. Burdick, Richard K. and José G. Ramirez. *Statistical Issues in Biosimilar Analytical Assessment: Perspectives on FDA ODAC Analysis* [Powerpoint]. DIA/FDA Statistics Forum 2015.
2. Chow SC (2014) *On Assessment of Analytical Similarity in Biosimilar Studies*. Drug Des 3: 119. doi:10.4172/2169-0138.1000e124.
3. Dong, Xiaoyu (January 13, 2015) *Statistical Revew and Evaluation of BLA No. 125553*, Center for Drug Evaluation and Research, United States FDA, Silver Spring, Maryland, USA.
4. FDA (2012a) Scientific considerations in demonstrating biosimilarity to a reference product. United States FDA, Silver Spring, Maryland, USA.
5. FDA (2012b) Quality considerations in demonstrating biosimilarity to a reference protein product. United States FDA, Silver Spring, Maryland, USA.
6. FDA (2012c) Biosimilars: Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009. United States FDA, Silver Spring, Maryland, USA.
7. FDA (2015) Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, Guidance for Industry, United States FDA, Silver Spring, Maryland, USA.
8. Feroz, J.; Hershenson, S.; Khan, M. A.; Martin-Moe, S. (2015) *Quality by Design for Biopharmaceutical Drug Product Development*. Springer: New York.
9. Tsong, Yi and OB CMC Analytical Biosimilar Method Development Team (Meiyu Shen, Cassie Xiaoyu Dong). *Development of Statistical Approaches for Analytical Biosimilarity Evaluation* [Powerpoint]. DIA/FDA Statistics Forum 2015.

Websites:

1. http://www.biopharma-reporter.com/Markets-Regulations/Where-are-tomorrow-s-biosimilar-hotspots
2. http://www.biosimilarz.com/?page_id=242
3. http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/SmallBusinessAssistance/ucm127615.pdf [Powerpoint].
4. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm436648.htm