Statistical Approaches to Evaluate Analytical Similarity

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Scott N. Goldie at 301-796-2055 or (CBER) Office of Communication, Outreach and Development at 800-835-4709 or 240-402-8010.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

September 2017
Biosimilars

Statistical Approaches to Evaluate Analytical Similarity

Guidance for Industry

Additional copies are available from:

Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Tel: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm

and/or

Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Email: ocod@fda.hhs.gov
https://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm
(Tel) 800-835-4709 or 240-402-8010

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

September 2017
Biosimilars

TABLE OF CONTENTS

Statistical Approaches to Evaluate Analytical Similarity Guidance for Industry

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

I. INTRODUCTION

This guidance is intended to provide advice on the evaluation of analytical similarity to sponsors interested in developing biosimilar products for licensure under section 351(k) of the Public Health Service Act (PHS Act) (42 U.S.C. 262(k)). This evaluation is to support the demonstration that a proposed biosimilar product (hereinafter *proposed biosimilar* or *biosimilar*) is highly similar to a reference product licensed under section 351(a) of the PHS Act. Specifically, this guidance describes the type of information a sponsor of a proposed biosimilar product should obtain about the structural/physicochemical and functional attributes of the reference product, how that information is used in the development of an analytical similarity assessment plan for the proposed biosimilar, and the statistical approaches recommended for evaluating analytical similarity.

This guidance is one in a series of guidance documents that FDA is developing or has developed to implement the Biologics Price Competition and Innovation Act of 2009 (BPCI Act). It serves as a companion document to the guidance for industry *Quality Considerations in Demonstrating Biosimilarity of a Therapeutic Protein Product to a Reference Product*.[1]

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidance documents describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in FDA guidance documents means that something is suggested or recommended, but not required.

II. BACKGROUND AND SCOPE

The BPCI Act created an abbreviated licensure pathway under section 351(k) of the PHS Act (42 U.S.C. 262(k)) for biological products shown to be biosimilar to or interchangeable with an U.S.-

[1] We update guidances periodically. To make sure you have the most recent version of a guidance, check the FDA Drugs guidance web page at https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm, or the CBER guidance web page at https://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/default.htm.

licensed biological reference product (see sections 7001 through 7003 of Public Law 111-148). Section 351(i) of the PHS Act defines *biosimilarity* to mean "that the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components" and that "there are no clinically meaningful differences between the biological product and the reference product in terms of the safety, purity and potency of the product."[2] A 351(k) application for a proposed biosimilar product must include information demonstrating biosimilarity based on data derived from, among other things, "analytical studies that demonstrate that the biological product is highly similar to the reference product notwithstanding minor differences in clinically inactive components."[3]

Since the passage of the BPCI Act in 2009, FDA has released a number of guidance documents on demonstrating biosimilarity, including the guidances for industry *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product* (final issued in 2015) and *Quality Considerations in Demonstrating Biosimilarity of a Therapeutic Protein Product to a Reference Product* (final issued in 2015). Based on the statutory definition of *biosimilarity*, these guidance documents are intended (1) to assist sponsors in demonstrating biosimilarity for submitting a marketing application under section 351(k) of the PHS Act and (2) to describe FDA's current thinking on scientific principles to be considered in determining biosimilarity. Specifically, in the *Scientific Considerations in Demonstrating Biosimilarity to a Reference Product* guidance for industry, FDA described the totality-of-the-evidence approach that FDA would use in the review of biosimilar applications. The results of statistical analyses conducted to support a demonstration that a proposed product is "highly similar" to U.S.-licensed reference product (hereinafter the *reference product* or the *U.S.-licensed reference product*) are considered within the context of totality-of-the-evidence in determining if a proposed product is biosimilar to a reference product. The *Quality Considerations in Demonstrating Biosimilarity of a Therapeutic Protein Product to a Reference Product* guidance for industry describes the Agency's recommendations to sponsors on the scientific and technical information (including analytical studies to support a demonstration that a proposed biosimilar is highly similar to the reference product), for the chemistry, manufacturing, and controls (CMC) section of a marketing application for a proposed product submitted under section 351(k) of the PHS Act.

The objective of this guidance is to assist sponsors in demonstrating, through an evaluation of the analytical similarity of the proposed biosimilar and reference product, that the proposed biosimilar and reference product are highly similar to support licensure under section 351(k) of the PHS Act. In general, an analytical similarity assessment involves a comparison of structural/physicochemical and functional attributes using multiple lots of the proposed biosimilar product and the reference product.

Conducting appropriate statistical analyses in the evaluation of analytical similarity can provide a high degree of confidence in the results and reduce the potential for bias. However, there are many challenges in designing the statistical analyses to be performed. First, there may be a limited number of reference product lots, and those obtained may be the result of biased sampling, leading to imprecise and possibly inaccurate estimates of the distributions of important quality attributes for the

[2] Section 7002(b)(3) of the Affordable Care Act, adding section 351(i)(2) of the PHS Act (42 U.S.C. 262(i)(2)).

[3] Section 351(k)(2)(A)(i)(I)(aa) of the PHS Act (42 U.S.C. 262(k)(2)(A)(i)(I)(aa)).

reference product. Second, there may also be a limited number of proposed biosimilar lots, and the available lots may not reflect the true variability of biosimilar product manufacturing. Third, there are a large number of potential quality attributes that can be compared in an evaluation of analytical similarity, and subjecting all of these attributes to formal statistical tests in the context of limited lots could lead to concluding incorrectly that a large number of truly highly similar products are not highly similar.

To address these challenges, the Agency recommends using a risk-based approach in the analytical similarity assessment of quality attributes. This approach to the evaluation of analytical similarity consists of several steps. The first step is a determination of the quality attributes that characterize the reference product in terms of its structural/physicochemical and functional properties. In the second step, these quality attributes are then ranked according to their risk of potential clinical impact. Third, these attributes/assays are evaluated according to one of three tiers of statistical approaches based on a consideration of risk ranking as well as other factors. It should be noted, however, that some attributes may be important but not amenable to quantitative evaluation.

This guidance is not intended to describe the Agency's expectations for determining the adequacy of similarity for initiating clinical studies in a biosimilar development program, nor is it intended to describe the expectations for developing the manufacturing control strategy.

The document is structured as follows: Section III describes the quantity and quality of both reference product and biosimilar lots that we generally believe are scientifically necessary for evaluating analytical similarity; Section IV describes general principles for the evaluation of analytical similarity, including the use of a risk assessment to rank attributes and a tiered approach to the evaluation of analytical similarity.

III. REFERENCE AND BIOSIMILAR PRODUCTS

The Agency recommends that the analytical similarity evaluation begin with an understanding of the structural/physicochemical and functional attributes of the reference product. Based on information obtained about these attributes during development of the proposed biosimilar, the sponsor should develop an analytical similarity assessment plan (see section IV.A). A key component of this plan is the description of lots available for similarity testing. The following factors should be considered when selecting lots to be used in the analytical similarity assessment:

- Number of Reference Product Lots - To establish meaningful similarity acceptance criteria, sponsors should acquire a sufficient number of reference product lots. We recommend a minimum of 10 reference product lots be sampled. In cases where limited numbers of reference product lots are available (e.g., for certain orphan drugs), alternate analytical similarity assessments should be proposed and discussed with the Agency.

- Number of Biosimilar Product Lots - To allow for meaningful comparisons, we recommend a minimum of 10 biosimilar lots be included in the analytical similarity assessment.

- Variability in Reference Product Lots - The reference product lots selected should represent the variability of the reference product. Lots with remaining expiry spanning the reference product shelf life should be selected. The date of the analytical testing as well as the product expiration date should be provided in the application. Expired reference product should not be included in the similarity assessment to avoid bias.

- Accounting for Reference Product and Biosimilar Product Lots - Sponsors should account for all of the reference product lots available to them. A list should be provided in the application of all lots that were evaluated in any manner even if a particular lot was not used in the final similarity assessment. The list should include the disposition of each lot and the specific physicochemical, functional, animal, and clinical studies for which a lot was used. When a lot is specifically selected to be included in or excluded from certain studies, a justification should be provided. Similar information on every manufactured drug substance and drug product lot of the proposed biosimilar product should also be provided.

- U.S.-Licensed Reference Product and Other Comparators - The analytical similarity acceptance criteria should be derived using data from an analysis of the U.S.-licensed reference product, and the similarity assessment should be based on a direct comparison of the proposed biosimilar product to the U.S.-licensed reference product. As a scientific matter, combining data from the U.S.-licensed reference product and comparator products approved outside of the United States to determine the acceptance criteria or to perform the analytical similarity assessment generally would not be expected to support a determination that the proposed biosimilar is highly similar to the U.S.-licensed reference product. For example, combining data from U.S.-licensed reference product and non-U.S.-licensed comparator products may result in broader similarity acceptance criteria than would be obtained by relying solely on U.S.-licensed reference product lots due to increased variability of the products. Sponsors are encouraged to discuss with FDA, during drug development, any plans to use data derived from products approved outside of the United States.[4]

- Biosimilar Lots Manufactured with Different Processes - It may be possible to combine data in the analytical similarity assessment from proposed biosimilar product lots manufactured with different processes and/or at different scales. However, data should be provided in the 351(k) biologics license application to support comparability of any materials manufactured with the different processes and/or scales.

[4] See the guidance for industry *Biosimilars:Questions and Answers Regarding Implementation of the Biologics Price Competition and Innovation Act of 2009*.

IV. GENERAL PRINCIPLES FOR EVALUATING ANALYTICAL SIMILARITY

Analytical similarity should be assessed by using appropriate statistical methods to evaluate the analytical data. Methods of varying statistical rigor should be applied depending on the risk ranking of the quality attributes. Sponsors should develop an analytical similarity assessment plan that includes their proposed statistical approach to evaluation and then should discuss this approach with the Agency as early in the development program as feasible. The final analytical similarity report, which should include the analytical similarity assessment plan, should be included when a 351(k) biologics license application is submitted. The development of the analytical similarity assessment plan is the topic of the first subsection below, followed by a discussion of FDA's current thinking on the statistical methods to be applied for evaluation.

A. Analytical Similarity Assessment Plan

We recommend that the analytical similarity assessment plan be carefully designed to identify and address all factors that could impact the determination about whether the proposed biosimilar is highly similar to the reference product. Some factors that may need to be considered include:

- Differences in age of the lots produced at testing: It is recognized that differences in the age of the proposed biosimilar and reference product lots at the time of testing may result in analytical differences. There should, therefore, be a pre-specified plan to address how changes in attributes over the shelf-life will be incorporated into the determination of the similarity acceptance criteria.

- Multiple testing results: When there are multiple testing results for the same lot with a given quality attribute or assay, the biosimilar applicant should pre-specify which results will be selected for analytical similarity assessment.

- Assay performance: The assay methodologies and assay designs used in the analytical similarity assessment should be carefully considered and optimized, as needed. Poor assay performance, including high assay variability, should not be used to justify selection of either a particular evaluation tier or an inappropriately broad similarity acceptance criteria.

- Differences in attributes that will be considered acceptable: It may be known in advance that a difference less than or equal to a certain amount for a particular quality attribute would not be expected to have a clinical impact. In this situation, supporting information and an adequate justification for the allowable differences should be provided in the application.

We recommend that the analytical similarity assessment plan be developed in four stages, corresponding to the following activities:

- Development of the risk ranking of the reference product's quality attributes based on the potential impact on the clinical performance categories (i.e., the product's activity as well as pharmacokinetic/pharmacodynamic (PK/PD), safety, and immunogenicity profiles)

- Determination of the statistical methods to be used for evaluating each quality attribute based on the risk ranking and on other factors

- Development of the statistical analysis plan

- Finalization of the analytical similarity assessment plan

These four stages are described in more detail in the following subsections.

1. Development of Risk Ranking of Attributes

FDA recommends that biosimilar sponsors develop a risk assessment tool to evaluate and rank the reference product quality attributes in terms of potential clinical impact.[5] The risk assessment tool should be developed considering, at a minimum, the following two factors:

- Potential impact of an attribute on clinical performance: Specifically, we recommend that sponsors consider the impact of an attribute on activity as well as on pharmacokinetic/pharmacodynamic (PK/PD), safety, and immunogenicity profiles. For example, sponsors should consider available public information, as well as the sponsor's characterization of the reference product, in determining the potential impact of an attribute on clinical performance.

- The degree of uncertainty around a certain quality attribute: For example, when there is limited understanding of the clinical impact of an attribute, we recommend that that attribute be ranked as having higher risk because of the uncertainty involved.

FDA recommends that an attribute that is a high risk for any one of the performance categories (i.e., activity, PK/PD, safety, or immunogenicity) should be classified as high risk. Ideally, the risk assessment tool should result in a list of attributes ordered by the risk to the patient. The risk scores for attributes should, therefore, be proportional to patient risk. Because there may be a limited number of attributes that can be evaluated with equivalence testing (see section IV.A.2), attributes that are known to be of high risk to patients (i.e., high impact attributes) should be a priority over attributes with unknown but potentially high risk (i.e., attributes with a high-risk ranking due to uncertainty). The scoring criteria used in the risk assessment should be clearly defined and justified in the analytical similarity assessment plan, and the risk ranking for each attribute should be justified with appropriate citations to the literature and data provided.

[5] Certain quality evaluations of the reference product—e.g., its degradation rates, which are determined from stability or forced degradation studies—generally would not be included in the risk ranking. However, these evaluations will still factor into the assessment of the analytical similarity of the proposed biosimilar and reference product.

2. Determination of the Statistical Methods to be Used

FDA's current approach to evaluating analytical similarity is to define three tiers corresponding to the use of three different methods for comparing attributes. FDA believes that the use of these three tiers with appropriate similarity acceptance criteria should help support a demonstration that the proposed biosimilar is highly similar to the reference product. Equivalence testing (Tier 1) is typically recommended for quality attributes with the highest risk ranking and should generally include assay(s) that evaluate clinically relevant mechanism(s) of action of the product for each indication for which approval is sought. The use of quality ranges (Tier 2) is recommended for quality attributes with a lower risk ranking, and an approach that uses visual comparisons (Tier 3) is recommended for quality attributes with the lowest risk ranking. The three methods are described in Section IV.B.

In addition to risk ranking, however, other factors should be considered in determining which tier of statistical evaluation should be applied to a particular attribute or assay. Although many attributes may be considered high risk, subjecting all of these attributes to Tier 1 testing may result in a false negative conclusion (i.e., a determination that a product is not highly similar when it truly is). Some additional factors, besides risk, that should be considered when determining the appropriate tier include:

- Level of the attribute: An attribute of the reference product known to be of high risk but present at a level that is unlikely to have significant clinical impact could potentially be assessed at a lower tier. To justify placing a high risk attribute in a lower tier for this reason, the level of the attribute should be confirmed in both the reference product (as determined by the proposed biosimilar sponsor's analysis of the reference product) and the proposed biosimilar product. The selected limits regarding the level of an attribute should be defined and justified. The justification should also include consideration of how the level of the attribute changes over time.

- Assays used for assessing the attribute: Although multiple, orthogonal assays are encouraged for assessing a single attribute, not all assays need to be included in the same tier of assessment. The assay with the best performance characteristics for detecting product differences should be used for testing with the highest tier methods, while other assays should be used for testing with lower tier methods. A justification should be provided for the assays selected for testing at each tier.

- Types of attributes/assays: Some attributes or the assays used to assess the attribute will, by their nature, be excluded from certain statistical evaluations. For example, compendial assays, qualitative assays, or limit assays might be excluded from evaluation with Tier 1 and, in some cases, Tier 2 methods. The analytical similarity assessment plan should clearly define the conditions used to exclude assays from evaluation at any tier.

Applicable data and cited literature should be provided in the application to support the use of any additional factors in determining the appropriate tier of statistical assessment.

3. Development of the Statistical Analysis Plan

A detailed statistical analysis plan should be developed and included in the analytical similarity assessment plan because the statistical aspects of the evaluation will impact whether or not the similarity acceptance criteria are ultimately met. The plan for the statistical evaluation of analytical similarity requires the selection of design features from among many possibilities. These design features include the following five factors:

- the choice and risk ranking of attributes;
- the statistical approach (tier) for assessing each attribute;
- the number of proposed biosimilar and reference product lots to be evaluated for each attribute, and the number of replicates to be evaluated per lot;
- for each attribute, a determination of the largest acceptable difference between the proposed biosimilar and reference product that is considered to not have clinical impact;
- the methods of statistical analysis for each tier, and the type of assay(s) used to evaluate each attribute.

It is well known that bias may be introduced when there is an opportunity to select the most desirable result from a number of results obtained; consequently, the probability of a false positive result may be increased, and any estimated differences between the products are likely to be biased toward equivalence. Therefore, to minimize bias and the chance of erroneous conclusions, the statistical analysis plan should be pre-specified to the fullest extent possible. In some cases, it may be necessary to first collect preliminary data (e.g., to get an initial estimate of the variability of the reference product's attribute or to select an assay at the outset before finalizing the statistical analysis plan).

4. Finalization of the Analytical Similarity Assessment Plan

The final analytical similarity assessment plan should include the risk ranking of attributes, the specification of tiers of evaluation to be used for each attribute/assay, and the final statistical analysis plan. The plan should specify the anticipated availability of both proposed biosimilar and reference product lots for evaluation of each attribute/assay and should include a rationale as to why the proposed number of lots will be sufficient for evaluation purposes. The analytical similarity assessment plan should be discussed with the Agency as early in the biosimilar development program as possible so that agreement can be reached on which attributes/assays should be evaluated in each tier. The final analytical similarity assessment plan should be submitted to the Agency prior to initiating the final analytical assessments; typically this would be done in connection with a meeting with the Agency.

B. Statistical Methods for Evaluation

The Agency's current thinking on the statistical evaluation of analytical similarity is described in this section. Sponsors that intend to propose alternative statistical approaches to the Agency should do so during the analysis planning stage.

1. Tier 1 (Equivalence Test)

a. Hypotheses and statistical tests

Analytical similarity of the quality attributes determined to have the highest potential clinical impact (based on the risk ranking and other factors, as described in section IV.A) should be evaluated through formal statistical tests of equivalence. Equivalence of attributes measured on a continuous scale can be assessed by testing the difference in means between the proposed biosimilar and reference product. In the following formulas, μ_T and μ_R denote the population means, and σ_T^2 and σ_R^2 denote the population variances of the proposed biosimilar and reference product, respectively. To test for equivalence in means, the null and alternative hypotheses are given by

$$H_0 : \mu_T - \mu_R \le -\delta \text{ or } \mu_T - \mu_R \ge \delta$$
$$H_a : -\delta < \mu_T - \mu_R < \delta$$

In these formulas, δ is a positive number denoting the largest acceptable difference between the proposed biosimilar and reference product that is considered to not have clinical impact (i.e., the "equivalence margin"). Analytical similarity is supported if the null hypothesis of non-equivalence, H_0, is rejected. In other words, the statistical equivalence in means is established if the results of the statistical analysis indicate, with high confidence, that

$$-\delta < \mu_T - \mu_R < \delta$$

A test of the equivalence hypothesis can be conducted by requiring the simultaneous rejection of the following two one-sided null hypotheses:

$$H_{01} : \mu_T - \mu_R \le -\delta \text{ vs. } H_{a1} : \mu_T - \mu_R > -\delta$$
$$H_{02} : \mu_T - \mu_R \ge \delta \text{ vs. } H_{a2} : \mu_T - \mu_R < \delta$$

The probability of making a Type I error (i.e., declaring incorrectly that a biosimilar product's particular attribute is equivalent to a reference product's particular attribute) for a test of the equivalence hypothesis is controlled at the prespecified level α, provided each of the two one-sided hypotheses, H_{01} and H_{02}, is tested at the same level α.[6]

A convenient way to simultaneously test the two null hypotheses defining equivalence is through a confidence-interval-based test. If the (1-2α)100% two-sided confidence interval of the mean difference lies within ($-\delta$, δ), then both null hypotheses are rejected and the Type I error probability is controlled at level α for a conclusion of equivalence. For example, a 5% Type I error probability is obtained by requiring a 90% confidence interval to lie within ($-\delta$, δ).

[6] Schuirmann, DJ, 1987, A Comparison of the Two One-Sided Test Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability, J Pharmacokinet Biopharm, 15(6):657-680.

b. Margin determination

Determining an appropriate margin is a critical but challenging step for equivalence testing in any setting. Ideally, it would be possible to establish and pre-specify a biologically or clinically meaningful equivalence margin based on scientific knowledge or past experience. Often, however, such a margin is not readily available for every quality attribute deemed important enough for Tier 1 testing in a biosimilar development program. With this limitation, FDA currently recommends use of an equivalence margin that is a function of the reference product's variability for the attribute being tested. Specifically, the equivalence margin should be in the form of $f \times \sigma_{R,}$ where f is a fixed constant, and σ_R is the standard deviation of the quality attribute of the reference product. This suggested form of the equivalence margin is based on three criteria: (1) the goal of ensuring that values of the attribute being tested for the proposed biosimilar tend to fall within the reference product distribution, (2) the desire to have a unified representation of the margin for all Tier 1 quality attributes despite different levels of variability, and (3) the goal of having sufficient power for practical sample sizes.

After examining a range of possible values for the constant f, FDA determined that a reasonable value should be 1.5. With $\delta = 1.5\ \sigma_R$, the test generally should support equivalence if the 90% confidence interval of the difference in means lies within the interval (-1.5 σ_R, 1.5 σ_R) (i.e., the lower limit of the 90% confidence interval for the difference in means is greater than -1.5 σ_R and the upper limit is less than 1.5 σ_R). Use of this multiplier in computing the equivalence margin results in a test with reasonable properties under what we feel are realistic conditions. For example, if 10 biosimilar and 10 reference product lots are available, and the variability of the attribute for the reference product (σ_R) is known and not estimated from the sponsor's data, this test has adequate power (i.e., at least 85%) to reject the null hypotheses in favor of equivalence when the true underlying mean difference between the proposed biosimilar and the reference products is small, namely, equal to σ_R / 8, assuming a test of size $\alpha = 0.05$. If the true difference between products is less than σ_R / 8, power will be increased.

A limitation of the proposed approach to setting the equivalence margin is that σ_R is usually not known and must be estimated from the current reference product lots available to the sponsor. If one uses a t-test and does not consider the uncertainty in the estimate of the margin, the Type I error probability may be inflated. Alternative tests can be constructed to account for this additional uncertainty, but additional research is needed to better understand the operating characteristics of these tests (such as the small sample size performance of a Wald[7] test based on large-sample approximations).

2. Tier 2 (Quality Range Approach)

For Tier 2, the similarity acceptance criteria based on reference product results for a specific quality attribute should be defined as $(\hat{\mu}_R - X\hat{\sigma}_R, \hat{\mu}_R + X\hat{\sigma}_R)$, where $\hat{\mu}_R$ is the sample mean and $\hat{\sigma}_R$ is the sample standard deviation based on the reference product lots. The multiplier (X) should be scientifically justified for that attribute and discussed with the Agency. Based on our experience to

[7] Bickel, P.J. and Doksum, K., 2007, Mathematical Statistics: Basic Concepts and Selected Ideas, Vol. I.

date, methods such as the tolerance interval approach and the min-max approach are not recommended.[8]

Analytical similarity generally should be demonstrated for a quality attribute if a sufficient percentage of test lot values (e.g., 90%) fall within the quality range defined above for that attribute. The lots used for Tier 2 testing should, if possible, be the same as those used for Tier 1 testing.

3. Tier 3 (Visual Displays)

Attributes to be evaluated in Tier 3 should correspond either to those of lowest risk for potential clinical impact or those attributes which are important but not amenable to formal tests of hypotheses or quantitative evaluation. Various forms of visual displays may be used to compare the distribution of values from the proposed biosimilar and reference lots, and a subjective determination of the similarity should be made based on those displays. The lots used for the Tier 3 evaluation should be the same as, or a subset of, the lots used for Tier 1 and Tier 2 evaluations. The number of lots needed for the Tier 3 evaluation can depend upon a number of factors, including, for example, the expected lot-to-lot variability of the attribute. In cases where limited lot-to-lot variability is expected, a single lot of the proposed biosimilar and reference product for the Tier 3 evaluation may be acceptable.

4. Additional Considerations

We also recommend considering the following:

- The variance of an attribute (e.g., σ_R^2) encompasses both the within-lot and between-lot variance components. It is recommended that sponsors examine the contribution of the two variance components, as estimated from their lots, to help understand the performance of the assay. High assay variability generally is not an appropriate justification for a large value of δ. Instead, the assay should be optimized and/or the number of replicates per lot should be increased to reduce variability. We note that, in either case, lots of both the proposed biosimilar and the reference product should be assessed with the same number of replicates for that attribute, and the margin and all subsequent calculations should be defined using all lot values.

- For all quantitative quality attributes, including those subject to Tier 1 and 2 evaluations, descriptive statistics and visual displays should be used to present the reference and proposed biosimilar product distributions. In addition, the sponsor should submit sufficient data in its application to allow the Agency to conduct independent analyses.

- When the calculated equivalence margins or quality ranges are too wide or narrow, the Agency may adjust them to more appropriate levels.

[8] Dong, X, Tsong, Y and M Shen, 2015, Statistical Considerations in Setting Product Specifications, J Biopharm Stat, 25(2):280-294.

457 It is important to note that FDA's final assessment as to whether a proposed biosimilar is highly
458 similar to the reference product is made upon the totality of the evidence, rather than the passing or
459 failing of the analytical similarity criteria of any one tier or any one attribute. For example, the
460 Agency generally will consider the impact of an enhanced manufacturing control strategy when
461 making this final assessment.