Biosimilars
FORUM

800 17th Street, NW Suite 1100, Washington, DC 20006

November 21, 2017

Dr. Scott Gottlieb
Commissioner
Food and Drug Administration
10903 New Hampshire Avenue
Silver Spring, MD 20993

Re: Comments to Draft Guidance: "Statistical Approaches to Evaluate Analytical Similarity" (Docket No. FDA-2017-D-5525)

Dear Dr. Gottlieb:

The Biosimilars Forum appreciates the opportunity to comment on the Food and Drug Administration ("FDA") Draft Guidance regarding statistical approaches to evaluate analytical similarity, as published in Docket No. FDA-2017-D-5525.

The Biosimilars Forum ("Forum") is a non-profit organization whose mission is to advance biosimilars in the United States with the intent of expanding access and availability of biological medicines and improving health care. The Forum works on a consensus basis to develop policy positions to ensure the United States has a competitive, safe and sustainable biosimilars market, providing more options to patients and physicians.

General Comments to the Proposed Guidance

The Forum appreciates the Agency's efforts to publish this Draft Guidance document regarding the role of statistical analysis in the evaluation of analytical similarity in support of a biosimilar product sponsor's demonstration that a product is highly similar to a reference product licensed under Section 351(a) of the Public Health Services Act. While the Forum provides specific comments below, in general, the Forum urges FDA to build appropriate flexibility into the final guidance document, taking into account the complexity of the subject matter and its applicability to diverse biosimilar development programs.

Timely Consultation between Sponsors and FDA

The approaches outlined in this Draft Guidance are complex and warrant product-specific discussion with the Agency very early in development. The Forum believes that the Draft Guidance should encourage sponsors to contact FDA early in the product development process to discuss lot selection and risk ranking of attributes, and initiate discussions regarding the analytical similarity plan and statistical analysis plan. Due to the stepwise process of biosimilar development, FDA should be prepared to facilitate discussion very early in development in order to ensure there is no misinterpretation of expectations from the outset, and to continue with follow up discussions as needed throughout development as more information is gained. In particular, the Forum anticipates that sponsors will need to consult with the FDA after assays are validated and multiple analyses of the reference product are undertaken, in order to finalize the statistical analysis plan.

Defining "Attributes"

The International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) Quality Guidelines Q8 defines a critical quality attribute (CQA) as the following, "A CQA is a physical, chemical, biological, or microbiological property or characteristic that should be within an appropriate limit, range, or distribution to ensure the desired product quality." However, some sponsors do not include biological

functions in the Product Quality Attribute Assessment (PQAA), since they are functional outcomes of the physical/chemical properties (attributes) of the molecule. For such sponsors, control strategies based on the Product Risk Assessment are designed to control the physical/chemical attributes of a molecule to target a specified range for relevant biological functions as defined in the Quality Target Product Profile (QTPP). Other sponsors, by contrast, consider biological functions to be biological properties of the molecule and, consequently, to be critical product quality attributes that are included as a component of their PQAA.

In order to accommodate both approaches, we recommend keeping the concepts of physical/chemical attributes separate from that of biological functions in the similarity guidance. We recommend that the text therefore refer to "physicochemical properties and functional activities" rather than simply "quality attributes." Separating physicochemical properties and functional activities in this manner would mean that sponsors who do not include functional activities in their risk assessments would not have to perform completely separate risk assessments for the similarity exercise. Rather, they could utilize their PQAA assessment, and assess functional activities as an additional component required for the similarity analysis. The suggested edits should not impact sponsors who consider functional activities to be critical quality attributes and include them as a component of their current PQAA practices. Further, this nomenclature is consistent with the terminology employed in other FDA biosimilars guidance documents.[1]

Examples throughout the text include:

- Lines 207-209: Recommended change: "Development of ~~the risk~~ **a** ranking **structure for the assays/attributes that will be used to assess** ~~**of**~~ the reference product's ~~**quality attributes**~~ **physicochemical properties and functional activities** 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)"
- Lines 211-212: Recommended change: "Determination of the statistical methods to be used for evaluating each ~~**quality attribute**~~ **physicochemical property and functional activity** based on the risk ranking and on other factors"
- Line 255-257: Recommended change: "Equivalence testing (Tier 1) is typically recommended for ~~**quality attributes**~~ **physicochemical properties or functional activities** 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."

Challenges and Limitations to Applying Statistical Analyses in the Evaluation of Analytical Similarity Data

The Forum respectfully disagrees with the statement in the Draft Guidance (Lines 76-77) indicating that, "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." The Forum believes that statistical analysis cannot generally provide a higher degree of confidence in analytical similarity results; rather, it represents a supportive tool when data are amenable to statistics and analysis of the data will be meaningful to the understanding or interpretation of data. The application of inferential statistical methods should be considered in the context of supportive information, to assist or facilitate comparative evaluation of quality attributes. The

[1] E.g., FDA Guidance for Industry: Quality Considerations in Demonstrating Biosimilarity of a Therapeutic Protein Product to a Reference Product (April 2015); FDA Guidance for Industry: Scientific Considerations in Demonstrating Biosimilarity to a Reference Product (April 2015).

use of orthogonal, comprehensive characterization techniques should represent the primary comparative data source for quality attribute comparison.

The Forum appreciates the acknowledgements in the Draft Guidance that that there are many challenges and limitations to applying statistical analyses in the evaluation of analytical similarity data (Lines 78-86; 94-95; 263-265; 378-382; 403-404). However, we note that there are other challenges that are not directly acknowledged or addressed in the Draft Guidance. For example, the most substantial risk to the biosimilar developers' ability to design the statistical analyses is the absence of control over the reference product, whose quality attribute levels may change at any time during the proposed biosimilar product's development. The Forum believes the approach recommended by the FDA for statistical analysis of analytical similarity can be a useful tool in supporting interpretation of data when applied appropriately. However, the interpretation should be considered complementary to the overall analytical similarity assessment, and not used as a pass or fail decision tool.

Analytical Similarity vs. Analytical Similarity Assessment Plan

Throughout the Draft Guidance, the term "analytical similarity assessment plan" is used both to describe the application of analytical procedures for demonstration of biosimilarity and to encompass the statistical assessment similarity plan. For clarity, the Forum recommends that these two concepts be clearly distinguished. Further, use of the term "analytical similarity assessment plan" may mislead readers as to the scope of this Draft Guidance, and may contribute to lack of clarity, as the reader may infer this to relate to the totality of analytical data collected to assess whether the proposed product is highly similar to the reference product, rather than the evaluation of these data. The Forum respectfully requests that the FDA consider other terminology, such as "analytical similarity data evaluation plan" or "statistical assessment plan" to encompass the risk ranking, method determination, and statistical analysis plan.

Challenges of Developing a Pre-Specified Statistical Analysis Plan

The Draft Guidance states that the statistical analysis plan should be pre-specified to the fullest extent possible, and notes preliminary data may be collected to provide an initial estimate of variability of the reference product's attribute. (Lines 315 – 318.) Early in development, reference product changes over the duration of shelf-life will not be known to the biosimilar manufacturer. An accurate estimation of the variability of the reference product's attributes may not become apparent until many years after initial lots are sampled. Biosimilar manufacturers could plot assay results of the stability-indicating attributes over time in the reference product, but this would not be known until testing is complete.

Further, the Forum believes that the statistical analysis plan does not necessarily need to provide numbers. For example, a sponsor might state in the plan that the statistical interval that is to be developed from the reference product data will be of the form "mean +/- 3*standard deviation" without having to give numerical values for the mean or standard deviation. Instead, the sponsor would provide information on how many reference lots it would sample and over what period of time it would obtain these lots. Once the sponsor has the required number of lots and performs the testing, the sponsor would then provide the actual numerical limits, with an update to the Agency if needed.

Thus, developing a pre-specified statistical analysis plan based on initial reference product lots may be irrelevant and/or even incomplete, as the reference product is sampled and tested throughout the development program of the proposed biosimilar product. Changes observed which conflict with early estimates of variability should be accounted for, and the plan adjusted accordingly. The Forum agrees that the statistical analysis plan should be pre-specified to the fullest extent possible, however, as the reference product variability

is outside the control of the biosimilar developer, circumstances may arise which require the original estimate of reference product variability to be adjusted.

Ensuring Consistency with the Totality of the Evidence Approach

The Draft Guidance states, "It is important to note that FDA's final assessment as to whether a proposed biosimilar is highly similar to the reference product is made upon the totality of the evidence, rather than the passing or failing of the analytical similarity criteria of any one tier or any one attribute." (Lines 457-461.) While it is helpful that the Agency reiterates its totality of the evidence approach to assessment and acknowledges that the pass/fail of any one attribute is not considered the sole basis of approval decisions, the conclusion could go further to clearly state how these data are and are not expected to be used.

The utility of the proposed statistical approach to analytical data in the context of the totality of evidence remains unclear. The limitations of the outlined statistical approaches should be reiterated in the conclusion of this Draft Guidance and the Agency should elaborate on how it intends to use these data in the context of application review.

Finally, the last sentence of the Draft Guidance indicating that "the Agency generally will consider the impact of an enhanced manufacturing control strategy when making this final assessment," is unclear. The Forum requests clarification of this statement, perhaps in the form of examples.

Specific Comments to the Proposed Guidance

1. **Background and Scope (Lines 36-105)**

As noted above, the Forum disagrees with the characterization at Lines 76-77 that, "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." The Forum respectfully recommends the following language to more accurately reflect the role of statistical analysis: "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**~~ **can assist the comparative evaluation of quality attribute data generated using comprehensive orthogonal analytical procedures**."

The Draft Guidance at Lines 77-86 acknowledges certain challenges that exist in the application of statistical analysis. However, as noted in our general comments, one of the most substantial risks to the biosimilar developer is absent from this description. The Forum suggests including a fourth challenge here, which is the potential for the reference product ranges to change at any time during the development of the biosimilar product, with the potential to impact the viable application of statistical analysis as a supportive comparison tool.

The Forum believes that the sentence in Lines 78-79 stating that, "there may be a limited number of reference product lots, and those obtained may be the result of biased sampling," conflates two separate issues. Biosimilar manufacturers do not have complete control over the reference product lots they are able to obtain, whereas biased sampling may be of concern regardless of the number of product lots. Accordingly, the Forum suggests that the FDA consider revising the language as follows: "First, there may be a limited number of reference product lots, ~~**and**~~ **or** those obtained may be the result of biased sampling..."

The Draft Guidance states that to address the challenges in Lines 77-86, "the Agency recommends using a risk-based approach in the analytical similarity assessment of quality attributes." (Lines 88-89.) While a risk-

based approach facilitates categorization of the importance of attributes based on potential clinical impact, it may not fully address the appropriate course of action when faced with challenges posed by a limited number of reference product and/or biosimilar lots. Therefore, the Forum respectfully requests that the Agency provide further clarity regarding how a risk-based approach in the analytical similarity assessment of quality attributes fully addresses the stated challenges. Further, the Forum recommends the Agency include acknowledgement that these challenges may need to be further discussed with the Agency to achieve full resolution.

Minor edits suggested for accuracy/consistency:

- At Line 85, the Forum believes that the FDA likely intended to state that subjecting all potential quality attributes to formal statistical tests based on limited lots could lead to incorrect conclusions regarding a lack of similarity for a high number of "attributes," rather than "products," and suggests that substituting "attributes" would be more appropriate.
- The Forum proposes a clarification in Lines 93-94, as evaluation of attributes in Tier 3 is not necessarily through statistics. The Forum recommends the following deletion: "…these attributes/assays are evaluated according to one of three tiers **~~of statistical approaches~~ of potential risk** based on a consideration of risk ranking as well as other factors."

2. Reference and Biosimilar Products (Lines 107-161)

The Forum understands that prior to the release of this Draft Guidance, the Agency has been requiring independent lots of biosimilars, derived from different lots of drug substances, to be used for statistical analysis. While lot selection is discussed in Section III of the Draft Guidance, the concept of lot independence is not mentioned in Section III, nor elsewhere in the document. The Forum would appreciate the Agency clarifying whether the requirement for lot independence for biosimilars is no longer aligned with its current thinking.

Reference Product Lots

In FDA's description of recommended variability of reference product lots, FDA is proposing to require the sponsor to select reference product lots that are in different stages of shelf-life. (Lines 127-128.) However, because the availability of reference product and the assigned expiry is outside the control of the biosimilar developer, and therefore requiring the reference product lots to span the shelf life may not always be achievable, the Forum requests that the Agency temper the language included in this bullet point to state, "Lots with remaining expiry **(at time of purchase), and lots ideally spanning** the reference product shelf life should be selected."

At Lines 132-133, FDA indicates that, "Sponsors should account for all of the reference product lots available to them." The Forum proposes a clarification because, as written, the text gives the impression that all reference product lots available on the market during the development window are to be listed. The Forum suggests replacing this sentence with, "Sponsors should account for all of the reference product lots **acquired during product development**."

The Agency goes on to state that the sponsor should create a list of all lots that were evaluated in any manner, even if a particular lot was not used in the final similarity assessment, including "the disposition of each lot and the specific physicochemical, functional, animal, and clinical studies for which a lot was used." (Lines 132-137.) The Forum believes that this could add substantial burden to the application process, and would thus ask that the Agency be more flexible, and allow for these data to be retained by the manufacturer and provided upon inspection.

On the topic of requiring the use of U.S. reference product for analytical similarity assessment, the Forum believes the Agency should allow the use of non-U.S. licensed reference products if comprehensive, sensitive and specific bridging data can be provided to justify the use of such batches, or if justified based on preliminary assessments showing no statistically significant difference in variability among U.S. versus non-U.S. lots. However, the studies to demonstrate similarity should be appropriately designed. In the event that ex-U.S. products are considered for similarity studies, the sponsor should engage in discussion with the Agency early in development.

The Draft Guidance encourages sponsors planning to use data derived from products approved outside of the U.S. to discuss this with the FDA during product development. (Lines 153-155.) However, there is not specific guidance regarding what data and information the FDA will require from the sponsor to support the use of reference product material obtained outside of the U.S. The Draft Guidance should be modified to more clearly outline what the Agency's expectations are for scientific justification of use of a non-U.S.-licensed comparator.

Biosimilar Product Lots

As noted above, our understanding is that the Agency has to date strongly recommended the use of independent drug product lots (drug product lots manufactured from different drug substance lots). As the concept of independent lots is not addressed in this guidance, the Forum requests clarification as to whether FDA has changed its thinking.

In Lines 122-124, the Agency recommends a minimum of 10 biosimilar lots for inclusion in the analytical similarity assessment. The Forum appreciates FDA's recognition that, in some cases, a biosimilars sponsor may not be able to obtain the recommended number of reference product lots, and that the opportunity to discuss alternative analytical similarity assessments with the Agency is available in such cases. (Lines 118-120.) However, we note that there also may be circumstances where a biosimilar sponsor may not have the recommended number of biosimilar product lots. The Forum therefore requests that the Agency add language to the bullet point at Lines 122-124 providing the same opportunity to discuss alternatives with FDA, i.e., "In cases where limited numbers of lots are available, alternate analytical similarity assessments should be proposed and discussed with the Agency."

3. General Principles for Evaluating Analytical Similarity (Lines 168-176)

The Draft Guidance seems to contain conflicting descriptions regarding development of the analytical similarity plan: Lines 110-111 state that the analytical similarity plan is developed "based on information obtained about these [structural/physicochemical and functional] attributes during development of the proposed biosimilar…" This suggests that the analytical similarity plan is developed *a posteriori*. However, Lines 315-316 state "…to minimize bias and the chance of erroneous conclusions, the statistical analysis plan should be pre-specified to the fullest extent possible." The Forum agrees that the statistical analysis plan should be pre-specified as much as possible, however, as the reference product variability is outside the control of the biosimilar developer, circumstances may arise which require the original estimate of reference product variability to be adjusted based on data. The Forum suggests that FDA include language that better addresses factors that may influence the timing of development and adjustment of the analytical similarity plan and guidance on how biosimilars developers should balance these factors.

The Forum recommends a clarification regarding the scope of the analytical similarity information that should be included with a 351(k) biologics license application. The Forum suggests that the Agency clarify that sponsors are not required to include all internal reports, which may contain references to documents which are not part of the submission. Accordingly, the Forum recommends the following change to Lines 172-173: "The

final analytical similarity ~~**report, which should include the analytical similarity**~~ assessment plan and results should be included when a 351(k) biologics license is submitted."

a. Analytical Similarity Assessment (Lines 180-218)

The Draft Guidance states at Lines 184-188, "There should…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." Pre-specifying the plan, which incorporates the effect of product age on the quality attributes, could pose difficulties for biosimilar manufacturers, since the correlation of age of the product and the quality attributes of the reference product is not clear at the initial stages of biosimilar development process. Early in development, reference product changes over the duration of shelf-life will not be known to the biosimilar manufacturer. Biosimilar manufacturers could plot assay results of those attributes which may be stability-indicating over time to gain a general idea of whether or not a particular attribute is changing, but even this would not be known until testing is complete.

Further, estimations of the variability of the reference product's quality attributes could also be impacted by reference product lot-to-lot variability, or even deliberate manufacturing changes, over which biosimilar developers have no control. These changes can only be gradually understood via the procurement and testing of multiple lots over a long period of time, often well into, or even toward the end of the course of the biosimilar development program.

The Forum also requests that the Agency either: (1) provide clear guidance as to how biosimilar sponsors should account for observed changes that conflict with early estimates of variability, and how the statistical analysis plan can be adjusted accordingly to account for additional data, or (2) reconsider the need for pre-specified criteria given the challenges for biosimilar manufacturers to establish meaningful pre-specified criteria, and allow sponsors to perform similarity assessment first, and then continue to monitor the effect of product age on the quality attributes throughout the biosimilar development process.

The Forum requests clarification of Lines 190-192 regarding multiple testing results and asks FDA to clarify whether pre-specifying which results will be selected refers to defining the reportable value.

With regard to differences in attributes that will be acceptable, the Draft Guidance states at Lines 199-202, "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." However, it is unclear how this statement relates to the determination of similarity acceptance criteria. The Forum asks for clarification on this point. For example, would knowledge that a difference in a particular quality attribute is not expected to have a clinical impact justify placement into Tier 2 rather than Tier 1, or be supportive for a sponsor to suggest a margin other than the $1.5\sigma_R$ margin?

i. Development of Risk Ranking of Attributes (Lines 222-246)

The Draft Guidance suggests in two places (Lines 111, 229-230) that the risk ranking should consider information obtained during the development of the proposed biosimilar, as well as the biosimilar sponsor's characterization of the reference product. While the Forum agrees that the justification for the risk ranking should be provided, the FDA's use of the term "scoring criteria" in Lines 237-246 could be interpreted as establishment of numerical values which may not be necessary to fully convey the justification of the risk ranking. Accordingly, the Forum recommends the following revisions to the language at Lines 244-246: "~~**The scoring criteria used in the risk assessment should be clearly defined and justified in the analytical**~~

~~similarity assessment plan, and the~~ **A justification of the proposed** risk ranking for each physicochemical property and functional activity should be ~~justified~~ **provided** with appropriate citations to the literature and data provided."

The Draft Guidance acknowledges that "there may be a limited number of attributes that can be evaluated with equivalence testing." Scientific criteria for the selection of Tier 1 attributes (e.g., direct relation to mode of action) could be more useful than the proposed tool with its described shortcomings. Risk ranking for assigning Tier 2 and Tier 3 attributes may be very well applicable.

ii. Determination of the Statistical Methods to be Used (Lines 252-292)

The Forum requests the addition of language that addresses quality attributes that are not amenable to statistical analysis by recommending comparison under Tier 3. Accordingly, Lines 259-260 of the Draft Guidance could be revised to read, "…an approach that uses visual comparisons (Tier 3) is recommended for quality attributes with the lowest risk ranking **or those not amenable to statistical analysis**."

At Lines 269-276, the Draft Guidance states: "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 …and the proposed biosimilar product." For some quality attributes, such as process impurities and some low-level product-related impurities, the Forum believes the focus should be on the attribute level in the biosimilar rather than the reference product. Accordingly, the Forum requests that this section be adjusted to acknowledge and account for the scenario where the level of the quality attribute in the biosimilar should be the deciding criterion for tier placement rather than the level of the attribute in the reference product.

The Forum also suggests a revision to the description of types of attributes/assays at Lines 286-288 to delete the reference to "compendial assays" as potential exclusions. Excluding all compendial assays is considered too broad as some assays may be used to measure critical attributes such as protein concentration.

Minor edits suggested for accuracy/consistency:

- The Forum recommends a minor clarification in Lines 252-253 to indicate that "FDA's current **recommended** approach to evaluating analytical similarity is to define three tiers corresponding to the use of three different methods for comparing attributes."
- The Forum recommends a revision in Line 276 for consistency: "The justification should also include consideration of how the level of the attribute changes over ~~time~~ **shelf life.**"

iii. Development of the Statistical Analysis Plan (Lines 298-318)

The Draft Guidance notes the statistical analysis plan should include selection of design features including certain factors; among these factors the Draft Guidance includes, "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…" (Lines 307-308.) The Forum requests that FDA delete this language. As articulated in the Draft Guidance, a pre-specified clinically meaningful equivalence margin is often not readily available for every Tier 1 quality attribute. (Lines 378-382.) For the same reason, it is difficult to pre-define the largest acceptable difference for **each** quality attribute. Further, it is unclear why an *a priori* determination of the largest acceptable difference between the proposed biosimilar and reference product that is considered to not have clinical impact is required for each attribute beyond the determination of the tiering classification and

associated justification of the statistical plan. An evaluation of the potential clinical impact of differences is needed only for attributes where a difference is found between the biosimilar and the reference product as an outcome of the analytical assessment. Consideration of acceptable differences in attributes is already discussed in Lines 199-202, which describe a more realistic approach to providing supporting information and an adequate justification in situations where allowable differences can be known in advance.

b. Statistical Methods for Evaluation (Lines 335-337)

The Forum generally appreciates the utility of dividing quality attributes into three tiers based on clinical criticality, but emphasizes that analytical data should be assessed with statistical methods when the analysis of the data will be meaningful to the understanding or interpretation of data.

However, the Forum disagrees with language in the Draft Guidance indicating that the lots used for testing should, if possible, be the same for all tiers. (Lines 420-421; 429-430.) The Forum notes that there are limitations to this suggestion, and recommends deletion of this language. Specifically:

- Not all analytical methods used in the final similarity assessment are necessarily available at the beginning of the biosimilar development. Analytical methods may be revised, replaced, or added.
- Some of the more difficult characterization methods may not be ready as early as other methods making it difficult to test the exact same lots, particularly when the reference product expiry is short.
- The lot selection for testing by individual method is often dictated by availability of the material (amount, expiry) and the final method.

i. Tier 1 (Equivalence Test) (Lines 340-408)

The Forum expresses some degree of concern with the proposed statistical recommendations for Tier 1. The Forum believes that the requested equivalence testing for Tier 1 attributes may pose the risk of restricting biosimilar approvals for random reasons only, if applied strictly as pass/fail criteria, which may have a substantial detrimental impact on the development of biosimilars. Strict application of the requested equivalence testing also would set separate regulatory standards which are in conflict with existing regulatory expectations as described in FDA-adopted guidelines such as ICH Q6B, ICH Q5E, ICH Q7, ICH Q8, ICH Q11.

Equivalence testing evaluates whether the mean of two data sets of quality attributes are equivalent. In cases where the mean of the reference product changes over time, linking the biosimilar approval strictly to the outcome of equivalence testing (i.e., requiring the biosimilar candidate to have an equivalent mean), would result in a random outcome of biosimilar assessment – depending on where the mean of the reference product is shifting over time.[2] Accordingly, the Forum requests that FDA include language in the guidance that encourages sponsors to discuss alternative methodologies with the Agency in such cases.

[2] For real life examples, *see* Schiestl *et al.,* Acceptable changes in quality attributes of glycosylated biopharmaceuticals, Nature Biotechnology (2011) 29:310-312; Kim *et al.,* Drifts in ADCC-related quality attributes of Herceptin®: Impact on development of a trastuzumab biosimilar, mAbs (2017), *available at* http://dx.doi.org/10.1080/19420862.2017.1305530; Lamanna *et al*., The structure-function relationship of disulfide bonds in etanercept, Scientific Reports (2017), http://www.nature.com/articles/s41598-017-04320-5.pdf.

Biosimilars Forum
Comments to FDA Draft Guidance:
Statistical Approaches to Evaluate Analytical Similarity

a. Margin Determination (Lines 376-408)

Section (b) under Tier 1 (Lines 376-408) provides FDA's rationale for the proposed approach to determining an appropriate margin, which the Draft Guidance states "is a critical but challenging step for equivalence testing in any setting." For this reason, we would suggest that the Agency engage stakeholders on appropriate margin determinations, perhaps by conducting a workshop on this topic to come up with more generally applicable criteria based on scientific evidence.

The Forum notes that the Agency acknowledges a limitation of the proposed approach in Lines 403-408, stating "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 test based on large-sample approximations)." The Forum posits that in order to account for the "uncertainty in the estimate of the margin," a measure such as effect size could be used to address this issue. A confidence interval on the effect size can be calculated and compared to a pre-determined equivalence margin.

We suggest the use of effect size

$$\frac{\mu_T - \mu_R}{\sigma_R}$$

to avoid the issue of 'uncertainty in the estimate of the margin.' The equivalence margin, δ, would then be a fixed constant (e.g., 1.5). We note that if effect size is to be used, the equation in Line 360 would become:

$$-1.5 < \frac{\mu_T - \mu_R}{\sigma_R} < 1.5$$

We note that while the control of Type I error is one of the major concerns of the current Tier 1 approach, it is not the only concern. For instance, the purpose of the equivalence test here is to test the mean difference between the reference product and the biosimilar product. The Forum notes that there may be instances where strict application of the Tier 1 equivalence testing is not suitable, and in such cases, the sponsor should consult with the Agency.

Another limitation worth pointing out is the equivalence margin, which is not fully science-based. As shown in the Draft Guidance, the margin 1.5 $\hat{\sigma}_R$ is based on the assumption that the true mean difference is no more than σ_R /8, which may or may not be a clinically relevant difference and therefore appears to be arbitrary.

The Forum believes that confusion may be caused by the use of σ_R in Lines 376-408, while Lines 413-414 use $\hat{\sigma}_R$ when discussing Tier 2. The Forum suggests that Lines 403-404 should be updated to explicitly state, "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**, therefore $\hat{\sigma}_R$ is used in practice."**

The Forum encourages FDA to consider the limitations of the proposed approach. Per Lines 199-202, it should be considered appropriate for a sponsor to scientifically justify an alternative margin if there is knowledge that a difference in a particular quality attribute is not expected to have a clinical impact.

ii. Tier 2 (Quality Range Approach) (Lines 410-421)

The Draft Guidance defines similarity acceptance criteria for Tier 2 based on reference product results for all quality attributes in the same manner. (Lines 412-414.) FDA is urged to consider allowing a one-sided range for attributes having a desired upper or lower limit only (i.e., aggregates and monomer).

Lines 414-415 of the Draft Guidance require that, "The multiplier (X) should be scientifically justified for that attribute and discussed with the Agency." Without framing what the Tier 2 Quality Range Approach is exactly evaluating, it is difficult to come up with a scientific justification of the multiplier. It is not clear what scientific factors need to be considered for the justification of X. For this reason, X=3 should apply as the "general rule," but the Forum requests that the Agency be open to discussions with the sponsor should different multipliers be scientifically justified.

Lines 415-417 of the Draft Guidance state that: "based on our experience to date, methods such as the tolerance interval approach and the min-max approach are not recommended." The Forum appreciates that the Agency believes that tolerance intervals and min-max are not suitable statistical approaches for Tier 2 quality attributes, but as written, the text may be misunderstood to mean that the tolerance interval and min-max approaches are not appropriate for any aspect of biosimilar development. The Forum believes that the tolerance interval approach can be justified when a relatively large number of reference product lots are tested. The min-max range represents the experimentally verified range of the reference product and has therefore its value as a descriptive tool (e.g., in defining the development target for the biosimilar development, which could include setting appropriate ranges in the QTPP), or as a supportive description in the final biosimilarity assessment. The Forum therefore recommends modifying this sentence as follows: "Based on our experience to date, methods such as the tolerance interval approach and the min max approach are**, in general,** not recommended **for use in a Tier 2 evaluation unless their use can be scientifically justified**."

iii. Tier 3 (Visual Displays) (Lines 425-433)

With regard to the attributes to be evaluated in Tier 3, the Agency states that it 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." (Lines 425-427.) However, as mentioned in Line 269-276, an attribute of the reference product known to be of high risk but present at a lower level could potentially be assessed at a lower tier (e.g., some attributes that have the potential for clinical impact but are present at low levels and do not represent a safety concern at the levels present could also be ranked as non-critical). Therefore, this should be included in the statement regarding the attributes to be evaluated in Tier 3. Accordingly, the statement at Lines 425-427 should be included in the description of attributes to be evaluated in Tier 3.

iv. Additional Considerations (Lines 435-461)

The Forum believes that Lines 442-444 confuse two method characteristics and should be revised. Assays should be both accurate and precise and these assay characteristics should not be mixed up. Assays can be optimized to increase accuracy, or the number of replicates may be increased to improve precision. Accordingly, the Forum recommends the following revision, "High assay variability generally is not an appropriate justification for a large value of δ. Instead, **the source of the variability should be investigated,**

the assay should be optimized and/or the number of replicates per lot should be increased to reduce variability as appropriate."

Lines 454-455 state, "When the calculated equivalence margins or quality ranges are too wide or narrow, the Agency may adjust them to more appropriate levels." The Forum requests clarification on this statement from the Agency, and suggests that it would be more appropriate for the Agency to discuss these situations with the sponsor and reach agreement on more appropriate levels.

Further, the statement made in Lines 454-455 seems to contradict the purpose of this guidance in providing advice on the statistical approaches recommended for evaluating analytical similarity. Since the Agency has provided specific statistical methods and margins for attributes in each tiering system in this guidance and stated that the similarity assessment plan should be discussed with the Agency, this statement seems unnecessary and could be misleading. In order for this statement to remain in the guideline, the Agency needs to provide clear description of the cases with statistical basis, in which the equivalence margins or quality ranges are deemed too wide or narrow.

Conclusion

The Forum appreciates the Agency's work to publish this Draft Guidance and the opportunity to provide comments. The Forum respectfully asks the FDA to carefully consider its comments and concerns as the Agency formulates its final guidance document.

If you have any questions or need any additional information, please contact Michael Werner at 202.419.2515 or at michael.werner@hklaw.com.