DESIGN AND ANALYSIS OF BIOSIMILAR STUDIES

3-part Course instruted by

Shein-Chung Chow, PhD

at the 1st
Annual Biosimilars Forum
Budapest, Hungary | October 6-7, 2016

1

Lecture 1:

Assessing Biosimilarity: Issues and Recent Development

2

Lecture 2:

Assessing Interchangeability: Issues, Designs and Statistical Methods

3

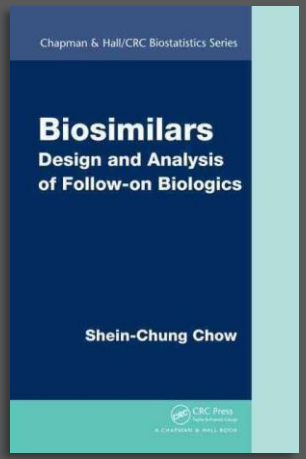
Lecture 3:

Analytical Similarity Assessment in Biosimilar Studies

DESIGN AND ANALYSIS OF BIOSIMILAR STUDIES

Annual Biosimilars Forum

This material includes the 3rd Lecture (entitled: **Analytical Similarity Assessment in Biosimilar Studies**) of the scientific course presented by ***Professor Shein-Chung Chow*** at the **1st Biosimilars Forum** in Budapest. The first and second parts of the course will be available to download separately, courtesy of Annual Biosimilars Forum event series at the Forum's official website: www.biosimsforum.com.



The materials of the course are developed based on

- the book entitled "**Biosimilars: Design and Analysis of Follow-on Biologics**" by *Chow SC* published in 2013 by Chapman and Hall/CRCPress, Taylor & Francis, New York,
- and the 3rd edition of the book entitled "**Design and Analysis of Bioavailability and Bioequivalence Studies**" by *Chow SC* and *Liu JP* published in 2008 by Chapman and Hall/CRC Press, Taylor & Francis, New York.

THE FIRST ANNUAL BIOSIMILARS FORUM
Budapest, Hungary, October 6-7, 2016

Design and Analysis of Biosimilar Studies

Lecture 1: Assessing Biosimilarity: Issues and Recent Development

Lecture 2: Assessing Interchangeability: Issues, Designs and Statistical Methods

Lecture 3: Analytical Similarity Assessment in Biosimilar Studies

DukeMedicine

Instructor

Shein-Chung Chow, PhD

Professor, Biostatistics and Bioinformatics, Duke University School of Medicine
Special Government Employee, United States Food and Drug Administration
Fellow, American Statistical Association, USA
Editor-in-Chief, Journal Biopharmaceutical Statistics and Drug Designing
Editor-in-Chief, Biostatistics Book Series, Chapman and Hall/CRC Press
More than 30 years of experience in pharmaceutical/clinical development
Author or co-author of over 280 methodology papers and 24 books
including *Biosimilars: Design and Analysis of Follow-on Biologics*

Design and Analysis of Biosimilar Studies

Lecture 3
Analytical Similarity Assessment in Biosimilar Studies

Shein-Chung Chow, PhD
Department of Biostatistics and Bioinformatics
Duke University School of Medicine
Durham, North Carolina, USA

October 6, 2016

DukeMedicine

Acknowledgements

- Some of the slides are borrowed from
  Dr.Yi Tsong (FDA Division Director)
  Dr. Xiaoyu Dong (FDA reviewer)
  presentations at 2015 Duke-Industry Statistics Symposium held in Durham, North Carolina, USA and 2016 DIA-FDA Statistics Forum held in North Bethesda, Maryland, USA

DukeMedicine

Outline

- Background
  - BPCI's definition of biosimilarity
  - Stepwise approach
  - Recent regulatory submission
- Analytical similarity assessment
  - Classification of critical quality attributes
  - Three-tier approach
    - Equivalence test for Tier 1 CQAs
    - Quality range approach for Tier 2 CQAs
    - Raw data and graphical comparison for Tier 3 CQAs
- FDA's current thinking on Tier approach

DukeMedicine

Recall
BPCI's definition of biosimilarity

A biosimilar product:

Is highly similar to the reference product notwithstanding minor differences in clinically inactive components

There are no clinically meaningful differences in terms of safety, purity and potency.

US BPCI Act, 2009

DukeMedicine

Recall scientific factors

- Highly similar and minor difference
  - Degree of similarity
- Clinical meaningful difference?
  - Equivalence limit = non-inferiority margin
  - How to determination non-inferiority margin?
  - Non-inferiority ≠ equivalence
- How many biosimilar studies are required?
  - Safety (PK/PD, safety/tolerability, immunogenicity)
  - *Purity (analytical similarity in critical quality attributes)*
  - Potency (efficacy)
- How to assess biosimilarity?

DukeMedicine

FDA's guidance

- FDA published three draft guisances on biosimilars in 2012 (finalized in early 2015)
  - Scientific considerations in demonstrating biosimilarity to a reference product
  - Quality considerations in demonstrating biosimilarity to a reference protein product
  - Biosimilars: questions and answers regarding implementation of the BPCI Act of 2009
- FDA recommends using stepwise approach in order to provide totality-of-the-evidence for demonstrating biosimilarity
- FDA's draft guidance on analytical similarity assessment is to be circulated for comments any time soon.

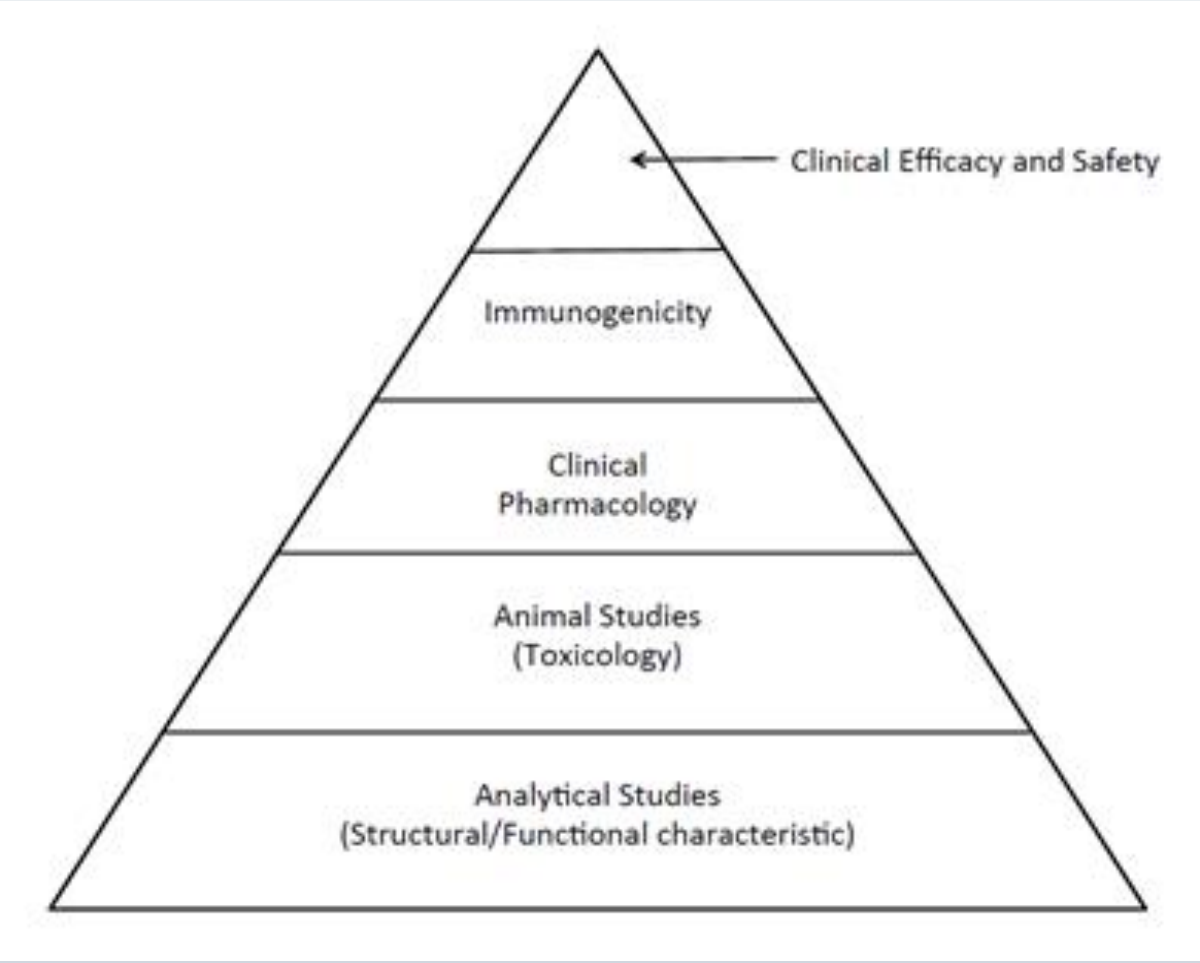
DukeMedicine

Recall
stepwise approach

- Analytical studies
  - Critical quality attributes at various stages of manufacturing process
- Animal studies
  - The assessment of toxicity
- Clinical pharmacology
  - Pharmacokinetics (PK) or pharmacodynamics (PD)
- Clinical studies
  - The assessment of immunogenicity
  - Safety/tolerability
  - Efficacy

DukeMedicine

Stepwise approach for obtaining totality-of-the-evidence



DukeMedicine

Recent regulatory submissions at FDA

Current status of biosimilar reviewers

- In total, FDA/CDER has **9 biosimilar BLA** submissions and about a hundred PIND/IND submissions.
- **2 Approved** biosimialr products: Zarxio (biosimilar to Neupogen) and Inflectra (biosimilar to Remicade)
- **A few pending biosimilar** BLAs are under review: the reference products are Humira, Enbrel, Neulasta)

First regulatory submission at US FDA

- Sandoz biosimilar filgrastim recommended for approval by FDA Oncologic Drugs Advisory Committee (ODAC) on January 7, 2015, which was subsequently approved by the FDA on March 6, 2015
- Biosimilar filgrastim recommended to be approved for use in all requested indications in the reference product's (Amgen's Neupogen®) label.
- Committee's recommendation based on review of extensive data from analytical, non-clinical, clinical studies and post-marketing pharmacovigilance.

DukeMedicine

FDA recommended 3-tier approach

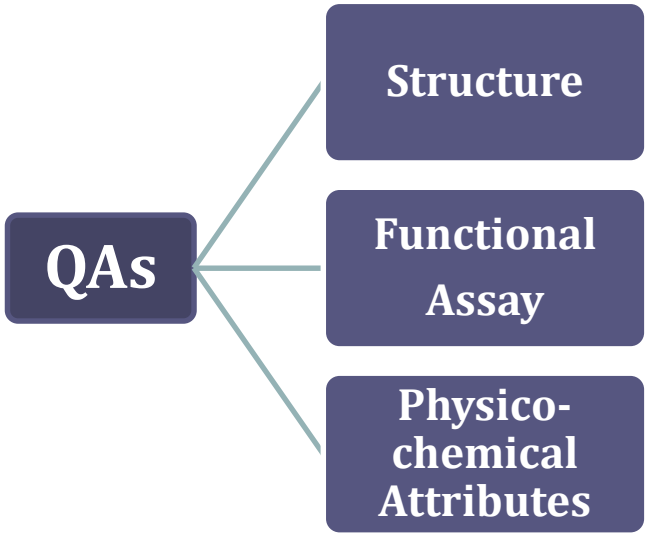
- Classification of critical quality attributes (CQAs) into three tiers according to their criticality or risking ranking relevant to clinical outcomes
- An appropriate statistical model or scoring system based on
  - mechanism of action (MOA) or
  - pharmacokinetic/pharmacodynamics (PK/PD)
  - Information available in the literature

  should be used whenever possible

DukeMedicine

FDA recommended 3-tier approach

- Analytical similarity study:
  - Characterize the proposed biosimilar and its reference product;
  - Tests for a number of quality attributes (QA);



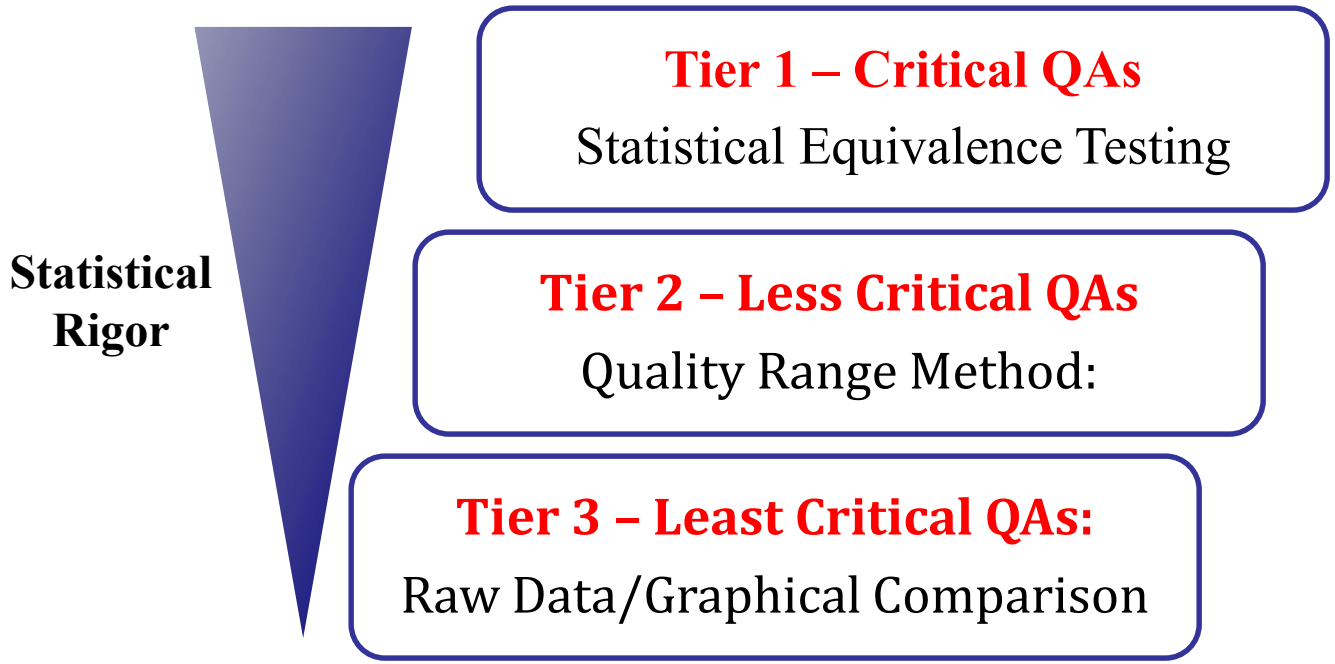
Quality Attribute	Methods
Primary structure	• N-terminal sequencing • Peptide mapping with UV and MS detection • Protein molecular mass by ESI MS • Protein molecular mass MALDI-TOF MS • DNA sequencing of construct cassette • Peptide mapping coupled with MS/MS
Bioactivity	• Proliferation of murine myelogenous leukemia cells (NFS-60)
Receptor binding	• Surface Plasmon Resonance
Protein content	• RP-HPLC
Clarity	• Nephelometry
Sub-visible particles	• Micro flow imaging
Higher Order Structure	• Far and Near UV circular dichroism • ^{1}H nuclear magnetic resonance • ^{1}H-^{15}N heteronuclear single quantum coherence spectroscopy • LC-MS (disulfide bond)
High molecular weight variants/aggregates	• Size exclusion chromatography • Reduced and non-reduced SDS-PAGE
Oxidized species	• RP-HPLC • LC/MS
Covalent dimers	• LC/MS
Partially reduced species	• LC/MS
fMet1 species	• RP-HPLC • LC/MS

Quality Attribute	Methods
Sequence variants: His→Gln Asp→Glu Thr→Asp	• RP-HPLC • LC/MS
Succinimide species	• RP-HPLC • LC/MS
Phosphoglucunoylation	• LC/MS
Acetylated species	• LC/MS
N-terminal truncated variants	• LC-MS/MS
Norleucine species	• RP-HPLC • LC/MS
Deamidated species	• RP-HPLC • LC/MS • IEF • CEX

Reference: BLA 125553

FDA recommended 3-tier approach

- The Tiered Approach (OB & OBP):
  - QAs are assigned to different tiers based on its criticality;
  - Different statistical/quantitative approaches are applied to each tier;



Summary – analytical similarity assessment

- Identify critical quality attributes (CQAs) which are relevant to clinical outcomes
  - Focus on structural/functional assays and physicochemical attributes
- Based on MOA and/or PK/PD, classify the identified CQAs into the following tiers depending upon their criticality (or risk ranking)
  - Tier 1: most relevant
  - Tier 2: mild to moderate (less) relevant
  - Tier 3: least relevant

DukeMedicine

Summary of analytical similarity assessment

- Tier 1 CQAs
  - Most relevant to clinical outcomes
  - Equivalence test
- Tier 2 CQAs
  - Mild-to-moderate relevant to clinical outcomes
  - Quality range approach
- Tier 3 CQAs
  - Least relevant to clinical outcomes
  - Raw data and graphical comparison

DukeMedicine

Equivalence test for Tier 1 CQAs

Interval hypotheses:

$$H_0: \mu_T - \mu_R \leq -\delta \quad \text{or} \quad \mu_T - \mu_R \geq \delta,$$

where $\delta > 0$ is the equivalence limit (or similarity margin), and μ_T and μ_R are the mean responses of the test (proposed biosimilar) product and the reference product lots, respectively.

DukeMedicine

Equivalence test for Tier 1 CQAs

- Analytical equivalence (similarity) is concluded if the null hypothesis of *in*equivalence (*dis*similarity) is rejected.
- Similar to the confidence interval approach for bioequivalence testing under the raw data model, analytical similarity would be accepted for the quality attribute if the (1-2α)100% two-sided confidence interval of the mean difference is within $(-\delta, \delta)$.

DukeMedicine

Equivalence test for Tier 1 CQAs

- Under the null hypothesis, FDA indicates that the equivalence limit (similarity margin), δ, would be a function of the variability of the reference product (σ_R).
- σ_R can be estimated based on some sampled lots randomly selected from a pool of reference lots for the statistical equivalence test.
- FDA recommends $\delta = c * \sigma_R$

DukeMedicine

EAC (equivalence acceptance criterion)

- FDA recommends $\mathrm{EAC} = \pm\, c * \sigma_R$
- FDA further recommends $c = 1.5$, which is considered a regulatory constant for equivalence test for Tier 1 CQAs
- Thus, $\mathrm{EAC} = \pm 1.5 * \sigma_R$, where σ_R can be estimated from available reference lots
- *Sample size*
  - How many lots should be used (i) for estimation of σ_R and (ii) for achieving a desired power for establishment of analytical similarity?

DukeMedicine

Justification of EAC $= 1.5 * \sigma_R$

- Recall criterion for IBE((individual bioequivalence equivalence)

$$\theta = \frac{(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WR}^2 - \sigma_{WT}^2)}{\max(\sigma_0^2, \sigma_{WR}^2)}$$

DukeMedicine

Justification of $\mathbf{EAC} = \pm \mathbf{1.5} * \boldsymbol{\sigma_R}$

- Step 1. We start with

$$0.8 = \delta_L \leq \mu_T - \mu_R \leq \delta_U = 1.25,$$

where μ_T and μ_R are reference mean and test mean (in log-scale), respectively.

- Step 2. For drug products with large variabilities (i.e., highly variable drug products), FDA recommends scaled average bioequivalence (SABE) criterion by adjusting the above bioequivalence limits for variability of the reference product (Haidar et al., 2008; Tothfalusi et al., 2009). This gives

$$0.8\sigma_R = \delta_L * \sigma_R \leq \mu_T - \mu_R \leq \delta_U * \sigma_R = 1.25 * \sigma_R$$

DukeMedicine

Justification of EAC $= \pm 1.5 * \sigma_R$

- Step 3. FDA assumes that the difference in mean is proportional to σ_R and allow a mean shift of

$$\frac{\sigma_R}{8} = 0.125,$$

which is half width of the margin. The worst possible scenario for the shift is that the true mean difference falls on $1.25 * \sigma_R$. In this case, FDA expands the margin $0.25 * \sigma_R$. Thus, the upper margin of EAC becomes

$$1.25 * \sigma_R + 0.25 * \sigma_R = 1.5 * \sigma_R.$$

DukeMedicine

Comments on Tier 1 approach

- Focus on *difference in means* (or ratio of means)
- Is the recommended equivalence test *consistent* with the usual bioequivalence test for PK parameters?
  - If not, why?
- Primary assumptions
  - Why difference in means is proportional to the standard deviation of the reference product, i.e., $\mu_T - \mu_R \propto \sigma_R$?
  - FDA allows a mean shift of $\frac{1}{8}\sigma_R$. Why?
  - Why EAC is selected as $1.5 * \sigma_R$?

DukeMedicine

Comments on Tier 1 approach

- How to estimate σ_R?
  - Although FDA's recommended approach (single test value from each reference lot) is an unbiased estimate of σ_R, it does not account for the variability of the estimate of σ_R (i.e., $\hat{\sigma}_R$)
  - Are there any benefit for having multiple test values from each reference lots
  - What if, by chance, best lots or worst lots are selected for establishing EAC?

DukeMedicine

FDA's recommended approach

- Suppose that there are *K* reference lots. Let $x_i, i = 1, 2., \ldots, K$ be the test result of the ith lot. $x_i, i = 1, 2, \ldots, K$ are assumed independently and identically distributed with mean μ_R and variance σ_R^2. FDA recommends that σ_R be estimated by

$$\hat{\sigma}_R = s = \sqrt{\frac{1}{K-1}\sum_{i=1}^{k}(x_i - \bar{x})^2}$$

- In fact, FDA assumes that $\mu_{Ri} = \mu_{Rj} = \mu_R$ and $\sigma_{Ri}^2 = \sigma_{Rj}^2 = \sigma_R^2$ for $i \neq j, i, j = 1,2, \ldots, K$.

DukeMedicine

Remarks

- $EAC = \pm c * \sigma_R = \pm 1.5 * \sigma_R$
- σ_R is estimated by s (sample standard deviation) which is obtained based on $x_i, i = 1, \ldots, K$
- s is an *unbiased* estimate of σ_R but it does not account for the *variability associated with* $s = \hat{\sigma}_R$
- $C = 1.5$ is a *regulatory constant* suggested by FDA
- The width of EAC depends upon the estimates of σ_R
- FDA indicates the choice of estimates of σ_R should be justified
- It is suggested that the *variability associated with* $s = \hat{\sigma}_R$ be taken into consideration to account for the *worst possible reference lots*.

DukeMedicine

Alternative estimates for σ_R

- Under the FDA's assumption, the expected value of $E(\bar{x}) = \mu_R$ and $Var(\bar{x}) = \sigma_R^2/K$.
- Under the assumption that $\mu_{Ri} \neq \mu_{Rj}$ and $\sigma_{Ri}^2 \neq \sigma_{Rj}^2$ for $i \neq j$, where μ_{Ri} and σ_{Ri}^2 be the mean and variance of the *i*th lot of the reference product. In this case, we have

$$\frac{\sigma_{(1)}^2}{K} \leq Var(\bar{x}) = \frac{\sigma_R^2}{K} \leq \frac{\sigma_{(K)}^2}{K},$$

where $\sigma_{(1)}^2$ and $\sigma_{(K)}^2$ are the smallest and largest within-lot variance among the *K* lots.

DukeMedicine

Alternative estimates for σ_R

- Current FDA's assumptions do not reflect real practices.
- Alternatively, may consider

$$\hat{\sigma}_R = \sqrt{\frac{n-1}{\chi^2_{1-\frac{\alpha}{2},n-1}}}\hat{\sigma}_x,$$

where $\hat{\sigma}_x$ is sample standard deviation obtained from the *n* reference lot test values and $\chi^2_{1-\frac{\alpha}{2},n-1}$ is the $(1-\frac{\alpha}{2})$th upper quantile of a chi-square distribution with n-1 degrees of freedom.

DukeMedicine

FDA's response to the question

- The equivalence margin is a key element for similarity assessment;
- It is very difficult to specify the margin:
  - A clinical or scientifically-justified margin is not available
  - No unified statistical method to determine the margin is desirable
  - We usually face a limited number of lots in submissions
- Thus, we need to develop a reasonable and feasible approach to determine the margin
- Our proposal:
  - Margin adjusted for sample size & variance
  - Larger variance ⇒ larger margin
  - Proper power for desirable similarity

FDA current thinking on equivalence test

Proposed Margin:

- Sample Size and Variance Adjusted Margin:
  - Give **lower** power γ* for **small** sample sizes
  - Solve for (symmetric) Margin so that Power(margin, Sample Sizes, True Mean Diff, Variability) = γ*

Rationale:

- A margin captures the product and quality attribute variability
- Allow additional acceptable CQA (critical quality attribute) shift $\boldsymbol{\delta_0} \geq 0$, with scientific justification;
- Not an approach with constant high power for all sample sizes;

Determination of equivalence margin

Proposed Margin - Sample Size and Variance Adjusted Margin

Step 1: Determine the variability of reference product σ_R^2

Our Proposed Margin: $\delta = \delta_0 + k\sigma_R$, k > 0

Remarks:

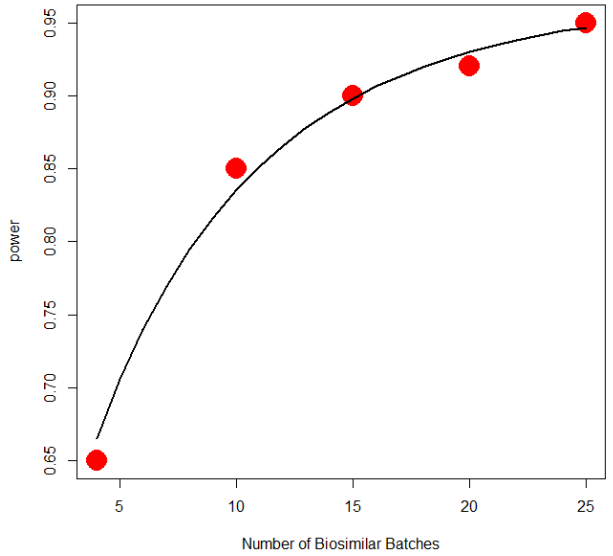
- σ_R^2 is the variability of the mean response from each lot
- σ_R^2 are subject to the approval from OBP reviewers for each CQA
- k is determined in order to have sufficient power for a targeted pre-specified highly similar product (i.e., differs from reference by Δ*), with practical sample sizes

Determination of equivalence margin

Proposed Margin : Sample Size and Variance Adjusted Margin

- Step 2: Consider by default, $\delta_0 = 0$
  Give **lower** power γ* when the number of biosimilar lots is small.

n_B	Power
4	65%
10	**85%**
15	90%
20	92%
25	95%



$$Power(n_B) = 1 - \exp(-0.53948 - 1.4694 n_B + 0.00205 n_B^{\,2})$$

Key concept:

- Provides high passing probability for "targeted" biosimilar product for practical sample sizes
- Also properly controls the passing probability for small sample sizes

Determination of equivalence margin

Deciding a target biosimilar product for margin setting:

Step 3: Compute the margin from the power function to achieve pre-assigned power at step 1 at a given number of batches, reference variability, using "targeted biosimilarity" using $\tau\,\sigma_R$ with = 1/16, 1/8, 3/16, ¼, and ½. After comparing power curves, "target biosimilar" is chosen $\mu_B - \mu_R = \sigma_R/8$.

Remarks:

- $\mu_B - \mu_R = \sigma_R/8$ was chosen as a reasonable value as highly similar; i.e. allowing for some difference; power assigned to approve for these quality.
- This approach also rewards a biosimilar product with good quality because the power would be higher than the pre-specified power when
  - the true mean difference is less than $\sigma_R/8$, or
  - the variability of biosimilar product is less than σ_R

Determination of equivalence margin

Proposed Margin : Sample Size and Variance Adjusted Margin

Calculation Examples: margin (δ) calculations are performed at equal sample sizes, equal variances, $\mu_B - \mu_R = \sigma_R/8$, type I error rate = 5% and power = assigned power at step 1;

$n_B = n_R$	4	5	6	7	8	9	**10**	11	12	13	14
Power	67%	71%	74%	77%	79%	82%	**84%**	85%	87%	88%	89%
δ/σ_R	2.11	1.89	1.74	1.64	1.55	1.50	**1.45**	1.39	1.36	1.33	1.30
$n_B = n_R$	15	16	17	18	19	20	21	22	23	24	25
Power	90%	91%	91%	92%	93%	93%	93%	94%	94%	94%	95%
δ/σ_R	1.27	1.25	1.21	1.20	1.19	1.16	1.13	1.13	1.11	1.09	1.09

- If we choose a consistent margin limit at $\delta = 1.5\sigma_R$, with $\mu_B - \mu_R = \sigma_R/8$ and $n_B = n_R = 10$, the passing probability of equivalence test (power) is recalculated to be 87%.
- This power increases with smaller mean difference.

II. Determination of Equivalence Margin (5)

- Confidence level used for fixed 1.50σ margin under equivalent sample sizes and variances

$n_T=n_R$	Margin	Power for $1/8\sigma_R$	Confidence Level (%)
6	$1.50\sigma_R$	78%	78.4
7	$1.50\sigma_R$	82%	81.4
8	$1.50\sigma_R$	84%	85.0
9	$1.50\sigma_R$	86%	87.4
10	$1.50\sigma_R$	87%	90.0
15	$1.50\sigma_R$	90%	90.0

$$\text{For } n_B \leq 10,\ Power(n_B) = 1 - \exp(-0.1139 - 0.2891 n_B + 0.0093 n_B^{\,2})$$

Summary and discussion

- Comparisons among margin determination methods for statistical equivalence test:

Fixed Margin	**Proposal**
Constant for all QAs; e.g. (80 ~ 125%) for ratio of means	Adjusted for sample sizes, variability, power
Tends to be too wide	Lower confidence level requirement for small sample sizes
Power depends on sample size	Control passing power for small sample sizes (e.g. 65%~90% for $n_b < 15$)
When variance is large, may need unreasonably large sample sizes;	Reward large sample sizes

Comments on FDA's response to question of sample size requirement

- Sample size is a function of
  - α (significance level),
  - $1-\beta$ (power),
  - $\delta = \mu_T - \mu_R$ (clinically meaningful difference), and
  - σ_R (variability of the reference product; it is usually assumed that $\sigma_R \approx \sigma_T$ though it is often not true)
  - That is, $n = f(\alpha, 1-\beta, \delta, \sigma_R)$

DukeMedicine

Comments on Tier 1 approach

- Fixed approach vs. random approach
  - Current proposal is a fixed approach depending upon the selected reference lots
  - The fixed approach is a conditional approach
- Sample sizes requirement for Tier 1 equivalence test
  - How many reference lots are required?
  - How many test lots are required?
  - Is there a need to match reference lots?

DukeMedicine

Fixed approach vs. random approach

- Fixed approach
  - Sensitive to the selected reference lots
  - Potential selection bias
  - Difficult to deal if the selected reference lots are either best lots or worst lots
  - Where to draw the line?
- Random approach
  - Reference lot follows similar distribution with similar but different mean and similar but different standard deviation
  - What if a selected reference lot is not biosimilar to another selected reference lot?

DukeMedicine

Sample size requirement

- Sample size is a function of
  - α (significance level),
  - $1-\beta$ (power),
  - $\delta = \mu_T - \mu_R$ (clinically meaningful difference), and
  - σ_R (variability of the reference product; it is usually assumed that $\sigma_R \approx \sigma_T$ though it is often not true)
  - That is, $n = f(\alpha, 1-\beta, \delta, \sigma_R)$

Sample size requirement

- In practice, it is not possible that we can select a sample size for achieving a desired power at a pre-specified level of significance while other parameters vary.
  - Under the assumption that $\mu_T - \mu_R = \frac{1}{8}\sigma_R$, the effect size adjusted for standard deviation is given by

$$\frac{\delta}{\sigma_R} = \frac{\mu_T - \mu_R}{\sigma_R} = \frac{\frac{1}{8}\sigma_R}{\sigma_R} = \frac{1}{8}$$

  - In this case, $n = f(\alpha, 1-\beta)$ which is independent of σ_R
  - In practice, the assumption that $\mu_T - \mu_R \propto \sigma_R$ cannot be verified.

DukeMedicine

FDA's current thinking

- For equivalence test in analytical similarity, we recommend sample size unbalance adjustment with a maximum ratio (Method B).
  - Encourage similar sample sizes
  - Reduce the impact of ad-hoc analysis after a failed test
- The proposed approach $n_R^* = \text{Min}(1.5 \times n_T, n_R)$:
  - Utilize all available information to estimate reference variabiliity and mean value
  - Controls the Type I Error probability
  - Ensure high power with decent mean difference
- Satterthwaite approximation is recommended for CI computation.

Chow, Song, and Bai (2016) proposal

- **Step 1: Selection minimum number of test lots required, n_T**
  - Under the assumptions that (i) $\mu_T - \mu_R = r\sigma_R = \frac{1}{8}\sigma_R$ and (ii) $\delta = EAC = 1.5 * \sigma_R$ (FDA's recommendation), we can determine n_T for achieving a desired power (i.e., $1-\beta$) at the α level of significance for various selections of k.
- **Step 2: Determination of k**
  - Depending upon the availability of the reference lots (N_R) and test lots (N_T), carefully evaluate the trade-off between controlling type I error rate ad achieving desired power with various selection of different ks.
- **Step 3: Selection of n_R lots from N_R available reference lots**
  - Once n_T and k have been determined, n_R can be obtained as $n_R = n_T/k$. The n_R lots, which will be randomly selected from the N_R available reference lots, will then be used for establishment of EAC for equivalence test.

Sample Size Requirement in Analytical Studies for Similarity Assessment

Shein-Chung Chow[1], Fuyu Song[2*], and He Bai[3]

[1]Duke University School of Medicine, Durham, North Carolina, USA
[2]Peking University Clinical Research Institute, Peking University, Beijing, China
[3]Department of Drug & Cosmetics Registration, China Food and Drug Administration, Beijing, China

Abstract

For assessment of biosimilar products, FDA recommends a stepwise approach for obtaining the totality-of-the-evidence for assessing biosimilarity between a proposed biosimilar product and its corresponding innovative biologic product (FDA, 2015). The stepwise approach starts with analytical studies for assessing similarity in critical quality attributes (CQAs) which are relevant to clinical outcomes at various stages of manufacturing process (Christl, 2015). For CQAs that are most relevant to clinical outcomes, FDA requires equivalence test be performed for similarity assessment based on an equivalence acceptance criterion that is obtained using single test value of some selected reference lots (Tsong, 2015). In practice, however, how many reference lots and test lots are necessarily selected for an unbiased and reliable assessment of similarity has become an important and interesting question to the sponsors of biosimilar products. To assist the sponsors, FDA proposes a rule for selection of the number of reference lots for establishment of EAC and consequently the number of test lots for equivalence test based on extensive simulation studies (Dong, Tsong, and Wang, 2016). This article not only provides statistical justification of the FDA's proposal, but also proposes an alternative method to the FDA's proposal sample size requirement for Tier 1 equivalence test.

Quality range approach for Tier 2 CQAs

- FDA suggests that analytical similarity be performed based on the concept of quality ranges, i.e., $\pm x\sigma$, where σ is standard deviation of reference product and x should be appropriately justified.
- Thus, the quality range of the reference product for a specific quality attribute is defined as

$$(\hat{\mu}_R - x\hat{\sigma}_R, \hat{\mu}_R - x\hat{\sigma}_R)$$

- Analytical similarity would be accepted for the quality attribute if a sufficient percentage of test lot values (e.g. 90%) fall within the quality range.

DukeMedicine

Comments on Tier 2 approach

- Focus on *population* (not population mean)
  - If test product is similar to the reference product, there is a high percentage of data will fall within the *quality range* constructed based on data observed from the reference product
  - We should expected there are about 95% (99%) of test data falls with ± 2 (3) standard deviations
  - *How to select an appropriate* x*?*
- FDA does not seem to allow *mean shift* for Tier 2 approach, why?
  - In practice, difference in means between test and reference could be substantial.

DukeMedicine

Comments on Tier 2 approach

- Tier 1 equivalence test supposes to be more rigorous than Tier 2 quality range approach. That is, passing Tier 1 test will pass Tier 2 test
  - In practice, *there is no guarantee that a given CQA which passes Tier 1 test will pass Tier 2 test and vice versa. Why?*
- Does FDA require all CQAs at Tier 2 pass the test?
  - If not, about what percentage of CQAs need to pass in order to pass Tier 2 test?
  - Are there any rule to follow?

DukeMedicine

Example 1 – First consider the case where $\mu_T \approx \mu_R$ and $\sigma_T \approx \sigma_R$. In this case, if we choose $x = 1.645$, we would expect 90% of the test results from test lots to lie within the quality range obtained based on the test values of the reference lots. This case is illustrated in Figure 2.

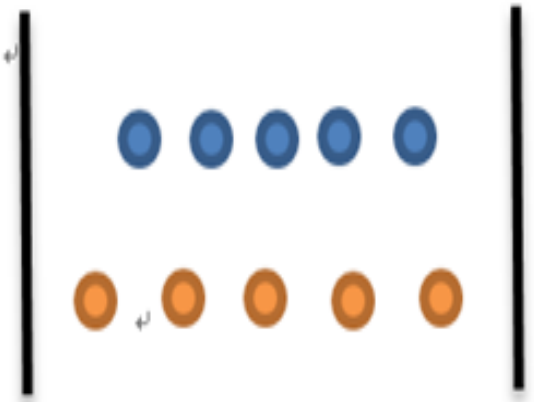


Figure 2. Quality Range Approach When $\mu_T \approx \mu_R$ and $\sigma_T \approx \sigma_R$

(blue represents test values of reference lots, while orange represents test values of test lots)

Example 2–When $\mu_T \approx \mu_R$ but $\sigma_T > \sigma_R$, if we choose $x = 1.645$, we would expect less than 90% of the test results from test lots to lie within the quality range obtained based on the test values of the reference lots. The percentage of test values from test lots decreases as $C = \sigma_T/\sigma_R > 1$ increases. This case is illustrated in Figure 3.

Figure 3. Quality Range Approach When $\mu_T \approx \mu_R$ and $\sigma_T > \sigma_R$
(blue represents test values of reference lots, while orange represents test values of test lots)

Example 3 – The case where $\mu_T > \mu_R$ but $\sigma_T \approx \sigma_R$ is illustrated in Figure 4. AS it can be seen from Figure 3, if we choose $x = 1.645$, we would expect less than 90% of the test results from test lots to lie within the quality range obtained based on the test values of the reference lots. The percentage of test values from test lots drop significantly if the difference between $\varepsilon = \mu_T - \mu_R$ increases (i.e., μ_T shifts away from μ_R).

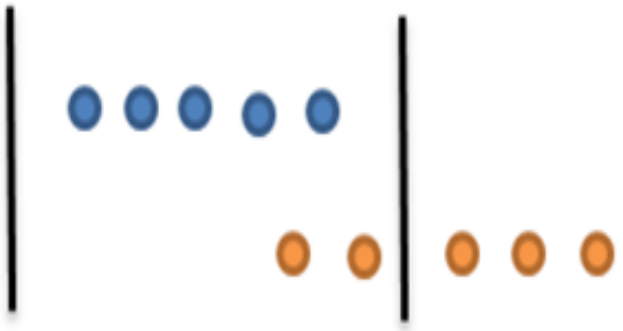


Figure 4. Quality Range Approach When $\mu_T > \mu_R$ and $\sigma_T \approx \sigma_R$
(blue represents test values of reference lots, while orange represents test values of test lots)

Example 4 – In practice, it is not uncommon to encounter the case where $\mu_T > \mu_R$ and $\sigma_T > \sigma_R$, which is illustrated in Figure 5. As it can be seen from Figure 4, if we choose $x = 1.645$, we would expect less than 90% of the test results from test lots to lie within the quality range obtained based on the test values of the reference lots. The percentage of test values from test lots could be very low, especially when both $C = \sigma_T / \sigma_R > 1$ and $\varepsilon = \mu_T - \mu_R$ increases.

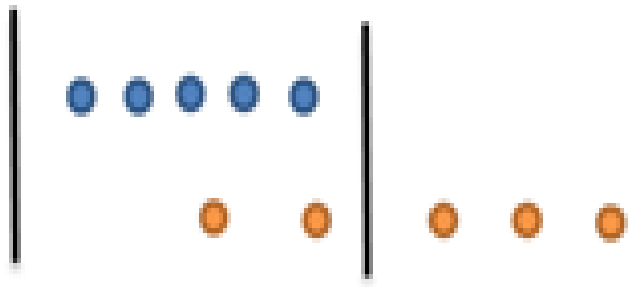


Figure 5. Quality Range Approach for the Case Where $\mu_T > \mu_R$ and $\sigma_T > \sigma_R$ (blue represents test values of reference lots, while orange represents test values of test lots).

Raw data and graphical comparison for Tier 3 CQAs

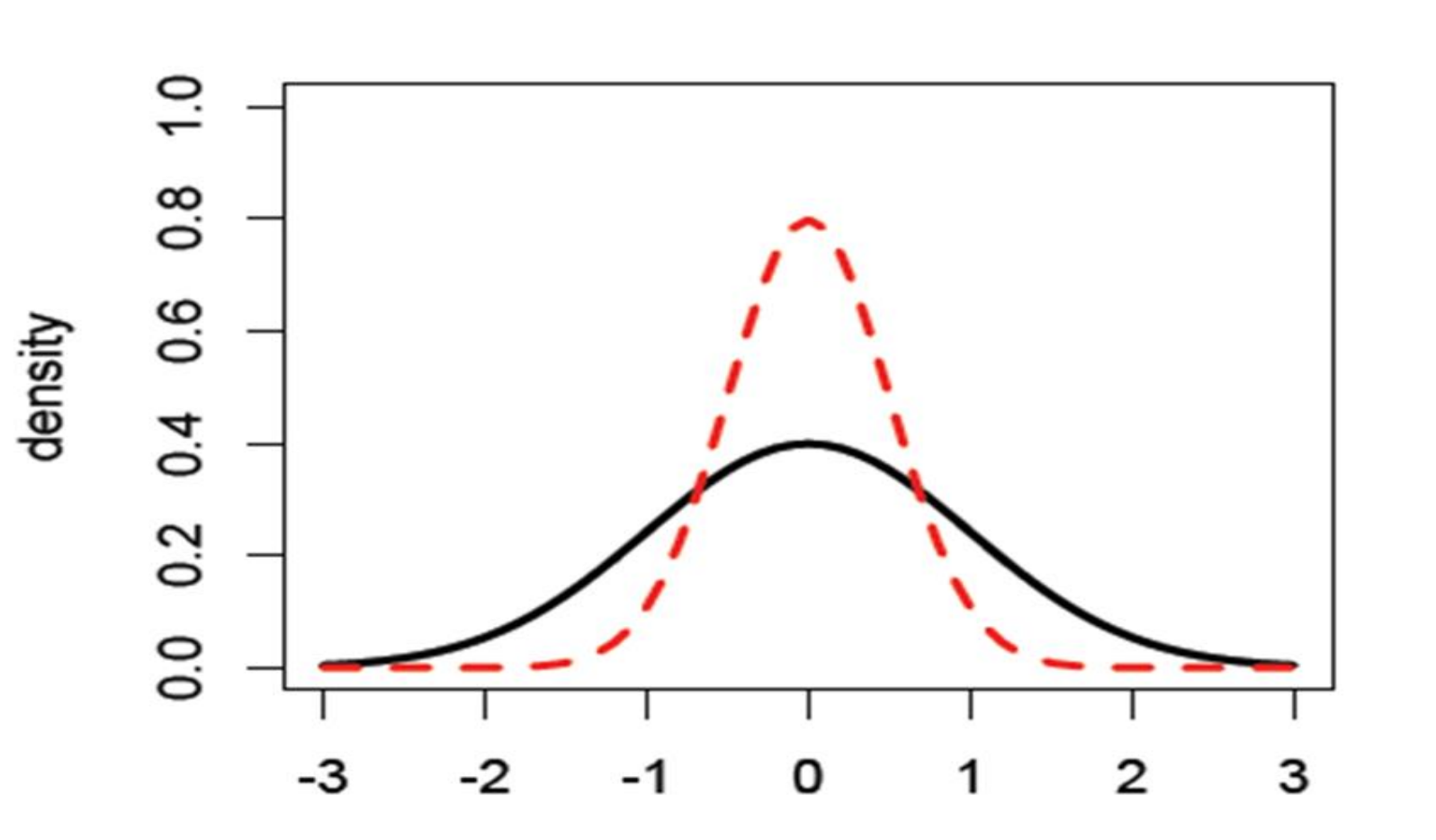
- For CQAs in Tier 3 with lowest risk ranking, FDA recommends an approach that uses raw data/graphical comparisons. The examination of similarity for CQAs in Tier 3 by no means is less stringent, which is acceptable because they have least impact on clinical outcomes in the sense that a notable dis-similarity will not affect clinical outcomes.

DukeMedicine

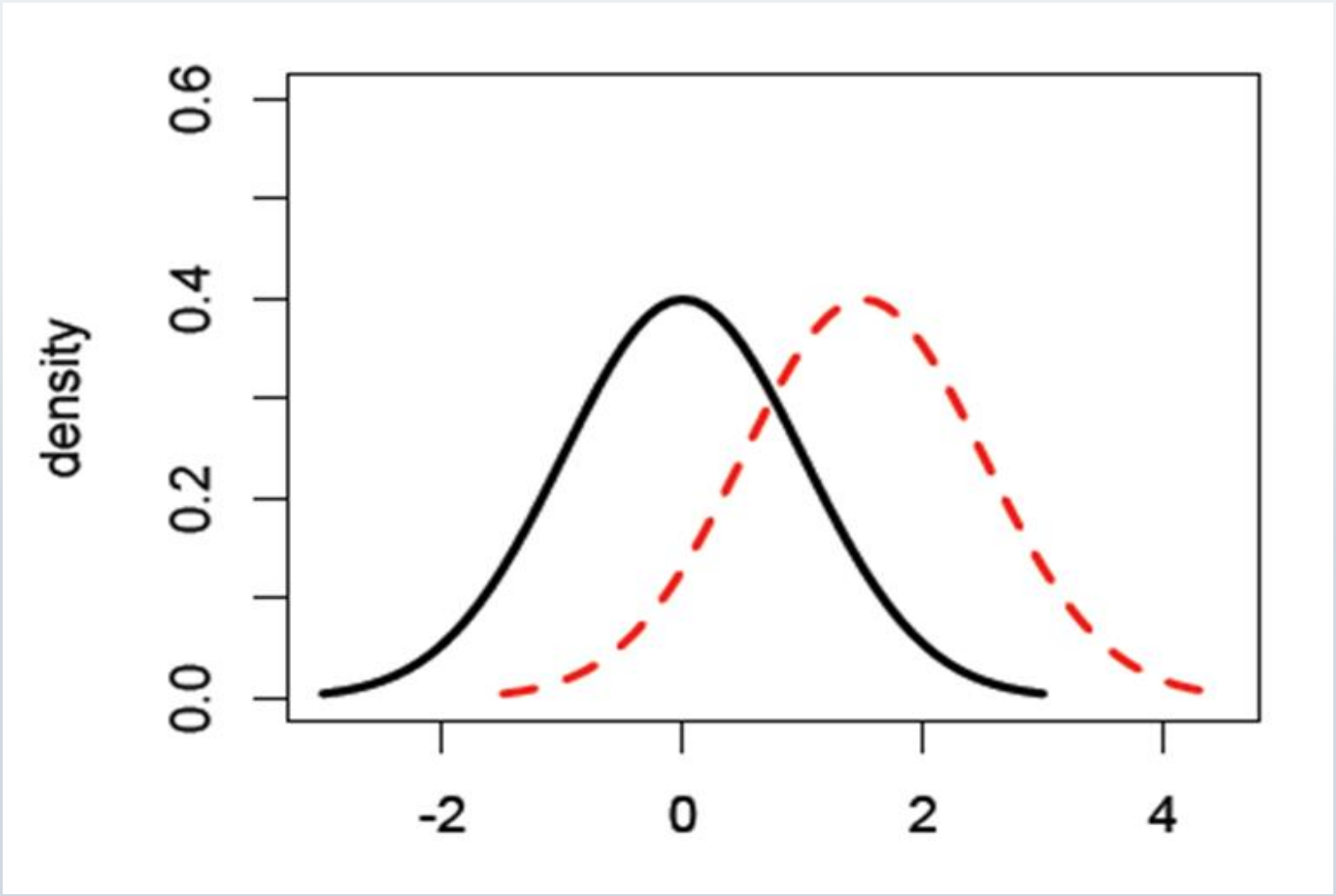
Comments on Tier 3 approach

- Evaluation based on raw data and graphical presentation, it is not only somewhat *subjective*, but also *biased*.
- Tier 1 and Tier 2 tests suppose to be more rigorous than Tier 3 approach. That is, passing Tier 1 and Tier 2 test will pass Tier 3 test
  - In practice, *there is no guarantee that a given CQA which passes Tier 1 or Tier 2 test will pass Tier 3 test and vice versa. Why?*
- Does FDA require all CQAs at Tier 3 pass the test?
  - If not, about what percentage of CQAs need to pass in order to pass Tier 3 test?
  - Are there any rule to follow?

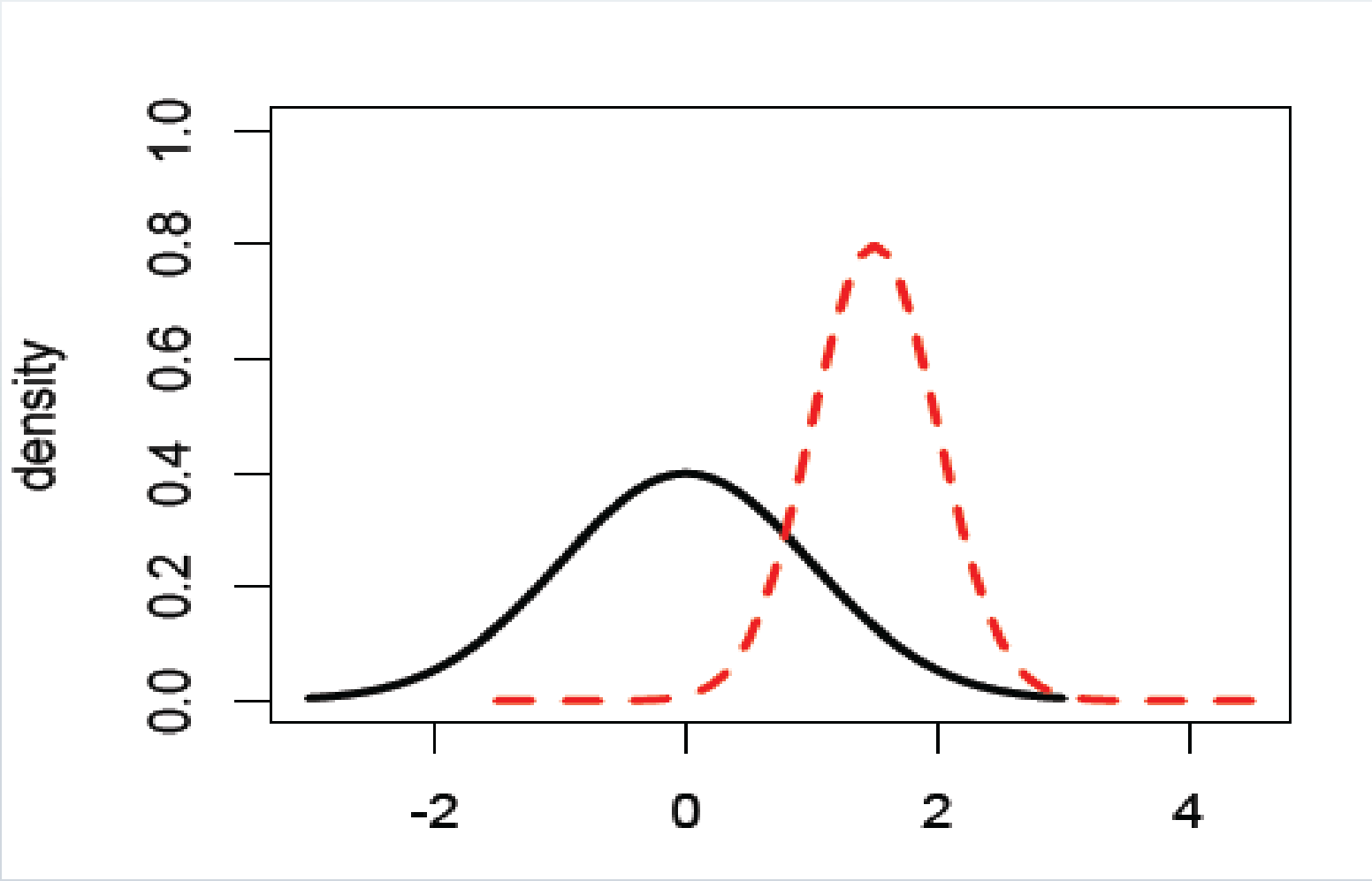
DukeMedicine



DukeMedicine



DukeMedicine



DukeMedicine

Comment on FDA's current thinking

- $EAC = \pm(1.5 * \sigma_R + \Delta)$
  - by default $\Delta = 0$
  - In practice, Δ depends upon scientific input (?)
  - Δ may be interpreted as the allowance in order to take into consideration of the worst possible reference lot (i.e., the lost with largest variability)
- Quality range approach for Tier 2 CQAs
  - The primary assumptions are that $\mu_T \approx \mu_R$ and $\sigma_T \approx \sigma_R$ (in practice, these assumptions are often not true)
  - FDA recommends x be chosen between 2 (95%) to 3 (99%)

DukeMedicine

The AAPS Journal (© 2016)
DOI: 10.1208/s12248-016-9882-5

Research Article

Analytical Similarity Assessment in Biosimilar Studies

Shein-Chung Chow,[1,4] Fuyu Song,[2] and He Bai[3]

Received 22 December 2015; accepted 27 January 2016

Abstract. For assessment of biosimilarity, the US Food and Drug Administration (FDA) recommends a stepwise approach for obtaining the totality-of-the-evidence for demonstrating biosimilarity between a proposed biosimilar product and an innovative (reference) biological product. The stepwise approach starts with analytical studies for functional and structural characterization at various stages of manufacturing process of the proposed biosimilar product. Analytical similarity assessment involves identification of critical quality attributes (CQAs) that are relevant to clinical outcomes. FDA proposes first classifying the identified CQAs into three tiers according to their criticality or risking ranking relevant to clinical outcomes and then performing equivalence test (for CQAs in Tier 1), quality range approach (for CQAs in Tier 2), and raw data or graphical presentation (for CQAs in Tier 3) for obtaining totality-of-the-evidence for demonstrating biosimilarity between the proposed biosimilar product with the reference product. In practice, some debatable issues are evitably raised due to this complicated process of analytical similarity assessment. In this article, these debatable are described and discussed.

KEY WORDS: equivalence test; fixed SD approach; quality range approach; stepwise approach; tiered approach; totality-of-the-evidence.

INTRODUCTION

In recent years, the assessment of biosimilarity for biosimilar products has received much attention by scientists, researchers, and reviewers from the pharmaceutical industry (biosimilar sponsors), academia, and regulatory ensure final product outputs remain within acceptable quality limits (see, e.g., (3)). For the analytical studies, FDA suggests that CQAs should be identified and classified into three tiers according to their criticality or risk ranking based on mechanism of action (MOA) or PK using appropriate statistical models or methods. CQAs with

Selected references

- Chow SC (2013). *Biosimilars: Design and Analysis of Follow-on Biologics*. Chapman and Hall/CRC Press, Taylor & Francis, New York, New York.
- Chow SC (2014). On Assessment of Analytical Similarity in Biosimilar Studies. *Drug Designing Open Access*, 3: e124. doi:10.4172/2169-0138.1000e124
- Chow SC (2015). Challenging Issues in Assessing Analytical Similarity in Biosimilar Studies. *Biosimilars*, 5, 33-39.
- Chow, S.C., Song, F.Y., Endrenyi, L. (2015). A note on Chinese draft guidance on Biosimilar Products. *Chinese Journal of Pharmaceutical Analysis*. 35(5), 762-767.
- Christl L (2015). Overview of regulatory pathway and FDA's guidance for development and approval of biosimilar products in US. Presented at FDA ODAC meeting, January 7, 2015, Silver Spring, Maryland.

DukeMedicine

DESIGN AND ANALYSIS OF BIOSIMILAR STUDIES

Annual Biosimilars Forum

All the 3 lectures of the whole course are available to download courtesy of Annual Biosimilars Forum event series at the Forum's official website: www.biosimsforum.com.

1

Lecture 1:

Assessing Biosimilarity: Issues and Recent Development

2

Lecture 2:

Assessing Interchangeability: Issues, Designs and Statistical Methods

3

Lecture 3:

Analytical Similarity Assessment in Biosimilar Studies

The event series will continue in 2017

Join us on 5-6 October, 2017 in Budapest for our 2nd Annual Biosimilars Forum regarding hot topics related to the drug development of Bio- and Nanosimilars with a strong scientific FOCUS ON Statistical and Regulatory perspectives. Visit:

www.biosimsforum.com

Annual Biosimilars Forum

The Annual Biosimilars Forum event series was founded in 2016 by two prestigious Central European scientific societies, the Viennese Section of the IBS and the Hungarian Society for Clinical Biostatistics in cooperation with the Accelsiors CRO Ltd., aimed at increasing effectiveness of clinical research and in order to provide even more effective support in sharing of recent scientific and practical knowledge for biosimilar drug development professionals.

Join us on 5-6 of October, 2017 at Budapest for the 2nd Biosimilars Forum and meet world's prominent biosimilar development experts!

CLICK HERE TO LEARN MORE!

ABOUT THE ORGANIZERS

The **Viennese Section of the International Biometric Society** is part of the ROeS, the Austrian Swiss Region of the International Biometric Society (IBS). WBS is an independent, non-profit organization which provides a professional forum for discussions of how to apply statistical methods in biological and medical science.

The **Hungarian Society for Clinical Biostatistics** is a national group of International Society for Clinical Biostatistics (ISCB), and it was founded to stimulate research into the principles and methodology used in the design and analysis of clinical research and to increase the relevance of statistical theory to the real world of clinical medicine.

Accelsiors Ltd. – as a scientific driven CRO – has been a committed supporter of biosimilar drug development, many of their professionals were involved into biosimilar drug development from the early beginnings, guided and managed the first biosimilar drug development projects and professionally supporting clinical trials as well as registration in this innovative field and being active in the clinical research arena in the past two decades.