On Analytical Similarity Assessment in Biosimilar Studies

Tongrong Wang, MS[1] and Shein-Chung Chow PhD[2]

[1] Master Candidate, Department of Biostatistics and Bioinformatics, Duke University School of Medicine

[2] Professor, Department of Biostatistics and Bioinformatics, Duke University School of Medicine

ABSTRACT

Abstract

For assessment of biosimilarity of biosimilar products, the United States (US) Food and Drug Administration (FDA) proposed a stepwise approach for providing totality-of-the-evidence of similarity between a proposed biosimilar product and a US-licensed (reference) product. The stepwise approach starts with assessment of critical quality attributes that are relevant to clinical outcomes in structural and functional characterization in manufacturing process of the proposed biosimilar product. FDA suggests that these critical quality relevant attributes be identified and classify into three tiers depending their criticality or risking ranking. To assist the sponsors, FDA also suggests some statistical approaches for assessment of analytical similarity for critical quality attributes (CQAs) from different tiers, namely equivalence test for Tier 1, quality range approach for Tier 2, and descriptive raw data and graphical comparison for Tier 3. Analytical similarity assessment for CQAs in Tier 1 is performed based on the equivalence acceptance criterion (EAC) which depends upon the estimate of variability of the reference product. The FDA's recommended approach often underestimates the variability of the reference product because it does not take the worst possible lots into consideration. In this poster, the statistical properties of the FDA's recommended approach is examined and alternative methods will be proposed in establishing a more accurate and reliable EAC for analytical similarity assessment.

Keywords:

Stepwise approach; Critical quality attribute (CQA); Equivalence test; Equivalence acceptance criterion (EAC).

CONTACT

Tongrong Wang

Email: tongrong.wang@duke.edu

ANALYTICAL SIMILARITY ASSESSMENT

For analytical similarity assessment of CQAs, FDA suggests that CQAs that are relevant to clinical outcomes be identified and classified into three tiers according to their criticality or risking ranking (e.g., most, mild to moderate, and least) relevant to clinical outcomes. FDA recommends equivalence test for CQAs from Tier 1, quality range approach for CQAs from Tier 2, and descriptive raw data and graphical presentation for CQAs from Tier 3 (Christl, 2015; Tsong, 2015).

Equivalence Test for Tier 1

For Tier 1, FDA recommends an equivalency testing for assessment of analytical similarity be performed. As indicated by the FDA, a potential approach could be a similar approach for bioequivalence testing for generic drug products (FDA, 2003; Chow and Liu, 2008). In other words, for a given critical attribute, we may test for equivalence of the following interval (null) hypothesis:

$$H_0: \mu_T - \mu_R < -\delta \text{ or } \mu_T - \mu_R > \delta$$

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. Analytical equivalence (similarity) is concluded if the null hypothesis of not-equivalence (notsimilarity) is rejected. Note that Yu (2004) defined in-equivalence as confidence interval falls entirely outside the equivalence limits. 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 (– δ, δ).

Quality Range Approach for Tier 2

For Tier 2, FDA suggests that analytical similarity be performed based on the concept of quality ranges, i.e., $\mu_R \pm x\sigma$, 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.

Raw Data and Graphical Comparison for Tier 3

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.

ESTABLISHMENT OF EAC

For analytical similarity assessment of CQAs in Tier 1, FDA recommends equivalence test under the null hypothesis. For testing the null hypothesis, FDA made the following assumptions. First, FDA assumes that the difference in mean responses between the reference product and the proposed biosimilar product is proportional to the variability of the reference product. In other words, $\Delta = \mu_T - \mu_R \propto \sigma_R$. Based on this assumption, FDA indicates that the equivalence limit is likely to be proportional the reference product variability as well, which means which means δ=c*σ_R. FDA chose c=1.5. However when establishing equivalency acceptance criterion (EAC= 1.5*σ_R), FDA consider the estimate of reference product variability obtained from the reference sample as a fixed constant, which may compromise the accuracy of the bioequivalence test.

Estimators of σ_R

Define x_{Rij} is the log transformed measured value of the j_{th} item in i_{th} lot and it follows normal distribution with mean μ_i and variance σ_i^2, where μ_i and σ_i^2 are also random variables. The expectations of μ_i and σ_i^2 are μ and σ^2 and the variances are σ_μ^2 and σ_σ^2. Then the variance of reference product

$$\sigma_R^2 = Var(x_{Rij}) = Var\left(E(x_{Rij}|\mu_i, \sigma_i^2)\right) + E\left(Var(x_{Rij}|\mu_i, \sigma_i^2)\right)$$
$$= Var(\mu_i) + E(\sigma_i^2) = \sigma_\mu^2 + \sigma^2$$

Define $\overline{x_{Ri\cdot}} = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{Rij}$, $\overline{\overline{x_{R\cdot\cdot}}} = \frac{1}{k}\sum_{i=1}^{k} \overline{x_{Ri\cdot}}$,

then FDA's estimation of σ_R^2 is

$$\widehat{\sigma_R^2} = \frac{1}{n'k-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{Rij} - \overline{\overline{x_{R\cdot\cdot}}})^2$$

$$E\left(\widehat{\sigma_R^2}\right) = \frac{1}{n'k-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Var(x_{Rij}) - Var(\overline{\overline{x_{R\cdot\cdot}}})\right)$$
$$= \sigma_\mu^2 + \sigma^2 - \frac{n'-1}{n'k-1}\sigma_\mu^2$$

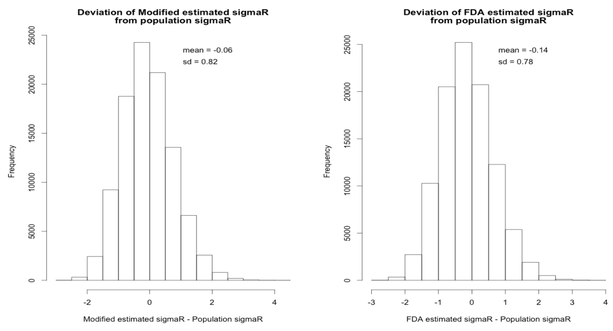
It is unbiased when $n' = 1$. But when $n' > 1$, the FDA approach tends to underestimate σ_R^2 leading to a conservative test which may not be a fair and reliable assessment of analytical similarity for a given quality attribute. Following is an alternative approach to adjust the estimate of σ_R^2 to unbiased.

$$\widehat{\sigma_R^2}' = \frac{1}{n'k}\sum_{i=1}^{k}\sum_{j=1}^{n'}(x_{Rij} - \overline{x_{Ri\cdot}})^2 + \frac{1}{k-1}\sum_{i=1}^{k}(\overline{\overline{x_{R\cdot\cdot}}} - \overline{x_{Ri\cdot}})^2$$

$$\begin{cases} \widehat{\sigma^2} = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k}\sum_{j=1}^{n'}\frac{(x_{Rij} - \overline{x_{Ri\cdot}})^2}{n'-1} \\ \widehat{\sigma_\mu^2} = \dfrac{1}{k-1}\displaystyle\sum_{i=1}^{k}(\overline{\overline{x_{R\cdot\cdot}}} - \overline{x_{Ri\cdot}})^2 - \frac{1}{kn'}\sum_{i=1}^{k}\sum_{j=1}^{n'}\frac{(x_{Rij} - \overline{x_{Ri\cdot}})^2}{n'-1} \end{cases}$$

SIMULATION RESULTS

The empirical distribution of $\widehat{\sigma_R'}$ and $\widehat{\sigma_R}$ under the same assumption is shown below. Every estimate is obtained using 10 references lots with sample size 3.



As an example, suppose that there are 20 RP lots and 6 TP lots. We first randomly select 6 out of the 20 RP lots to match the 6 TP lots. Also, suppose that the true difference between the biosimilar product and the reference product is proportional to σ_R.

Number of TP lots	$\frac{\mu_T - \mu_R}{\sigma_R}$	Test size (Confidence Interval)	Statistical Power using FDA estimated σ_R	Statistical Power using modified estimated σ_R	Statistical Power using population σ_R
6	0	5%(90%)	98.68%	98.62%	97.63%
6	1/8	5%(90%)	90.77%	90.53%	89.22%
6	1/4	5%(90%)	89.87%	90.22%	87.96%
6	1/2	5%(90%)	86.51%	86.23%	82.21%
6	1	5%(90%)	71.76%	71.28%	61.48%

From result above, the tests using estimated $\boldsymbol{\sigma_R}$ are more likely to conclude that the test product is similar to reference product than the test using the true population $\boldsymbol{\sigma_R}$, which leads to an inflated alpha. For instance, when the difference is 2 σ_R , i.e . the null hypothesis is true, the type 1 error of test using population $\boldsymbol{\sigma_R}$ is 10.09%. While, the type 1 error for the other two tests is respectively 30.5% and 29.8%.

REFERENCES

[1] FDA (2012).Scientific considerations in demonstrating biosimilarity to a reference product.The United States Food and Drug Administration, Silver Spring, Maryland, USA.

[2] Christl, L. (2015). Overview of the regulatory pathway and FDA's guidance for the development and approval of biosimilar products in the US. Presented at the Oncologic Drugs Advisory Committee meeting, January 7, 2015, Silver Spring, Maryland.

[3] Chow, S.C. and Liu, J.P. (2008).*Design and Analysis of Bioavailability and Bioequivalence Studies*.3rd edition, Chapman Hall/CRC Press, Taylor& Francis, New York, New York, USA.

[4] Chow, S.C. (2015). Challenging Issues in Assessing Analytical Similarity in Biosimilar Studies. Biosimilars, 5, 1-7.

[5] FDA (2003). Guidance on Bioavailability and Bioequivalence Studies for Orally Administrated Drug Products – General Considerations, Center for Drug Evaluation and Research, the US Food and Drug Administration, Rockville, Maryland, USA.

[6]. Haidar, S.H., Davit, B., Chen, M.L., Conner, D., Lee, L., Li, Q.H., Lionberger, R., Makhlouf, F., Patel, D., Schuirmann, D.J., and Yu, L.X. (2008). Bioequivalence approaches for highly variable drugs and drug products. *Pharmaceutical Research,* 25: 237-241.