Enliven Archive
Intensifying Thoughts

Research Article | www.enlivenarchive.org | Enliven: Biosimilars and Bioavailability

Confidence Interval Approach for Analytical Similarity Assessment - A Comparison between US FDA and EU EMA Approach

Yuqing Liu*, and Shein-Chung Chow

Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA

***Corresponding author**: Yuqing Liu, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, 2424 Erwin Road, Durham, North Carolina, USA, E-mail: yuqing.liu96@duke.edu

Received Date: 29th November 2021
Accepted Date: 15th December 2021
Published Date: 26th December 2021

Citation: Liu Y, Chow S-C. Confidence Interval Approach for Analytical Similarity Assessment - A Comparison between US FDA and EU EMA Approach. Enliven: Biosimilars Bioavailab. 2020; 5(1): 005.

Abstract

For analytical similarity assessment in biosimilar studies, a two one-sided tests (TOST) procedure and a confidence interval (CI) approach are commonly employed to demonstrate biosimilarity between a proposed biosimilar (test) product and an innovative (reference) biological product. TOST is considered an official method for biosimilarity assessment by most regulatory agencies including the United States (US) Food and Drug Administrations (FDA) and the European Medicines Agency (EMA) of the European Union (EU). Since TOST is operationally equivalent to the CI approach, the CI approach is often used for biosimilarity assessment. However, FDA and EMA use the CI approach based on relative change and absolute change in mean response, respectively. This difference may lead to different conclusions regarding biosimilarity between the test product and the reference product, which causes some confusion among sponsors, regulatory agencies, and researchers in academia. This article attempts to clarify the confusion points and compare the FDA and EMA's approach in terms of false positive rate and false negative rate when performing analytical similarity assessment.

Keywords: Equivalence test; Analytical similarity assessment; Sample size calculation

Introduction

It is well-known that the discovery, research, and development of medicine are complex and require a huge investment of time and money. However, once the patent of the original product expires, the approved biosimilar and generic drugs can be manufactured and enter the market. Due to competition, the prices of branded drugs are usually significantly reduced. Consequently, the approval of these drugs becomes particularly important and needs to be taken seriously. Biosimilarity, the active substance in the test drug is similar to the reference medicine in a high degree, is the goal of biosimilar development [1]. According to the Public Health Services Act, FDA [2] would only approve biological products that are demonstrated to be biosimilar on the safety, purity, and potency of the reference product under certain existing scientific knowledge. However, compare with the bioequivalence study on generic drugs, the small-molecule chemical synthesis, the biosimilarity evaluation of biologics are harder due to their large molecular size, complex molecular structure, and difficult manufacturing processes.

In biomedical research, various approaches are used to demonstrate biosimilarity to conduct the comparison of the proposed product and the reference product. Among them, TOST (two one-sided tests) is recommended by both FDA (the United States Food and Drug Administrations) and EMA (European Medicines Agency of European Union) as one of the formal statistical tests of equivalence for the statistical evaluation of analytical similarity and they used confidence interval approach to simultaneously test these two null hypotheses for convenient calculation [3,4]. However, FDA and EMA used a confidence interval approach with different statistical models, measurements, confidence levels, and criteria which have varied statistical properties [5]. Thus, the fundamental difference between the two statistical methods used by the FDA and EMA for similarity evaluation has caused some confusion among sponsors and regulatory agencies when choosing methods [6]. Meanwhile, it is not easy to compare the pros and cons of two methods with different standards and to choose the better one. Therefore, this paper will focus on the comparison between the US FDA approach and the EU EMA approach.

Therefore, the first aim of this article is to study the statistical properties of the confidence interval approach for analytical similarity assessment in biosimilar studies. Second, the purpose of this article is to compare the US FDA approach (based on 90% confidence interval on μ_T / μ_R) and the EU EMA approach (based on 95% confidence interval on $\mu_T - \mu_R$). By simulation, this article proposed a table of the conditional approved probability under the criteria under these two methods of similarity assessment for generic biosimilar products in biomedical research. At the same time, it provides the relationship between the requirement of sample size and desired power. From the article structure aspect, the definition and characteristic of the confidence interval approach of biosimilar testing based on parallel design and standard 2×2 crossover design will be given in the next section. In section 3, a comparison of the US FDA approach and the EU EMA approach will be briefly outlined, whereas the procedures for the calculation of two conditional probabilities will be purposed. Then, statistical properties and justification of the proposed procedure will be studied, which is illustrated applied and verified through a simulation study in section 4. Afterward, the discussion and conclusion remarks will be provided in the last section of this article.

Confidence Interval Approach and Two One-Sided Tests

The confidence interval approach is the method to determine the criterion of the bioequivalence and biosimilarity comparison. Generally, there are two confidence interval approaches, mean difference and geometric mean ratio, which work on two different study designs, crossover design and parallel design.

The confidence interval approach claims the bioequivalence between the reference product and test product when the (1-2α) ×100% confidence interval for the mean difference or the ratio of averages totally falls in the bioequivalence or biosimilarity limit. The concept of using confidence intervals to assess bioavailability was first suggested by Westlake [16], who proposed the symmetrical confidence intervals method that modified the conventional confidence interval of the mean difference of two normal populations to symmetric to zero for bioequivalence trials [17]. After discussion and research many scholars [8,9], Westlake modified the confidence interval approach from (1-α) ×100% confidence interval into (1-2α) ×100% confidence interval to be consistent with the standard of the efficacy testing of FDA [7].

To compare the pharmacokinetic measures of generics and biosimilars, the two one-sided tests procedure was recommended by FDA as the official statistical method for testing interval hypotheses of bioequivalence or biosimilarity since then. The TOST tests procedure was proposed by Schuirmann [10] for the assessment of the equivalence of average bioavailability that separates the interval hypotheses into two one-sided hypotheses. Since the TOST procedure is a size-α test [11], two one-sided tests with α significant level for each side is algebraically equivalent to (1− 2α) ×100% CI approach in many cases. However, TOST is based on the power while the CI approach is based on type I error, so the mixed up of these two methods may cause the confusion in some practical cases [12].

At the same time, the average bioequivalence approach and the log-transformation of pharmacokinetic data were also suggested [13]. Since then, there are several rules have been proposed to set the bioequivalence limit for the average bioequivalence assessment. This article focuses on two mainstream rules, ±20 rule and 80/125 rule, which are generally used to assess raw data and log-transformed data, respectively.

Considering the study design, the FDA guideline suggested using the crossover design to conduct the comparison between two formulations or two test conditions while the parallel design may be more appropriate for the long half-life drug [14]. For parallel design, suppose Y_{ik}, i=1,... n_k, k=1,2, where Y_{i1} and Y_{i2} are the response of the test lots and the reference lots, respectively. For crossover design, since we focus on the average bioequivalence approach, the replicated crossover design is not necessary [14]. Thus, 2×2 crossover design is the research subject of this this article, and then use general linear model to calculate the confidence interval. Suppose Y_{ijk}, i = subject 1,...n_k, j = period 1, 2, k = sequence 1,2 are the response of two sequences RT and TR, respectively. For the confidence interval approach, let [L,U] be the (1− 2α) ×100% confidence interval for μ_T/μ_R or $\mu_T - \mu_R$. If confidence interval is entirely within BE limits or BS margin, then we claim that test formulation and reference formulation are bioequivalence/biosimilar [15]. Table 1 demonstrates the criteria of the confidence interval approach, where

$$S_p^2 = \frac{\sum_{k=1}^{2}\sum_{i=1}^{n_k}\left(Y_{ik}-\frac{\sum_{i=1}^{n_k}Y_{ik}}{n_k}\right)^2}{n_1+n_2-2}$$

Table1. BE/BS Limit of Confidence Interval Approach

Hypothesis	Confidence Interval	BE/BS Limit
Crossover		
$\mu_T - \mu_R$	$[L,U] = \hat{\mu}_T - \hat{\mu}_R \pm t_{1-\alpha,n_1+n_2-2}\sqrt{\hat{\sigma}_d^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$	$[L,U] \subseteq (-20\%\,\hat{\mu}_R, 20\%\,\hat{\mu}_R)$
μ_T/μ_R*	$[L,U] = \exp\left[\hat{\mu}_T - \hat{\mu}_R - t_{1-\alpha,n_1+n_2-2}\sqrt{\hat{\sigma}_d^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right]$	$[L,U] \subseteq (80\%, 125\%)$
Parallel		
$\mu_T - \mu_R$	$[L,U] = \hat{\mu}_T - \hat{\mu}_R \pm t_{1-\alpha,n_1+n_2-2}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$	$[L,U] \subseteq (-20\%\,\hat{\mu}_R, 20\%\,\hat{\mu}_R)$
μ_T/μ_R*	$[L,U] = \exp\left[\hat{\mu}_T - \hat{\mu}_R \pm t_{1-\alpha,n_1+n_2-2}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right]$	$[L,U] \subseteq (80\%, 125\%)$

* Log-transformation prior to data analysis

is pooled variance of the subjects from two treatments for parallel design, and

$$\hat{\sigma}_d^2 = \frac{\sum_{k=1}^{2}\sum_{i=1}^{n_k}\left[\frac{1}{2}(Y_{i2k}-Y_{i1k})-\frac{\sum_{i=1}^{n_k}\frac{1}{2}(Y_{i2k}-Y_{i1k})}{n_k}\right]^2}{n_1+n_2-2}$$

is the pooled sample variance of period differences from both sequences for crossover design.

It seems reasonable to transform the confidence interval of the mean difference into the confidence interval of the ratio of two treatment mean by dividing the mean of the reference group and plus 1, which means $\frac{\mu_T-\mu_R}{\mu_R}+1=\frac{\mu_T}{\mu_R}$. However, these two methods are different because usually we could only use the sample mean estimate of the reference group instead of the true μ_R, which may lead to different conclusions [15].

By the definition of (1-2α) ×100% confidence interval, there is (1-2α) ×100% confident that the true value of the difference of mean or the geometric mean ratio in the confidence interval. However, with limited sample size, this probability of confidence interval contain true value is unequal to the probability of the (1-2α) ×100% confidence interval within the BE/BS limit [15]. Let p be the probability that (1-2α) ×100% confidence interval within BE/BS limit. Select an appropriate sample size (i.e., n_1 and n_2, without loss of generality, assume $n_1 = n_2 = n$) such that $p \geq p_0$ when p_0 is desired probability of concluding BE or BS. Based on this probability, next section will compare the difference between the analytical similarity assessment of US FDA and EU EMA.

$$1-2\alpha = P(\mu_T-\mu_R \text{ or } \mu_T/\mu_R \in [L,U]) > P([L,U] \subseteq (-20\%\,\hat{\mu}_R, 20\%\,\hat{\mu}_R) or (80\%, 125\%)) = p$$

A Comparison between US FDA and EU EMA with Simulation Study

As the description in section 2, FDA suggested logarithmic transformation of pharmacokinetic data on the 2003 guidance [13]. Nowadays, the log-transformation method is the only approach certified by FDA for analytical similarity assessment. However, the methods and standards of the assessment of generic drugs and biological products remain diverse around the world. For example, the procedures of two confidence interval approaches used by FDA and EMA have already been given in detail, however, these methods are based on different statistical models, measurements, confidence levels, and criteria as shown in the following table 2 [6]. The most obvious difference between these two methods is that they use different confidence interval approach with different significant level.

A simulation of AUC data with 1000 data sets each was used to demonstrate the comparison of the FDA approach and EMA approach under parallel design and 2×2 crossover design. Assume the AUC of test treatment and reference treatment are followed the normal distribution with true mean 80 and 82 with no period effect and carryover effect. The sample size for two treatments in parallel design and two sequences in 2×2 crossover design was considered to be the same, starting from 12 to 50, since 12 is the minimum number of evaluable subjects in BE study [3].

First, these data were used to show the relationship of the inequation mentioned in section 2, that is the probability $p < 1 - 2\alpha$. From the result in Table 3, the accuracy of the confidence interval approach to assess bioequivalence decreases as the variability increases.

Secondly, to give a more intuitive comparison, we compare the two methods in terms of the following theoretically, where FDA+ and EMA+ represent conclude bioequivalence/biosimilarity under FDA and EMA methods, respectively. Similarly, FDA- and EMA+ means reject the bioequivalence/biosimilarity under FDA and EMA methods, respectively. Table 4 shows four possible results of BE/BS assessment under FDA and EMA methods. Let $p_1,...p_4$ be the probability of each result. Based on this, two conditional probabilities are generated to compare the FDA and EMA approaches.

$$p_A = \text{P(US FDA claim BE/BS | EU EMA claim BE/BS)} = \text{P(FDA +|EMA+)} = \frac{p_1}{p_1+p_2}$$

$$p_B = \text{P(EU EMA claim BE/BS | US FDA claim BE/BS)} = \text{P(EMA + |FDA +)} = \frac{p_1}{p_1+p_3}$$

Table 2 Comparison of analytical similarity assessment between US FDA and EU EMA

Characteristics	US FDA	EU EMA
Statistical Model	Log-transformed Model	Raw data Model
Official Method	TOST ($\alpha = 0.05$)	TOST ($\alpha = 0.025$)
Sample Size requirement	Based on TOST	Based on TOST
Interval Hypotheses Testing	$H_{01}: \mu_T/\mu_R \leq 0.8$ $H_{02}: \mu_T/\mu_R \geq 1.25$	$H_{01}: \mu_T - \mu_R \leq -0.2\mu_R$ $H_{02}: \mu_T - \mu_R \geq 0.2\mu_R$
Criteria of CI Approach	(1-2α) × 100% CI for μ_T/μ_R $\subset (80\%, 125\%)$	(1-2α) × 100% CI for $\mu_T - \mu_R$ within $\pm 20\%$ of μ_R

Note: TOST = two one-sided tests, α = level of significance

Table 3: Comparison between the probability of CI contain 0 and within certain limit

		$\sigma/\mu = 0.2$		$\sigma/\mu = 0.3$		$\sigma/\mu = 0.4$	
Study Design	Sample Size	$p_{0\in CI}$	p	$p_{0\in CI}$	p	$p_{0\in CI}$	p
Parallel	12	0.947	0.322	0.945	0.009	0.942	0.000
EMA	16	0.943	0.565	0.946	0.035	0.948	0.001
	20	0.939	0.730	0.948	0.130	0.951	0.006
	30	0.931	0.918	0.936	0.443	0.944	0.064
	40	0.910	0.980	0.933	0.669	0.940	0.260
	50	0.928	0.997	0.940	0.823	0.947	0.451
Parallel	12	0.895	0.628	0.897	0.091	0.893	0.003
FDA	16	0.885	0.782	0.902	0.191	0.903	0.014
	20	0.878	0.870	0.887	0.328	0.906	0.044
	30	0.873	0.970	0.890	0.550	0.894	0.146
	40	0.842	0.993	0.888	0.735	0.901	0.296
	50	0.863	0.999	0.887	0.828	0.902	0.450
Crossover	12	0.931	0.795	0.939	0.233	0.943	0.027
EMA	16	0.923	0.942	0.943	0.455	0.950	0.099
	20	0.934	0.979	0.949	0.637	0.947	0.231
	30	0.897	0.999	0.938	0.894	0.941	0.605
	40	0.890	1.000	0.919	0.965	0.932	0.797
	50	0.866	1.000	0.904	0.993	0.913	0.878
Crossover	12	0.869	0.897	0.889	0.402	0.899	0.074
FDA	16	0.874	0.973	0.888	0.609	0.899	0.186
	20	0.866	0.989	0.885	0.698	0.890	0.269
	30	0.843	1.000	0.871	0.902	0.889	0.540
	40	0.803	1.000	0.861	0.956	0.893	0.719
	50	0.798	1.000	0.854	0.990	0.881	0.821

Note: $p_{0\in CI}$ is the probability of CI contain *0, p* is the probability of CI within BE/BS limit

Table 4: Probability of conclude BE/BS

	FDA+	FDA-
EMA+	p_1	p_2
EMA-	p_3	p_4

Then, these two conditional probabilities could be estimated through the same simulation data sets. From the simulation result in Table 5, when the sample size is small, $p_B < p_A$ for both two study designs, then with the increase of sample size, p_B get close to p_A and even larger than p_A. Therefore, EMA approach is more likely to approve the bioequivalence and biosimilarity for the small sample size. While EMA approach and FDA approach have similar probabilities to approve the bioequivalence and biosimilarity for a large size. Table 6 shows the relationship between power and sample size of TOST for different study designs under fixed coefficient of variations.

Table 5: Simulation result

		$\sigma/\mu = 0.2$		$\sigma/\mu = 0.3$		$\sigma/\mu = 0.4$	
Study Design	Sample Size	p_A	p_B	p_A	p_B	p_A	p_B
Parallel	12	0.997	0.511	1.000	0.099	NA	0.000
	16	0.998	0.721	1.000	0.183	1.000	0.071
	20	0.997	0.837	0.962	0.381	1.000	0.136
	30	1.000	0.946	0.883	0.711	0.656	0.288
	40	1.000	0.987	0.948	0.863	0.669	0.588
	50	1.000	0.998	0.953	0.947	0.736	0.738
Crossover	12	0.999	0.885	0.940	0.545	0.741	0.270
	16	0.999	0.967	0.927	0.693	0.747	0.398
	20	0.998	0.988	0.929	0.848	0.671	0.576
	30	1.000	0.999	0.962	0.953	0.764	0.856
	40	1.000	1.000	0.987	0.996	0.867	0.961
	50	1.000	1.000	0.995	0.998	0.912	0.976

Note: Sample Size: $n_1=n_2=n$, $p_A=1$ means $p_2=0$, $p_A= NA$ means $p_1= p_2=0$.

Table 6. Power of TOST for different sample size and study design under fixed coefficient of variations.

	FDA (α=0.05)		EMA (α=0.025)	
Sample Size	Parallel	Crossover	Parallel	Crossover
12	18.61	56.60	3.17	22.89
16	34.78	73.54	7.62	41.80
20	49.70	83.47	15.68	57.39
24	61.05	89.60	26.32	68.48
26	65.60	91.76	31.83	72.81
30	72.98	94.86	42.02	79.73
32	76.00	95.95	46.51	82.51
36	80.99	97.51	54.32	87.02
40	84.92	98.48	60.81	90.44
48	90.50	99.45	70.89	94.91
50	91.54	99.58	72.95	95.67
60	95.30	99.89	81.24	98.12
78	98.43	99.99	90.46	99.61

Note: Assume CV (coefficient of variations) $= \frac{\sqrt{MSE}}{\mu_R} \times 100\% = 20\%$.

Discussion and Conclusion

In this study, we take statistical approaches towards studying the properties of the confidence interval approach and comparing the differences between the EMA approach and FDA approach that were used to assess biosimilar. To investigate the statistical features of the confidence interval approach for assessing analytical similarity in biosimilar studies, we illustrated the criterion of the mean difference and geometric mean ratio confidence interval approaches based on parallel design and standard 2×2 crossover design. In addition, one of our primary aims was to compare the US FDA approach and the EU EMA. From the data requirement perspective, the EMA approach has an easier evaluation in data processing without log-transformation that is required by the FDA approach. However, the bioequivalence limit of the EMA approach contains the mean of the reference drug, which is unknown and has to be estimated by the sample, adding to the uncertainty of its result. Furthermore, the simulation results of the probability of concluding bioequivalence/biosimilar show that the FDA confidence interval approach has larger accuracy than the EMA approach with the same study design, sample size, and variability. In terms of false positive and false negative rates, the EMA approach and FDA approach have similar results in approving bioequivalence and biosimilarity for the large sample size, whereas the EMA approach has a higher probability of ratifying BE/BS for the small sample size, particularly for the parallel design, which requires a larger sample size to achieve desired power of demonstration bioequivalence with the same coefficient of variation. Therefore, the FDA method is more reliable for analytical similarity, especially when the bioequivalence is marginal.

However, in our attempt to uncover the differences between EMA and FDA CI approach, we find the comparison remains some limitations. Thereinto, this article focuses on the average bioequivalence, which is based on the population average of the bioequivalence measure and is recommended by both FDA and EMA. However, it does not consider the comparisons of the variances for the compared formulations. In addition, we discuss the FDA confidence interval approach and EMA confidence interval approach for analytical similarity assessment under two-arm parallel design and 2×2 crossover design. In practice, with the increased complexity of study designs, such as period effect, carryover effect, and replication, and other uncontrollable factors such as unbalanced sample size, the average bioequivalence method maybe not be appropriate. In order to fill these gaps, FDA gave two new approaches, the population bioequivalence approach and the individual bioequivalence approach, which included total variability, within-subject variability, and the subject-by-formulation interaction [13]. Therefore, the availability of conclusions for more complex situations needs further study.

References

1. EMA. Biosimilars in the EU: information guide for healthcare professionals. European Medicines Agency (EMA), Amsterdam, Netherlands. 2017.

2. FDA. Guidance for industry – Scientific Considerations in Demonstrating Biosimilarity to a Reference Product. Maryland: FDA. Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER), the United States Food and Drug Administrations, Silver Spring, Maryland. 2015.

3. FDA. Guidance for Industry – Statistical Approaches to Evaluate Analytical Similarity. Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER), the United States Food and Drug Administrations, Silver Spring, Maryland. 2017.

4. EMA. Guideline on the investigation of bioequivalence. CPMP/EWP/QWP/1401/98 Rev. 1/Corr**, London, 20 January 2010.

5. Chow SC, Song F, Bai H. Analytical similarity assessment in biosimilar studies. AAPS Journal. 2016, 18: 670-677.

6. Chow SC, Lee SJ. Current issues in analytical similarity assessment. Statistics in Biopharmaceutical Research. 2020, 13: 203-209.

7. Westlake WJ. Response to Kirkwood TBL: Bioequivalence testing - A need to rethink. Biometrics. 1981, 37: 589-594.

8. Metzler CM. Bioavailability: A problem in bioequivalence. Biometrics. 1974, 30: 309-317.

9. Kirkwood TB. Bioequivalence testing - A need to rethink. Biometrics. 1981, 37: 589-591.

10. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987, 15: 657-680.

11. Chow SC, Shao J. A note on statistical methods for assessing therapeutic equivalence. Control Clin Trials. 2002, 23: 515-520.

12. Chow SC, Zheng JY. The use of 95% CI or 90% CI for drug product development — a controversial issue? J Biopharm Stat. 2019, 29: 834-844.

13. FDA. Guidance for Industry – Bioavailability and Bioequivalence Studies for Orally Administered Drug Products — General Considerations. Center for Drug Evaluation and Research (CDER), U.S. Food and Drug Administration, Rockville, Maryland. 2003.

14. FDA. Guidance for Industry – Development of Therapeutic Protein Biosimilars: Comparative Analytical Assessment and Other Quality-Related Considerations. Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER), the United States Food and Drug Administrations, Silver Spring, Maryland. 2019.

15. Chow SC, Liu JP. Design and Analysis of Bioavailability and Bioequivalence Studies. 3rd Edition, Chapman and Hall/CRC Press, Taylor & Francis, New York. 2008.

16. Westlake WJ. Use of confidence intervals in analysis of comparative bioavailability trials. J Pharm Sci. 1972, 61: 1340-1341.

17. Westlake WJ. Symmetric confidence intervals for bioequivalence trials. Biometrics. 1976, 32: 741-744.