# Estimators and Confidence Intervals of f_2 Using Bootstrap Methodology for the Comparison of Dissolution Profiles

**Xu Z[1], Merino San Juan M[1,2], Mangas San Juan V[1,2], García Arieta A[3]**

[1] Department of Pharmacy and Pharmaceutical Technology and Parasitology, University of Valencia, Valencia, Spain.

[2] Interuniversity Research Institute for Molecular Recognition and Technological Development, Polytechnic University of Valencia–University of Valencia, Spain.

[3] División de Farmacología y Evaluación Clínica, Departamento de Medicamentos de Uso Humano, Agencia Española de Medicamentos y Productos Sanitarios, Spain.

## 1. Introduction

Dissolution tests are essential in the development of medicinal product, but there are many methods to compare dissolution profiles. The most widely used one is the similarity factor f_2. Nevertheless, the f_2 method has several drawbacks, which lead to certain restrictions described in regulatory guidelines, e.g. when variability of dissolution data is more than 20 % and 10 % for early and later time points, respectively, the f_2 method cannot be used. In such circumstance, alternative methods are recommended in several regulatory guidelines, for instance, the model-independent multivariate statistical distance methods (MSD) [1–2]. However, members from US FDA indicated in 2013 that the model-independent MSD method is less discriminative and sensitive than the f_2 method. Therefore, they recommended the confidence interval of f_2 approach using bootstrap [3]. Recent studies comparing the MSD method with the confidence interval of f_2 approach with bootstrap method confirmed these findings [4–5]. However, the guidelines neither specify the estimator nor the type of confidence interval to be used and literature with this regard is scarce. Therefore, we investigated the accuracy and precision of several estimators and types of confidence intervals by simulation.

## 2. Materials and Methods

One million individual dissolution profiles of the test and the reference product were simulated for each of the following target population f_2 values of 25, 35, 45–75. Random samples of size 6, 12, 18 or 24 units were chosen to estimate the f_2 value and its confidence intervals with 5000 bootstraps. Comparisons with low and high variability were also included. The whole process is repeated 10000 times, as shown in figure 1. Five f_2 estimators as shown in Equation 1 to 5 were calculated and fourteen confidence intervals were investigated according to literature: Normal interval, basic interval, percentile interval (Type 1 to Type 9, and another one using the interpolationmethod from R's boot package), two bias-corrected and accelerated (BCa) intervals (one by regression and another by jackknife). Accuracy was expressed as bias and precision was evaluated by the root of mean square error. In addition, type I error rate was evaluated by calculating the percentage of similarity for each target population.

**Figure 1.** Workflow of simulation. HV and LV denots high and low variability, respectively

## 3. Results

Results of accuracy and precision of the estimators are summarized in figure 2 showing bias and root of mean square error. All percentile results are similar, therefore, Type 7 was used as a representative one. Percentage of similarity was represented in figure 3 for population f2 values of 45 to 55 with a sample size of 12 units.

$$\hat{f}_2 = 100 - 25 \log \left( 1 + \frac{1}{P} \sum_{i=1}^{P} \left( \bar{X}_{T,i} - \bar{X}_{R,i} \right)^2 \right) \quad (1)$$

$$\hat{f}_{2,bc} = 100 - 25 \log \left( 1 + \frac{1}{P} \left( \sum_{i=1}^{P} \left( \bar{X}_{T,i} - \bar{X}_{R,i} \right)^2 - \frac{1}{n} \sum_{i=1}^{P} \left( S_{T,i}^2 + S_{R,i}^2 \right) \right) \right) \quad (2)$$

$$\hat{f}_{2,vcbc} = 100 - 25 \log \left( 1 + \frac{1}{P} \left( \sum_{i=1}^{P} \left( \bar{X}_{T,i} - \bar{X}_{R,i} \right)^2 - \frac{1}{n} \sum_{i=1}^{P} \left( w_{T,i} S_{T,i}^2 + w_{R,i} S_{R,i}^2 \right) \right) \right) \quad (3)$$

$$\hat{f}_{2,exp} = 100 - 25 \log \left( 1 + \frac{1}{P} \left( \sum_{i=1}^{P} \left( \bar{X}_{T,i} - \bar{X}_{R,i} \right)^2 + \frac{1}{n} \sum_{i=1}^{P} \left( S_{T,i}^2 + S_{R,i}^2 \right) \right) \right) \quad (4)$$

$$\hat{f}_{2,vcexp} = 100 - 25 \log \left( 1 + \frac{1}{P} \left( \sum_{i=1}^{P} \left( \bar{X}_{T,i} - \bar{X}_{R,i} \right)^2 + \frac{1}{n} \sum_{i=1}^{P} \left( w_{T,i} \cdot S_{T,i}^2 + w_{R,i} \cdot S_{R,i}^2 \right) \right) \right) \quad (5)$$

**Figure 2.** Accuracy and precision results using different estimators of f2 and number of units per formulation throughout the range of population f2 values.
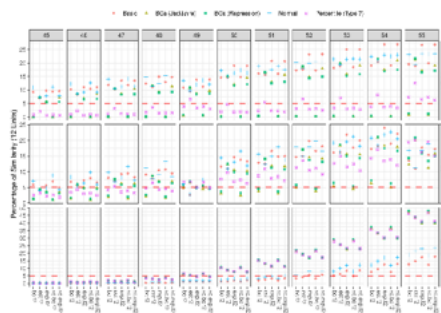


**Figure 3.** Probability of declaring similarity according to the variability conditions and the 90 % confidence interval method.

## 4. Discussion

In general, when f_2 values are high, all estimators under-estimate the population value, which is consistent with the literature findings. When variance is high, a large proportion of $f_{2,bc}$ and $f_{2,vcbc}$ cannot be calculated as the domain of the log function is negative in such cases, which makes them unsuitable to be used for profiles with high variability. When the population $f_2$ values are low, the estimator $f_2$ slightly over-estimates the population value; with the most conservative percentile interval, this estimator still over-estimate the population value at $f_2$ of 47, 48 and 49, with type I error more than 7 %. Estimators $f_{2,exp}$ and $f_{2,vcexp}$ showed better precision than other estimators, but with slightly larger bias when the population $f_2$ values are high. However, from a regulatory perspective, this is less relevant. Regarding the types of intervals, basic and Normal intervals are not suitable as those intervals showed almost 15 % Type I error. BCa intervals with estimator $f_2$, $f_{2,exp}$ and $f_{2,vcexp}$ also led to a type I error higher than 10 %, depending on the sample size and variability. The best combination of estimator and type of confidence interval were $f_{2,exp}$ and $f_{2,vcexp}$ with percentile interval that have type I error around 5 %. The drawback of this combination is its low power; therefore, larger sample sizes may be necessary to have sufficient power.

## 5. Conclusion

We recommend the use of the percentile interval with $f_{2,exp}$ or $f_{2,vcexp}$ to compare dissolution profiles when the conventional $f_2$ similarity factor is not applicable.

## References

1.   U.S. FDA, Guidance for industry. Dissolution testing of immediate release solid oral dosage forms (1997).

2.   European Medicines Agency, Guideline on the investigation of bioequivalence (2010).

3.   B. M. Davit, E. Stier, X. Jiang, O. Anand, Expectations of the US FDA regarding dissolution data in generic drug regulatory submissions, Biopharma Asia 2 (2) (2013).

4.   V. Mangas-Sanjuan, S. Colon-Useche, I. Gonzalez-Alvarez, M. Bermejo, A. García-Arieta, Assessment of the regulatory methods for the comparison of highly variable dissolution profiles., AAPS J. 2016;18(6):1550–61.

5.   P. Paixão, L. F. Gouveia, N. Silva, J. A. G. Morais, Evaluation of dissolution profile similarity — comparison between the f2, the multivariate statistical distance and the f2 bootstrapping methods. Eur J Pharm Biopharm. 2017;112:67–74.

Este trabajo debe ser citado como:

Xu Z, Merino San Juan M, Mangas San Juan V, García Arieta A. Estimators and Confidence Intervals of f_2 Using Bootstrap Methodology for the Comparison of Dissolution Profiles. Rev Esp Cien Farm. 2021;2(2):54-6.