

ARTICLE

Trapezoid bioequivalence: A rational bioavailability evaluation approach on account of the pharmaceutical-driven balance of population average and variability

Sara Soufsaf | Fahima Nekka | Jun Li

Université de Montréal, Montréal,
Québec, Canada**Correspondence**Sara Soufsaf and Jun Li, Faculté de
Pharmacie, Université de Montréal, C.P.
6128, Succ. Centre-ville, Montréal, H3C
3J7, Québec, Canada.Emails: sara.soufsaf@umontreal.ca and
jun.li.2@umontreal.ca**Funding information**Sara Soufsaf reports a research grant
from Fonds de Recherche du Québec–
Santé. Fahima Nekka and Jun Li
report research grants from the Natural
Sciences and Engineering Research
Council of Canada**Abstract**

Among the current approaches for the analysis of bioequivalence, the average bioequivalence (ABE) is limited only to the mean bioavailability, whereas the population bioequivalence (PBE) criterion aggregates both mean and variance in a general comparison formula. However, a rational bioequivalence criterion capable of judging specific drug considerations is always still preferred. As an alternative approach, we introduce an aggregate criterion, namely, the trapezoid bioequivalence (TBE), which includes the consideration of both mean and variance of the bioavailability and adapted weighting of a drug's therapeutic properties. We first applied our method to specific simulated scenarios to compare the strengths and weaknesses of current bioequivalence approaches and demonstrate the improvements brought by TBE. As well, the impact of sample size and variability on ABE, PBE, and TBE are assessed using a population pharmacokinetic model of methylphenidate. Our results indicate that TBE inherits the advantages of both ABE and PBE while greatly reducing their inadequacies. Through simulations with population pharmacokinetic models of specific scenarios, we confirm that (1) TBE does not encounter the overly permissiveness issue of PBE, (2) TBE respects the hierarchy to ABE (TBE \Rightarrow ABE), and (3) TBE assesses bioequivalence with a restriction on $\sigma_T^2 - \sigma_R^2$ without an increase to type 2 errors. The clinically inspired simulations demonstrate TBE's superiority in a realistic context and its potential usefulness in practice. Moreover, the parameter choice in TBE may be adapted according to the specific context of a drug's pharmacological and pharmacodynamic properties.

Study Highlights**WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

Average bioequivalence (ABE) and population bioequivalence (PBE) are the current statistical analyses of bioequivalence. ABE does not consider the variability of bioavailability. PBE is an aggregate criterion that considers variability but poses hierarchy problems with ABE.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

WHAT QUESTION DID THIS STUDY ADDRESS?

Can we propose an aggregate bioequivalence criterion that addresses the flaws of ABE and PBE without adding limitations of its own?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

We propose the trapezoid bioequivalence (TBE) as a criterion that considers the mean and variance of bioavailability with flexible and drug-specific weights. We show that TBE can effectively be applied to compare formulations and respect hierarchy with ABE.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

TBE might be implemented in bioequivalence studies as a flexible approach when a drug's interindividual variability is a limiting factor in prescribability and switchability.

INTRODUCTION

When the patent of an innovative drug expires, a generic can be approved with an abbreviated new drug application, which states that the generic is bioequivalent to the brand name formulation in terms of efficacy and safety.¹ Indeed, only the absorption process might differ, and it must be assessed through the bioavailability.² The bioavailability is measured by the rate and extent of drug absorption, represented by the maximum concentration (C_{\max}) and the area under the curve (AUC) of the plasma concentrations, respectively. Statistical analyses have been proposed to determine the therapeutical equivalence between the test and reference formulations.²

Among the aggregate approaches, the US Food and Drug Administration (FDA) proposes the following three levels of bioequivalence: average bioequivalence (ABE), population bioequivalence (PBE), and individual bioequivalence (IBE).² Because IBE is less used, we focus our work on ABE and PBE.

ABE applies to the averages of the bioavailability metrics on the logarithmic scale for the test and reference formulations (μ_T and μ_R , respectively). ABE states that the test may be a substitute for the reference formulation if the difference between μ_R and μ_T is within 20%. Because ABE is a simple comparison of averages, the declared bioequivalent event can be challenged by largely different variances of two formulations. Consequently, ABE has been questioned for its limited applicability.³⁻⁵ To correct the situation, an additional consideration was proposed by Sheiner³ and Hauck and Anderson,⁴ for example. These authors pointed out that the variability in bioavailability evidently translates to a low precision in predicting the efficacy, which led to the introduction of the PBE criterion.^{2-4,6-8}

PBE considers the drug variability by accounting for the distribution of bioavailability metrics. Compared with ABE, it aggregates the mean and variance (σ_T^2 and σ_R^2) into a one-step comparison by simultaneously considering $\sigma_T^2 - \sigma_R^2$ and $\mu_T - \mu_R$.^{2,3} PBE finds its use when addressing the issue of drug prescribability, which is defined as the substitutability of a test drug to a reference drug for the treatment of naïve patients.⁴

Nonetheless, PBE does not automatically imply ABE, which leads to overly permissive and contradictory results.⁹⁻¹² The non-respect of hierarchy is a fundamental issue when combining two elements into one criterion. In fact, if σ_T^2 is smaller than σ_R^2 , a larger difference between μ_R and μ_T is accepted with PBE,⁹⁻¹¹ thus offsetting the benefit of adding the variance in the evaluation. Hence, a better criterion is needed for a fair trade-off between average and variance to respect the natural hierarchical property.^{8,13} Indeed, several adaptations were proposed in the literature, with some questioning the idea that PBE gives equal importance to μ and σ^2 in the assessment of bioequivalence.^{7,12,14-17} As a solution, Hauck et al.¹³ and Midha¹⁸ proposed to add a weight to σ^2 that can be modified to alter the acceptable threshold of bioequivalence.

The objective of our work is to propose a new bioequivalence criterion, named trapezoid bioequivalence (TBE), which simultaneously takes into account the average (μ) and variance (σ^2) of bioavailability by addressing the flaws of ABE and PBE without adding new limitations. Moreover, the goalpost of the new criterion for establishing bioequivalence should not become more permissive as the within-subject variability of the test drug is reduced, contrarily to PBE whose performance deteriorates in these cases. Finally, we add a trade-off between mean and variance that can be adjusted according to specific drug properties.

As a concrete drug example, we show how we can directly apply our proposed approach to methylphenidate (MPH), the main drug for attention deficit hyperactivity disorder. Because the interindividual variability (IIV) is very large for MPH, the dose individualization is especially difficult.^{19,20} By adopting TBE, we show how we can effectively establish bioequivalence between various formulations while reducing uncertainty related to substitutability. For an objective evaluation of bioequivalence, we have also used enriched clinical trial data for specific scenarios by incorporating population pharmacokinetic (pop-PK) modeling and simulation in our investigation.^{21,22}

METHODS

The detailed ABE and PBE approaches are described in Supplementary Material S1.

Trapezoid bioequivalence

TBE is our proposed strategy to address the role of average and variance in bioequivalence evaluation, and most important the trade-off between both. Contrary to ABE or PBE, which use a single metric, TBE includes a trapezoid zone of acceptance outlined by two distinct sets of inequalities. This zone is expressed as:

$$\text{zone}_{\text{TBE}} = \left\{ (x, y) \left| \begin{array}{l} -\sigma_R^2 \leq x < \alpha_1 \text{ and } 0 \leq y < \theta^{\text{TBE}}; \\ \text{or} \\ \alpha_1 \leq x < \alpha_2 \text{ and } 0 \leq y < -\lambda_1 x + \lambda_2 \end{array} \right. \right\} \quad (1)$$

The explanation for each variable of TBE is given next, whereas the specific values chosen for these variables are detailed in the Scenario-Based Simulations methods section.

For the purpose of bioequivalence, we specifically define $x = \sigma_T^2 - \sigma_R^2$ and $y = (\mu_T - \mu_R)^2$. θ^{TBE} is the maximal squared difference of μ allowed for bioequivalence; α_1 is the therapeutically acceptable difference of σ^2 where TBE can be judged solely based on the difference in μ , and α_2 is the therapeutically unacceptable difference of σ^2 beyond which TBE will directly fail. α_1 and α_2 are used to control the trade-off between μ and σ^2 . Given these parameters, λ_1 and λ_2 are weights which regulate the trade-off between mean and variance and are computed as follows:

$$\lambda_1 = \frac{\theta^{\text{TBE}}}{\alpha_2 - \alpha_1} \quad (2)$$

$$\lambda_2 = \lambda_1 \alpha_2 \quad (3)$$

Specific values for θ^{TBE} , α_1 , α_2 are to be defined by regulatory agencies according to the drug's pharmacological properties and its tolerance for IIV. In this work, we assigned values that would respect general clinical significance and would allow an agreement with ABE and PBE. Each of these values is explained in the Scenario-Based Simulations section in the Methods.

To facilitate the hypothesis test, Equation (1) can be transformed as:

$$\begin{aligned} & \bullet (\mu_T - \mu_R)^2 - \theta^{\text{TBE}} < 0, \text{ when } -\sigma_R^2 \leq \sigma_T^2 - \sigma_R^2 < \alpha_1 \\ & \bullet (\mu_T - \mu_R)^2 + \lambda_1 (\sigma_T^2 - \sigma_R^2) - \lambda_2 < 0, \text{ when } \alpha_1 \leq \sigma_T^2 - \sigma_R^2 < \alpha_2 \end{aligned}$$

TBE can be dynamically accessed in <https://mphss.shinyapps.io/SoufsafTrapezoidBE/>.

TBE conclusions are drawn based on the upper 90% confidence interval (CI) of the previous inequalities. The bootstrap procedure was used to compute the CI, and 2000 replicates were used for the bootstrap.² As in PBE, TBE is declared if this value falls below 0. Otherwise, TBE will not be concluded. The TBE acceptance zone (zone_{TBE}) is illustrated in Figure 1a, and a flowchart of TBE computation and decision is presented in Figure 1b.

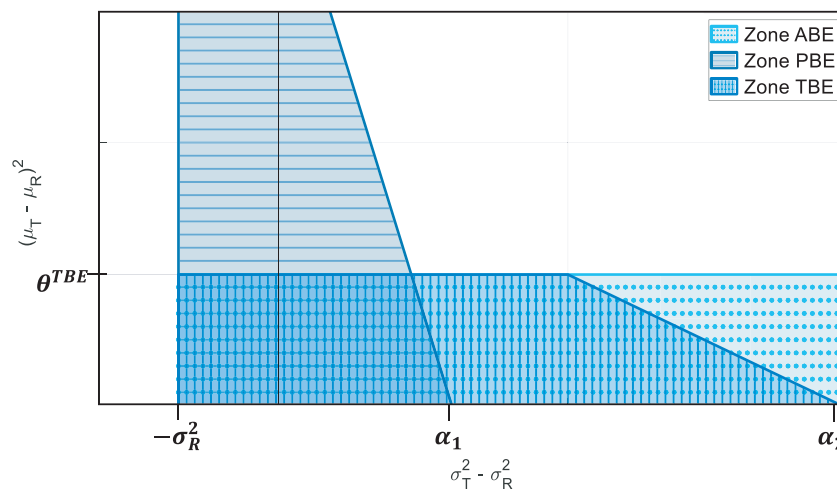
Similar to PBE and in accordance with Equation (1), zone_{TBE} is defined by limits in terms of $(\mu_T - \mu_R)^2$ and $\sigma_T^2 - \sigma_R^2$. When $\sigma_T^2 - \sigma_R^2 > 0$, zone_{TBE} corrects the drawback of ABE and adds a consideration of variability to bioequivalence. As well, when $\sigma_T^2 - \sigma_R^2 < 0$, zone_{TBE} corrects the drawback of PBE and imposes a limit to $(\mu_T - \mu_R)^2$.

As defined in Equation (1), TBE possesses multiple favorable properties. First, its limits are defined with a clinical significance for each parameter (θ^{TBE} and α_1, α_2). Indeed, θ^{TBE} may be set in accordance with current ABE criteria, and α_1, α_2 may be set in accordance with clinically acceptable limits of variability. Second, it can be shown that TBE is reduced to ABE assuming $\sigma_T^2 = \sigma_R^2$ and fixing $\theta^{\text{TBE}} = \ln^2(1.25)$. Third, TBE respects the hierarchy with ABE. Indeed, TBE introduces θ^{TBE} to prevent widening the acceptable limits of $(\mu_T - \mu_R)^2$ when σ_T^2 is reduced with respect to σ_R^2 . Finally, TBE allows a flexible trade-off between mean and variance with the weights λ_1 and λ_2 . Indeed, TBE permits a control on the weight and importance given to $(\mu_T - \mu_R)^2$ and $\sigma_T^2 - \sigma_R^2$. These favorable properties are examined in this work.

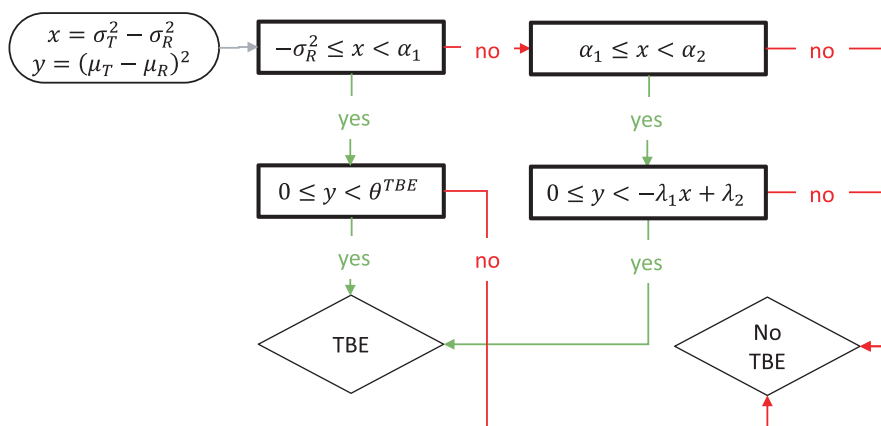
ABE, PBE and TBE were applied in three ways. First, we computed statistical methods of bioequivalence to broad scenario-based simulations applicable to all drugs. Second, we used MPH as a specific drug to exemplify our work. Third, we computed the type 1 and 2 errors.

FIGURE 1 (a) Zones of acceptance of bioequivalence for average bioequivalence (ABE), population bioequivalence (PBE), and trapezoid bioequivalence (TBE) as shaded areas. μ_T and μ_R are the averages of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively; σ_T^2 and σ_R^2 are the variances of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively; θ^{TBE} is the maximal squared difference of μ allowed for bioequivalence; α_1 is the therapeutically acceptable difference of σ^2 ; α_2 is the therapeutically unacceptable difference of σ^2 ; λ_1 and λ_2 are weights applied to control the trade-off between μ and σ^2 . (b) Flowchart of bioequivalence decisions with TBE

(a) Bioequivalence zones



(b) Flowchart of trapezoid bioequivalence



Scenario-based simulations

The scenarios chosen for simulation are combinations of the following situations: a relatively small variability for the test formulation ($\sigma_T^2 - \sigma_R^2 < 0$), a large mean difference between the test and reference formulations ($(\mu_T - \mu_R)^2 > \theta^{TBE}$), a therapeutically tolerable difference of variability ($\alpha_1 \leq \sigma_T^2 - \sigma_R^2 < \alpha_2$), and nonsubstitutable test and reference formulations ($\sigma_T^2 - \sigma_R^2 > \alpha_2$). The first and second situations identify failures of PBE, whereas the third and fourth situations identify failures of ABE.

For each patient, the bioavailability metric values (AUC or C_{max}) of test (or reference) formulations are drawn from normal distributions with means μ_T (or μ_R) and variances σ_T^2 (or σ_R^2). The fixed values of μ_R , μ_T , σ_R^2 , and σ_T^2 are reported in Table 1.

For each scenario, 40 subjects are used. Thus, 80 AUC measurements (two per subject) are simulated for a cross-over and nonreplicated clinical trial. ABE, PBE, and TBE results are then computed for each scenario as described

in the previous section. For the sake of respecting the desirable properties of bioequivalence criterion, we fixed TBE parameters in accordance with ABE’s and PBE’s FDA goalposts.

θ^{TBE} : by fixing $\theta^{TBE} = \ln^2(1.25)$, we respect ABE’s acceptable $\pm 20\%$ mean difference on the log scale (or 80%–125% on the original scale).

α_1 : we fix $\pm 30\%$ as the tolerable difference of σ^2 on the log scale as suggested.²³ Clinically, this is a range of variance between 70%–143% on the original scale.

Using Equations (2) and (3), we have $\lambda_1 = 0.1480$ and $\lambda_2 = 0.1026$.

We repeated each sampling scenario 100 times.

MPH model-based simulations

In addition to the scenario-based simulations, we applied the described bioequivalence methods to the specific context of MPH. Indeed, as a drug with a higher IIV, it

TABLE 1 Results for the scenario-based simulations

Scenario	ABE			PBE			TBE		
	$(\mu_T - \mu_R)^2$	$\sigma_T^2 - \sigma_R^2$	σ_R^2	90% CI	Probability of passing bioequivalence (%)	Mean upper 90% CI (CV %)	Probability of passing bioequivalence (%)	Mean upper 90% CI (CV %)	Probability of passing bioequivalence (%)
1	0.0111	0.3	0.0225	0.9529–1.2977	1	0.4070 (12.86)	0	–0.0271 (–18.67)	100
2	0.0900	0.3	0.0225	1.1572–1.5751	0	0.5069 (9.86)	0	0.1317 (29.40)	0
3	0.0225	0.8	0.0225	0.9174–1.4834	0	1.1530 (11.68)	0	0.0953 (11.34)	0
4	0.0900	–0.01	0.0225	1.2894–1.4217	0	0.1918 (12.17)	0	0.1347 (22.30)	0
5	0.0111	–0.01	0.0225	1.0574–1.1651	100	0.0142 (74.343)	10	–0.0278 (–17.69)	100
6	0.0111	0.3	0.1225	0.9188–1.3461	0	0.3605 (15.42)	0	–0.0270 (–18.803)	100
7	0.0900	0.3	0.1225	1.1146–1.6374	0	0.4330 (15.28)	0	0.1327 (33.158)	0
8	0.0225	0.8	0.1225	0.8885–1.5235	0	1.1009 (14.67)	0	0.0948 (15.63)	0
9	0.0900	–0.1	0.1225	1.2287–1.5012	0	–0.0476 (–80.4)	85	0.1388 (26.29)	0
10	0.0111	–0.1	0.1225	1.0059–1.2293	86	–0.1827 (–12.45)	100	–0.0271 (–18.67)	100

Note: Values in bold signify that the approach passes bioequivalence. 90% CI indicates mean 90% CI across all replications.

Abbreviations: μ_T and μ_R , averages of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively; σ_T^2 and σ_R^2 , variances of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively; ABE, average bioequivalence; CI, confidence interval; CV, coefficient of variation across all replications; PBE, population bioequivalence; TBE, trapezoid bioequivalence.

exemplifies the added value of using TBE instead of ABE or PBE.

The bioequivalence methods were applied to two types of data. First, it was applied to the analysis of a randomized clinical trial (Supplementary Material S2). Second, as the available clinical trial data were limited, we used model-based simulations to explore additional considerations pertinent to bioequivalence: IIV and sample size.

Interindividual variability

To explore the impact of IIV on the bioequivalence methods, we used the published MPH pop-PK model to simulate databases that incorporate interindividual and intraindividual variability.¹⁹ Each simulated pair of test and reference formulations is chosen with a random variance of IIV (ω^2) listed in Table 2 while the fixed effects and residual variability remained unchanged. Consequently, the σ_R^2 and σ_T^2 evaluated by PBE and TBE still depend on

interindividual and intraindividual variability. However, as total IIV is the only difference between the reference and test databases, only the impact of IIV on ABE, PBE, and TBE is evaluated. The magnitude of IIV on each parameter was chosen according to a reasonable scale observed in pop-PK models. In staying true to the original pop-PK model, we did not explore any IIV on lag and fixed it for all simulations.

Sample size

Subsequently, we explored the impact of sample size on ABE, PBE, and TBE. The same methods as noted previously were applied with 40 and 100 subjects in each simulation. These numbers were chosen to investigate a realistic range observed in clinical trials.

TBE parameters used in the MPH model-based simulations were chosen exactly as in the Scenario-Based Simulations section in the Methods.

TABLE 2 Results for MPH model-based simulations

Number of patients in clinical study	Sum of IIV ^a for the reference formulation	Sum of IIV ^a for the test formulation	Results of bioequivalence		
			ABE	PBE	TBE
40		IIV _T = IIV _R			
	0.1	0.1	YES	YES	YES
	1	1	YES	YES	YES
	1.5	1.5	YES	YES	YES
			IIV _T > IIV _R		
	0.1	1	YES	NO	YES
	0.1	1.5	<u>NO</u>	NO	YES
			IIV _T < IIV _R		
	1	0.1	YES	YES	YES
1.5	0.1	<u>NO</u>	YES	YES	
100		IIV _T = IIV _R			
	0.1	0.1	YES	YES	YES
	1	1	YES	YES	YES
	1.5	1.5	YES	YES	YES
			IIV _T > IIV _R		
	0.1	1	YES	NO	YES
	0.1	1.5	<u>YES</u>	NO	YES
			IIV _T < IIV _R		
	1	0.1	YES	YES	YES
1.5	0.1	<u>YES</u>	YES	YES	

Note: Bold signifies that the approach passes bioequivalence. Italics and underline signify a result that changes according to the number of patients in the clinical study.

Abbreviations: ABE, average bioequivalence; IIV_T and IIV_R, interindividual variability for the test and reference formulations, respectively; MPH, methylphenidate; PBE, population bioequivalence; TBE, trapezoid bioequivalence.

^aIIV expressed as the sum of variance (ω^2) on ka_1 (first absorption constant); ka_2 (first absorption constant); F_1 (immediate release fraction of MPH), where ω^2 is the variance of the normally distributed IIV $\eta \sim N(0, \omega^2)$ and $\omega_{ka1}^2 \in \{0.05, 0.15, 0.35, 0.5\}$; $\omega_{ka2}^2 \in \{0.05, 0.15, 0.35, 0.5\}$; $\omega_{F1}^2 \in \{0.001, 0.3, 0.5, 0.7\}$; $\omega_{lag}^2 = 0$.

Type 1 and type 2 errors

The type 1 and type 2 errors of ABE, PBE, and TBE were evaluated through simulations of 1000 trials with a crossover and nonreplicated design. To evaluate the impact of sample size, the type 1 and type 2 errors were computed separately with sample sizes of 10, 20, 40, 60, 80, and 100.

First, we computed the type 1 error for simulated trials that follow the null hypothesis of bioinequivalence. Specifically, we chose to simulate pharmacokinetic measures from distributions where $(\mu_T - \mu_R)^2 = 0.0498$. In keeping with the scenario-based simulations and the range observed in bioequivalence studies in Nakai et al.,⁹ the value of σ_R^2 was fixed to 0.0225. The type 1 error was computed as the proportion of the simulated trials that reject the bioequivalence.

Second, we computed the type 2 error for simulated trials that follow the alternative hypothesis of bioequivalence where $(\mu_T - \mu_R)^2 = 0$, $\sigma_T^2 = \sigma_R^2 = 0.0225$. The type 2 error was computed as the proportion of the simulated trials that accept the bioequivalence. In addition, we computed power curves by simulating trials with varying levels of $(\mu_T - \mu_R)^2 \in \{0, 0.01, 0.04, 0.0498, 0.09\}$ with $\sigma_T^2 = \sigma_R^2 = 0.0225$.

RESULTS

Scenario-based simulations

For the three bioequivalence methods, five specific scenarios are chosen and tested as reported in Table 1. Each scenario is applied to a small variability ($\sigma_R^2 = 0.0225$) and a large variability ($\sigma_R^2 = 0.1225$). Thus, we have 10 scenarios. To have more reliable conclusions, all scenarios were repeated 100 times. The results are also illustrated in Figure 2.

In Table 1, the results of ABE, PBE, and TBE are presented pertaining to the CI and the percentage of all replications that conclude bioequivalence. Specifically, we report the average 90% CI across all replications for ABE analysis, and the average upper limit of the linearized 90% CI across all replications for PBE and TBE analyses. As well, to account for the results of each replication, we report the probability of passing bioequivalence as the percentage of all replications that conclude bioequivalence.

The results are summarized as follows, and the details are displayed in Figure 2:

- As expected, the scenarios where the clinically acceptable limits of $(\mu_T - \mu_R)^2$ and $\sigma_T^2 - \sigma_R^2$ are exceeded never conclude bioequivalence (Scenarios 2 and 7).

- The scenarios where solely the acceptable limits of $\sigma_T^2 - \sigma_R^2$ are exceeded show cases where ABE has been criticized (Scenarios 3 and 8). Indeed, ABE only considers the mean μ and does not take into account the variability σ^2 . By implementing α_1 in Equation (1), TBE corrects these situations. However, although these scenarios are found in $zone_{ABE}$, the probability of concluding bioequivalence with ABE is null. This contradiction stems from its 90% CI's sensitivity to sample size (Equation S2). When the sample size is large enough, 90% CI tightens and ABE can conclude bioequivalence. A larger sample size and the results are discussed further in the next section.
- A scenario where the limits of ABE, PBE, and TBE are all respected concludes to bioequivalence for all methods (Scenario 10) with $\sigma_R^2 = 0.1225$. With $\sigma_R^2 = 0.0225$, we can identify the first drawback of PBE. Indeed, due to its reference-scaled Equation (S4), PBE's permissiveness depends on σ_R^2 .⁹ Thus, the criterion PBE is expected to be stricter when σ_R^2 is decreased. In fact, PBE's probability of concluding bioequivalence drops from 100% to 6% (Scenario 10 vs. 5). TBE corrects this drawback as its probability of concluding bioequivalence remains at 100% for either scenario.
- For scenarios where the clinically acceptable limits of $(\mu_T - \mu_R)^2$ and $\sigma_T^2 - \sigma_R^2$ are respected (Scenarios 1 and 6), we highlight cases where TBE accepts bioequivalence while ABE and PBE do not. Indeed, these scenarios do not fall inside $zone_{PBE}$, and the null probability of passing bioequivalence with PBE is expected. In addition, similar to Scenarios 3 and 8, ABE does not pass bioequivalence due to a small sample size. Nonetheless, TBE still successfully concludes bioequivalence.
- For scenarios where solely the acceptable limits of $(\mu_T - \mu_R)^2$ are exceeded, we show the second drawback of PBE. In fact, these are cases where PBE has been criticized and deemed too permissive compared with ABE.⁹⁻¹¹ It is clearly shown that, with $\sigma_R^2 = 0.1225$ and $(\mu_T - \mu_R)^2 > \theta^{ABE}$ (Scenario 9), PBE is the only approach that concludes bioequivalence. Conversely, if σ_R^2 is reduced to $\sigma_R^2 = 0.0225$ (Scenario 4), PBE no longer concludes bioequivalence. TBE corrects this contradiction through θ^{TBE} and its probability of concluding bioequivalence is the same as ABE's regardless of σ_R^2 .

MPH model-based simulations

To complement the results obtained from the clinical trial data (Supplementary Material S2), we used the model-based simulations to examine various levels of

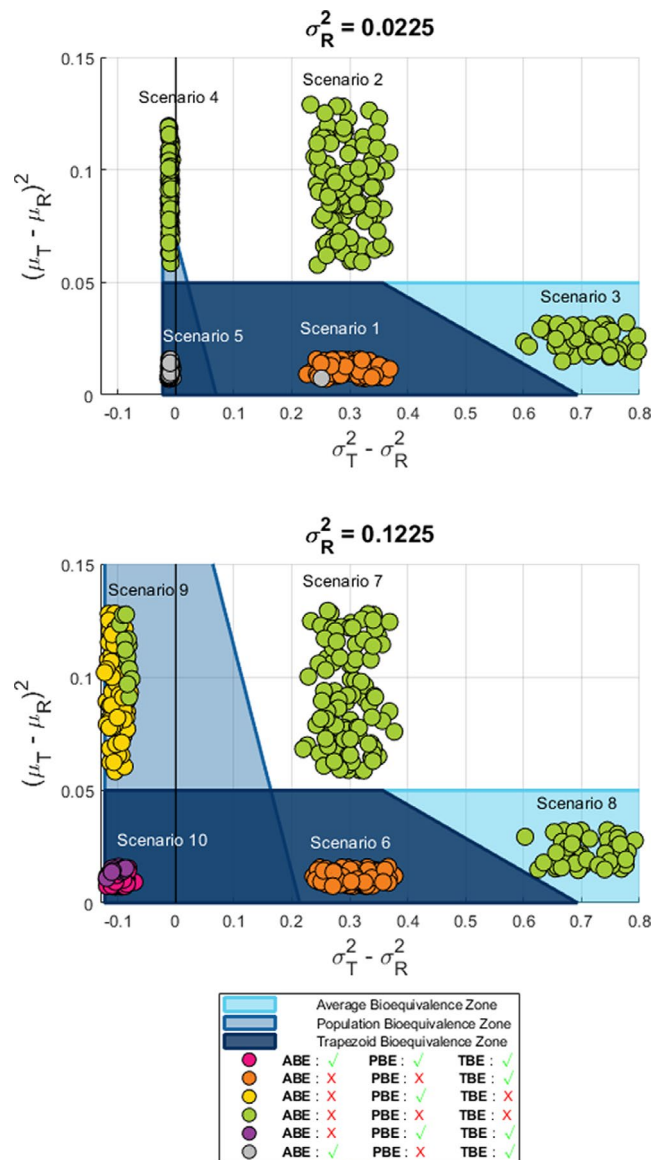


FIGURE 2 The conclusion of bioequivalence for average bioequivalence (ABE), population bioequivalence (PBE), and trapezoid bioequivalence (TBE) are represented as a scatter, and the bioequivalence zones are illustrated as shaded areas. Each cluster is identified with a text box referring to the scenario number in Table 1. μ_T and μ_R are the averages of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively; σ_T^2 and σ_R^2 are the variances of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively. ✓, the approach passes bioequivalence; X, the approach fails to demonstrate bioequivalence. Top, scenarios for $\sigma_R^2 = 0.0225$; bottom, scenarios for $\sigma_R^2 = 0.1225$

IIV, which were not observed in the MPH clinical trial data. The IIV was modified for three pharmacokinetic parameters: the first and second absorption (ka_1 and ka_2 , respectively) and the release of the external MPH fraction (F_1).

TABLE 3 Type 1 and Type 2 errors

Sample size	ABE	PBE	TBE
Type 1 error ^a (%)			
10	5.4	2.3	15.6
20	4.3	0.3	7.1
40	4.8	0	2.3
60	5.4	0	0.5
80	4	0	0.1
100	4	0	0
Type 2 error ^b (%)			
10	15.2	55.1	1.2
20	0	23.1	0
40	0	7.3	0
60	0	2.2	0
80	0	0.7	0
100	0	0.6	0

Abbreviations: μ_T and μ_R , averages of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively; σ_T^2 and σ_R^2 , variances of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively; ABE, average bioequivalence; PBE, population bioequivalence; TBE, trapezoid bioequivalence.

^aThe type 1 error was computed from simulations with $(\mu_T - \mu_R)^2 = 0.0498$ and $\sigma_T^2 = \sigma_R^2 = 0.0225$.

^bThe type 2 error was computed from simulations with $(\mu_T - \mu_R)^2 = 0$ and $\sigma_T^2 = \sigma_R^2 = 0.0225$.

Because the typical values for all parameters did not change, the mean pharmacokinetic profile was the same for all MPH model-based simulations. Thus, this section demonstrates differences between ABE, PBE, and TBE solely when the IIV is involved. Two sample sizes were tested to represent realistic numbers of patients enrolled in the MPH clinical trial and general bioequivalence studies.¹¹

The total IIV for ka_1 , ka_2 , and F_1 and the bioequivalence results for ABE, PBE, and TBE are presented in Table 2. As expected, ABE, PBE, and TBE always conclude to bioequivalence when the IIV is unchanged between the test and reference formulations ($IIV_T = IIV_R$) regardless of sample size.

When $IIV_T > IIV_R$, PBE does not conclude bioequivalence in either of the two examples given in Table 2 ($IIV_T = 1$; $IIV_R = 1.5$). In fact, this situation precisely represents the restrictiveness of PBE and its lack of drug-specific flexibility. On the other hand, ABE passes bioequivalence only if $IIV_T = 1$, which can be explained with ABE's sensitivity to sample size (this property is mentioned in the Scenario-Based Simulations section in the Results). In fact, when the sample size is increased to 100 and $IIV_T =$

FIGURE 3 Power curve for $(\mu_T - \mu_R)^2 = \{0, 0.01, 0.04, 0.0498, 0.09\}$ applied to average bioequivalence (ABE), population bioequivalence (PBE), and trapezoid bioequivalence (TBE). The power of ABE, PBE, and TBE were evaluated through the simulations of 1000 trials with a crossover and nonreplicated design. Each simulation was applied to sample sizes of 10, 20, 40, 60, 80, and 100. μ_T and μ_R are the averages of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively

1, ABE passes bioequivalence. By contrast, TBE concludes to bioequivalence for both examples and both sample sizes because the IIV values respect the chosen TBE parameter used for MPH: $\theta^{\text{TBE}} = \ln^2(1.25)$, $\alpha_1 = \ln 1.4286$, $\alpha_2 = \ln 2$, $\lambda_1 = 0.1480$, and $\lambda_2 = 0.1026$. Nonetheless, these parameter values may be changed by regulatory agencies to restrict the tolerated IIV.

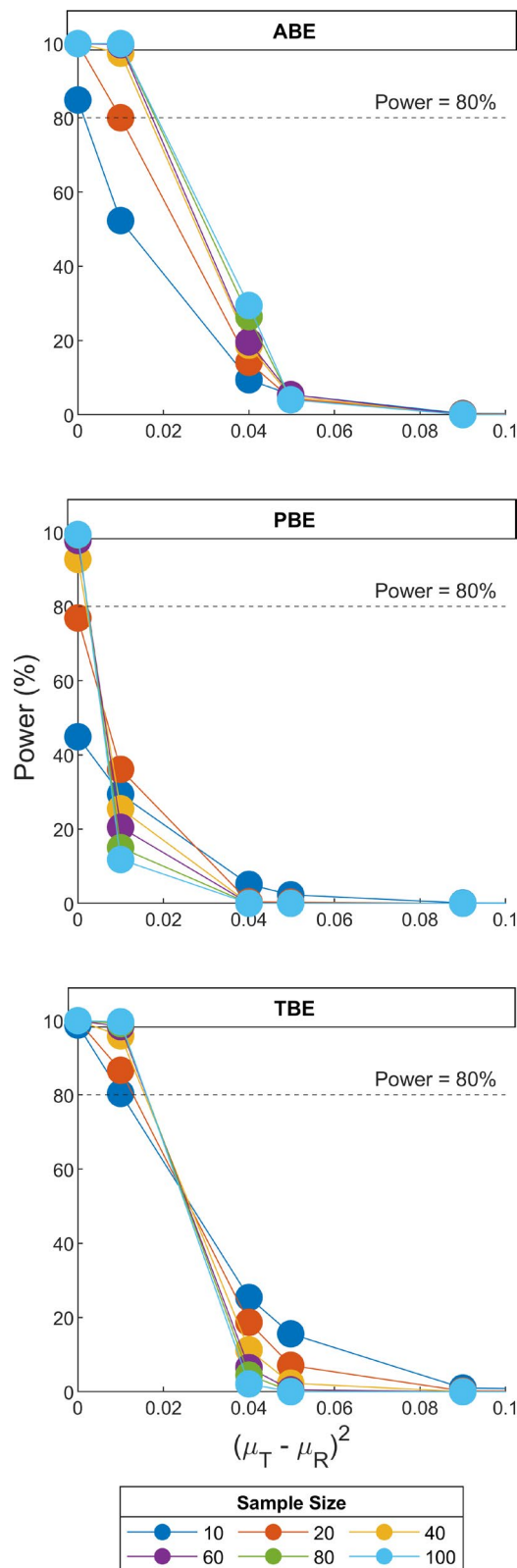
Finally, when $\text{IIV}_T < \text{IIV}_R$, ABE draws once again different conclusions depending on IIV_T and sample size. If $\text{IIV}_T = 1$, ABE concludes to bioequivalence. Contrarily, if $\text{IIV}_T = 1.5$, ABE does not conclude to bioequivalence for a small sample size. However, when the sample size is increased to 100, ABE can conclude to bioequivalence. PBE and TBE always conclude to bioequivalence regardless of sample size.

Type 1 and type 2 errors

Table 3 provides the type 1 and type 2 errors for all tested sample sizes. When the sample size is 10, the type 1 error exceeds 5% for ABE and TBE. It is 5.4%, 2.3%, and 15.6% for ABE, PBE, and TBE, respectively. We also note that the type 1 error with TBE is greater when the sample size is small (10 and 20 patients). For all other sample sizes, the type 1 error is below the acceptable 5% threshold and shows a satisfactory level of rejection of the null hypothesis. The type 1 error was also computed for cases where $\sigma_T^2 - \sigma_R^2 \in \{0.3, 0.7\}$ and was evaluated at 0% for all tested sample sizes (results not shown).

The type 2 error presented in Table 3 shows that the power of TBE is greater than that of ABE and PBE for all sample sizes. Specifically, we note that the type 2 error of TBE is very low for all sample sizes analyzed. Notably, the type 2 error with TBE is 1.2% for a sample size of 10, whereas it is 15.2% with ABE and 55.1% with PBE. When the sample sizes are greater than 10, the type 2 error is null for ABE and TBE, whereas those for PBE are higher.

Figure 3 provides the power curves of ABE, PBE, and TBE, explicitly the probability of concluding bioequivalence for different values of $(\mu_T - \mu_R)^2$ and samples. Among all simulations and sample sizes, TBE's power was



higher or similar to ABE's and PBE's power. Specifically, the minimal sample sizes that allow a power larger than 80% are 20, 40, and 10, respectively, for ABE, PBE, and TBE. We note that the power for larger sample sizes

decreased as $(\mu_T - \mu_R)^2$ approaches its maximum threshold. In other words, larger sample sizes allow a more precise and accurate assessment of true BIE.

DISCUSSION

Several bioequivalence methods have been proposed to solve the issues of ABE and PBE.^{13,18,24} As an alternative bioequivalence method, we propose the aggregate bioequivalence criterion TBE that encompasses ABE and PBE. TBE can balance the similarity in both formulations' average (μ) and variance (σ^2) of bioavailability at the same time. Moreover, this balance can be adapted to specific pharmacological characteristics through weights applied in TBE. For example, the weight of the difference of variance may be chosen larger for drugs with large IIV to ensure the best substitutability for a patient.

Table 4 summarizes TBE's appealing properties as observed from the scenario-based simulations and the MPH case studies. For the former, each scenario was chosen to demonstrate TBE's properties in various clinical settings and to compare them to ABE and PBE. We analyzed small and large σ_R^2 , large mean difference between the test and reference formulation, substitutable and nonsubstitutable formulations according to a therapeutically defined tolerance of variability. These scenarios allowed us to identify inadequacies of ABE and PBE and to confirm TBE's strengths. Namely, we first showed that an increase in σ_R^2 did not change TBE's acceptable limits of bioequivalence. Indeed, TBE's probability of passing bioequivalence was identical whether $\sigma_R^2 = 0.0225$ or $\sigma_R^2 = 0.1225$. This was not

the case for PBE, where comparable levels of $(\mu_T - \mu_R)^2$ and $(\sigma_T^2 - \sigma_R^2)$ resulted in opposite conclusions of bioequivalence depending on σ_R^2 . In addition, we showed that TBE is not overly permissive compared with ABE when $\sigma_T^2 - \sigma_R^2 < 0$, contrary to PBE. On the other hand, when $\sigma_T^2 - \sigma_R^2 > 0$, we define TBE to differ from ABE by imposing a clinically defined threshold on $\sigma_T^2 - \sigma_R^2$ and proportionally reducing the threshold on $(\mu_T - \mu_R)^2$. This property was only observed when the sample size was high enough for ABE to lead to bioequivalence (results not shown with scenario-based simulations).

We explored additional conditions of variability and sample size through a pop-PK model of extended release MPH.¹⁹ By varying the IIV on absorption parameters and testing a larger sample size, we simulated $IIV_T = IIV_R$, $IIV_T > IIV_R$, and $IIV_T < IIV_R$ when the sample size is 40 or 100 subjects. All results concurred with the scenario-based method. We observed again PBE's increased permissiveness as we reduce IIV_T . Furthermore, as the sample size increased, we observed an agreement between the results from ABE and TBE. These results complemented the scenario-based simulations and confirmed that TBE's permissiveness was only apparent and dependent on the study sample size.

An interesting feature of TBE was demonstrated in the type 2 error computations. Indeed, higher or similar power may be achieved with TBE across all sample sizes compared with ABE and PBE. Thus, although ABE supposes that there is no difference in variance between both drugs, TBE imposes an explicit restriction to $\sigma_T^2 - \sigma_R^2$ without additional cost to the type 2 errors. We note that the use of 90% CI reduces the fixed maximum threshold of $(\mu_T - \mu_R)^2$. Namely, we find that only simulations where the sample $(\mu_T - \mu_R)^2 < 0.0365$, $(\mu_T - \mu_R)^2 < 0.0055$ or $(\mu_T - \mu_R)^2 < 0.025$ were declared bioequivalent with ABE, PBE, and TBE, respectively. Because the $(\mu_T - \mu_R)^2$ limit of TBE is further removed from its theoretical value than in the case of ABE, future work should improve TBE's computation of 90% CI and replace the current use of the bootstrap.

In addition, the type 1 error is dependent on sample size for TBE, which is contrary to statistical definitions of the type 1 error. We hypothesize that it is due to the nonparametric bootstrap in TBE's CI, which transfers its dependance on sample size to the type 1 error.²⁵⁻²⁷ The authors chose a nonparametric bootstrap to compute the TBE's CI as it does not rely on any assumptions and it is the most accessible.

We acknowledge that the application of TBE is limited to cases where the variability has a significant impact on treatment. The example we chose in our work is MPH, whose dose individualization is challenging because of

TABLE 4 Desirable properties of ABE, PBE, and TBE

Properties	ABE	PBE	TBE
Sensitive to μ and σ^2	X	√	√
Interpreted on the normal scale	√	X	X
Stable results with different n	X	√	√
Stable results with different σ_R^2	X ^a	X ^b	√
Stable results when $\sigma_T^2 < \sigma_R^2$	√	X	√

Note:√ Signifies that the property applies to the bioequivalence method and X signifies that the property does not apply to the bioequivalence method
Abbreviations: ABE, average bioequivalence; PBE, population bioequivalence; TBE, trapezoid bioequivalence; μ , average of the bioavailability metrics on the logarithmic scale; σ , variance of the bioavailability metrics on the logarithmic scale; σ_T^2 and σ_R^2 , the variances of the bioavailability metrics on the logarithmic scale for the test and reference formulations, respectively.

^aIf σ_R^2 is large, bioequivalence is less permissive.

^bIf σ_R^2 is large, bioequivalence is more permissive.

its large IIV. As well, it is known that MPH's therapeutic effect closely follows its pharmacokinetics. Thus, it is our hypothesis that controlling for variability between test and reference formulations will reduce the titration period by increasing drug substitutability.

In conclusion, the clinically inspired simulations showed TBE's superiority in a reasonable context and its potential usefulness in practice. Indeed, TBE is mathematically accessible, and its statistical analysis is not more complex than PBE. As well, a standard 2×2 crossover design is sufficient to estimate TBE. Furthermore, TBE's parameters (θ^{TBE} and α_1, α_2) permit a highly flexible approach. Although these parameters were specifically fixed to values justified for MPH in this work, they can be modified to reflect any drug's pharmacological and pharmacodynamic characteristics. For example, stricter limits can be established by regulatory agencies for narrow therapeutic drugs that require close titration (insulin, blood thinners, anticonvulsants, etc.). Further work on TBE should involve work on the calculation of the CI, and estimation of optimal sample size must be applied to complete work on TBE.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

S.S., J.L., and F.N. wrote the manuscript. S.S., J.L., and F.N. designed the research. S.S. performed the research and analyzed the data.

REFERENCES

1. US Food and Drug Administration. *Abbreviated New Drug Application (ANDA)*. <https://www.fda.gov/drugs/types-applications/abbreviated-new-drug-application-anda>. Published 2019. Accessed May 21, 2021.
2. US Food and Drug Administration. *Guidance for industry. Statistical approaches to establishing bioequivalence*. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-approaches-establishing-bioequivalence>. Published 2001. Accessed October 22, 2020.
3. Sheiner LB. Bioequivalence revisited. *Stat Med*. 1992;11(13):1777-1788.
4. Hauck WW, Anderson S. Measuring switchability and prescribability: when is average bioequivalence sufficient? *J Pharmacokinet Biopharm*. 1994;22(6):551-564.
5. Steinijans VW. Some conceptual issues in the evaluation of average, population, and individual bioequivalence. *Therapeutic Innovat Regulat Sci*. 2001;35(3):893-899.
6. Hyslop T, Hsuan F, Holder DJ. A small sample confidence interval approach to assess individual bioequivalence. *Stat Med*. 2000;19(13):2885-2897.
7. Schall R. Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. *Biometrics*. 1995;51(2):615-626.
8. Endrenyi L, Hao Y. Asymmetry of the mean-variability trade-off raises questions about the model in investigations of individual bioequivalence. *Int J Clin Pharmacol Ther*. 1998;36(8):450-457.
9. Nakai K, Fujita M, Tomita M. Comparison of average and population bioequivalence approach. *Int J Clin Pharmacol Ther*. 2002;40(9):431-438.
10. Barrett JS, Batra V, Chow A, et al. PhRMA perspective on population and individual bioequivalence. *J Clin Pharmacol*. 2000;40(6):561-570.
11. Zariifa NM, Patterson SD. Population and individual bioequivalence: lessons from real data and simulation studies. *J Clin Pharmacol*. 2001;41(8):811-822.
12. Dragalin V, Fedorov V, Patterson S, Jones B. Kullback-Leibler divergence for evaluating bioequivalence. *Stat Med*. 2003;22(6):913-930.
13. Hauck WW, Chen ML, Hyslop T, Patnaik R, Schuirmann D, Williams R. Mean difference vs. variability reduction: trade-offs in aggregate measures for individual bioequivalence. FDA Individual Bioequivalence Working Group. *Int J Clin Pharmacol Ther*. 1996;34(12):535-541.
14. Karalis V, Symillides M, Macheras P. Novel methods to assess bioequivalence. *Expert Opin Drug Metab Toxicol*. 2011;7(1):79-88.
15. Pereira LM. Bioequivalence testing by statistical shape analysis. *J Pharmacokinet Pharmacodyn*. 2007;34(4):451-484.
16. Polli JE, McLean AM. Novel direct curve comparison metrics for bioequivalence. *Pharm Res*. 2001;18(6):734-741.
17. Marston SA, Polli JE. Evaluation of direct curve comparison metrics applied to pharmacokinetic profiles and relative bioavailability and bioequivalence. *Pharm Res*. 1997;14(10):1363-1369.
18. Midha KK, Rawson MJ, Hubbard JW. Individual and average bioequivalence of highly variable drugs and drug products. *J Pharm Sci*. 1997;86(11):1193-1197.
19. Soufsaf S, Robaey P, Bonnefois G, Nekka F, Li J. A quantitative comparison approach for methylphenidate drug regimens in attention-deficit/hyperactivity disorder treatment. *J Child Adolesc Psychopharmacol*. 2019;29(3):220-234.
20. Ermer JC, Adeyi BA, Pucci ML. Pharmacokinetic variability of long-acting stimulants in the treatment of children and adults with attention-deficit hyperactivity disorder. *CNS Drugs*. 2010;24(12):1009-1025.
21. Yue CS, Ozdin D, Selber-Hnatiw S, Ducharme MP. Opportunities and challenges related to the implementation of model - based bioequivalence criteria. *Clin Pharm*. 2019;105(2):13.
22. Dubois A, Gsteiger S, Pigeolet E, Mentré F. Bioequivalence Tests Based on Individual Estimates Using Non-compartmental or Model-Based Analyses: Evaluation of Estimates of Sample Means and Type I Error for Different Designs; 13.
23. Mould DR, Upton RN. Basic concepts in population modelling, simulation, and model-based drug development-part 2: introduction to pharmacokinetic modelling methods. *CPT Pharmacometrics Syst Pharmacol*. 2013;17(2):e38.
24. Wellek S. On a reasonable disaggregate criterion of population bioequivalence admitting of resampling-free testing procedures. *Stat Med*. 2000;19(20):2755-2767.
25. Chernick MR. *Bootstrap Methods: A Guide for Practitioners and Researchers*. 2nd ed. Wiley-Interscience; 2008.

26. Hesterberg TC. What teachers should know about the bootstrap: resampling in the undergraduate statistics curriculum. *Am Stat.* 2015;69(4):371-386.
27. Bradley E, Tibshirani RJ. *An Introduction to the Bootstrap*. Chapman and Hall; 1993. Monographs on Statistics and Applied Probability; vol 57.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Soufsaf S, Nekka F, Li J. Trapezoid bioequivalence: A rational bioavailability evaluation approach on account of the pharmaceutical-driven balance of population average and variability. *CPT Pharmacometrics Syst Pharmacol.* 2022;11:482-493. doi:[10.1002/psp4.12775](https://doi.org/10.1002/psp4.12775)