Research Article

A Bioequivalence Test by the Direct Comparison of Concentration-versus-Time Curves Using Local Polynomial Smoothers

Suyan Tian,[1,2,3] Howard H. Chang,[4] Dana Orange,[5,6] Jingkai Gu,[7,8] and Mayte Suárez-Fariñas[2,9,10]

[1] *Division of Clinical Research, First Hospital of Jilin University, 71 Xinmin Street, Changchun, Jilin 130021, China*

[2] *Center for Clinical and Translational Science, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA*

[3] *School of Mathematics, Jilin University, 2699 Qianjin Street, Changchun, Jilin 130012, China*

[4] *Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, USA*

[5] *Laboratory of Molecular Neurooncology, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA*

[6] *Division of Rheumatology, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021, USA*

[7] *Clinical Pharmacology Center, Research Institute of Translational Medicine, First Hospital of Jilin University, Dongminzhu Street, Changchun 130021, China*

[8] *College of Life Science, Jilin University, 2699 Qianjin Street, Changchun 130012, China*

[9] *Center for Biostatistics, Department of Population, Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA*

[10] *Department of Genetics and Genomics Science, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA*

Correspondence should be addressed to Jingkai Gu; gujk@jlu.edu.cn and Mayte Suárez-Fariñas; mayte.suarezfarinas@mssm.edu

In order to test if two chemically or pharmaceutically equivalent products have the same efficacy and/or toxicity, a bioequivalence (BE) study is conducted. The 80%/125% rule is the most commonly used criteria for BE and states that BE cannot be claimed unless the 90% CIs for the ratio of selected pharmacokinetics (PK) parameters of the tested to the reference drug are within 0.8 to 1.25. Considering that estimates of these PK parameters are derived from the concentration-versus-time curves, a direct comparison between these curves motivates an alternative and more flexible approach to test BE. Here, we propose to frame the BE test in terms of an equivalence of concentration-versus-time curves which are constructed using local polynomial smoother (LPS). A metric is presented to quantify the distance between the curves and its 90% CIs are calculated via bootstrapping. Then, we applied the proposed procedures to data from an animal study and found that BE between a generic drug and its brand name cannot be concluded, which was consistent with the results by applying the 80%/125% rule. However, the proposed procedure has the advantage of testing only on a single metric, instead of all PK parameters.

1. Introduction

In order to test if two chemically or pharmaceutically equivalent products, for example, a generic drug and its brand name, have the same efficacy and/or toxicity, a bioequivalence (BE) study is usually conducted [1–3]. The objective of

a BE trial is to determine whether the test (T) and the reference (R) formulation of a pharmaceutical product are "equivalent" with respect to blood concentration $\times$ time profiles. In contrast to a difference test, the null hypothesis in an equivalence test states that two agents differ in terms of the endpoint under consideration by at least the minimum

tolerable amount, called the equivalence margin Δ, whereas the alternative hypothesis states that such difference is less than the equivalence margin Δ:

$$H_0 : |\mu| \geq \Delta,$$
$$\text{versus } H_1 : |\mu| < \Delta. \tag{1}$$

In a BE test, some parameters derived from the concentration-versus-time curves are evaluated. Those parameters traditionally include the area under plasma concentration-versus-time curves (AUC_0^t), peak plasma concentrations ($C_{\max}$), and its corresponding time ($T_{\max}$). Among them, AUC is the most accepted measure of absorption rate. $C_{\max}$ is also of importance because, for some drugs, a certain level of concentration needs to be reached to guarantee the desired therapeutic effect. $T_{\max}$ is a relevant measure for drugs such as antibiotics that must reach the peak concentration quickly, but it may not be an appropriate measure for drugs requiring multiple dosage before a therapeutic effect is observed. Since none of these three parameters are universally superior to the others, a BE test usually considers them together. For example, under FDA regulations, BE can be claimed only when the 90% confidence intervals (CIs) for the ratio of $C_{\max}$, $\text{AUC}_{(0-t)}$, and $\text{AUC}_{(0-\infty)}$ of the tested (e.g., generic) to the reference drug (e.g., brand name) are within 80% to 125% [4]. This is referred to as 80%/125% rule, corresponding to a ± 0.223 rule on the logarithmic scale (of the T/R ratio). Hence, the value of Δ is usually set to 0.223. Of note, the 90% CIs of the BE-endpoint represent a 0.05 significance level on the equivalence test since the hypothesis-testing problem in an equivalence test is divided into two one-sided hypothesis tests [2, 5, 6]: one where the null hypothesis states that the difference between two agents is less than $-\Delta$ whereas the other one assumes such difference is larger than Δ:

$$H_{01} : \mu < -\Delta,$$
$$\text{versus } H_{a1} : \mu \geq -\Delta,$$
$$H_{02} : \mu > \Delta,$$
$$\text{versus } H_{a2} : \mu \leq \Delta. \tag{2}$$

Although the BE parameters can be easily obtained from either one of the concentration-versus-time curves using a suitable pharmacokinetic (PK) model such as a one-compartment model or a nonparametric method [7], there are many parameters to be tested, which inflates the type I error rate, requiring the adjustment for multiple comparisons. Because these parameters may be highly correlated, such an adjustment is challenging.

On the other hand, some researchers had pointed out that the requirement of all confidence intervals falling within the equivalence bounds might lead to a conservative result, depending on the correlations among the PK parameters and the study power [8, 9]. To address this, simultaneous testing of all PK parameters had been explored and developed; see, for example, the semiparametric Bayesian approach proposed by Ghosh and Gönen [10]. However, such multivariate methods are less popular than the univariate approach of testing PK

parameters, partially because of the modeling complexity associated with multivariate methods as compared with their univariate counterparts.

More importantly, when the confidence intervals of $C_{\max}$, $T_{\max}$, and AUC between two drugs all fall within the equivalence boundary (such that BE is concluded), it does not imply that the drugs are clinically equivalent since the overall shapes of the concentration-versus-time curves may in fact differ [11]. A falsely determined BE between two drugs when they are not clinically equivalent may be very harmful to the public. To alleviate these limitations and drawbacks, we propose to make a direct comparison of the concentration-versus-time curves that accounts for the differences between the profile shapes for BE testing. The main objective of this paper is to present a strategy to test BE with the aid of local polynomial smoother (LPS), which is used to construct concentration-versus-time curves. Then, a summary statistic, whose standard error is estimated by bootstrapping, is defined. This allows the calculation of CIs upon which decisions over equivalence between such two curves are made.

LPS is a flexible nonparametric regression method to model curves or surfaces. Specifically, the fitted regression function at a covariate value x is based only on observations within a prespecified neighborhood of x. LPS can be traced back to the late 19th century in actuarial sciences where it was used to estimate the gradation of mortality rates and in time series modeling [12]. In contrast, another popular nonparametric regression method, locally weighted scatterplot smoothing (LOWESS), obtained smooth fitted values by a weighted linear least squares regression over the prespecified spans. In this study, however, we do not emphasize their differences and instead treat them as synonymous.

With advances in computing, the potential applications of LPS continue to expand. For example, it has been used for dye normalization of two-color microarray experiments [13]. Recently, Anders and Huber [14] used LPS in R's DESeq package to estimate the mean-variance relationship in a negative binomial model, a technique now commonly used to model RNA-Seq data. The popularity of LPS in statistical applications may be due to its several advantages including satisfactory boundary behavior and straightforward interpretability of the nonlinear relationships [15, 16].

2. Methods and Materials

2.1. Experimental Data. The study included 24 beagle dogs weighting 9–11 kg, and those dogs were randomly assigned to receive either single or multiple (every four weeks for 3 times) administrations of either 1.4 mg/kg Sandostatin or a generic drug developed by GenSci. There were 6 dogs including 3 females and 3 males per treatment-dosage group. Blood samples for PK were collected into tubes containing heparin sodium and centrifuged at room temperature with 3500 rpm for 20 min to obtain plasma samples. This animal study was conducted in accordance with the rules and regulations of Chinese Pharmacopoeia, 2000 Edition/Version 2 and approved by the Committee on the Ethics of Animal Experiments of Jilin University. The plasma concentrations were

determined using an Agilent 1100 liquid chromatography-tandem mass spectrometry. In the single-dose regimen, blood samples were collected at the initial time corresponding to hour 0 and at the hours of 0.25, 0.5, 1, 1.5, 2, 4, 8, 24, 48, 96, 144, 192, 240, 336, 432, 528, 624, 720, 816, 912, 1008, 1104, and 1200.

Based on [7], PK parameters in this study were calculated as follows. For each dog, the maximum plasma concentration (C_{max}) and its corresponding time (T_{max}) were determined by visual inspection of the profiles. The apparent terminal elimination rate constant (λ) was calculated by linear regression of the natural logarithms of the terminal plasma concentrations. The terminal half-life ($t_{1/2}$) was derived as $0.693/\lambda$. The area under the curve (AUC) to the last measured point (AUC_0^t) was calculated using the trapezoidal rule. The AUC for the plasma concentration-versus-time function from 0 hours to infinity ($AUC_{0-\infty}$) was calculated as the sum of AUC_0^t and C_t/λ, while C_t was the last quantifiable concentration.

2.2. The LPS Model.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be covariate-response pairs for n observations. To simplify the notation, suppose there is only one covariate. Then, the response Y_i is related to the covariate X_i through the following model:

$$Y_i = g(X_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3}$$

where the error term ε_i is assumed to be independently distributed with zero mean and variance equal to σ^2. The goal of a local polynomial smoother is to estimate the smooth function $g(x) = E(Y \mid X = x)$ using an approximation provided by a local polynomial of low order in the neighborhood of the point x. Usually, the model includes only first- and second-degree polynomials, that is, either locally linear or locally quadratic. This is because any function can be well approximated in a small neighborhood by low-order polynomials and the inclusion of higher order terms often does not improve the model fit dramatically and even may lead to overfitting instead. Specifically, for $x - h \leq x_i \leq x + h$,

$$g(x_i) \approx \sum_{j=0}^{p} \beta_j (x_i - x)^j. \tag{4}$$

The abovementioned approximation is fitted by local weighted least squares, where the estimated coefficients are those that minimize the following function:

$$\sum_{i=1}^{n} W\left(\frac{(x_i - x)}{h}\right) \left(y_i - \sum_{j=0}^{p} \beta_j (x_i - x)^j\right)^2, \tag{5}$$

where the weighted function W gives greater weight to x_i in the vicinity of x and is usually specified as a symmetric bounded function such as a normal kernel function.

There are two commonly used approaches for the selection of bandwidth h in the equation above [12]. One choice is simply to make the bandwidth a constant, which is adequate if the regression function behaves smoothly. The other one is to allow the bandwidth to change as a function of x. In our applications, we chose the latter by using a nearest-neighbor bandwidth selection method [17]. Briefly, in the nearest-neighbor bandwidth selection method, a new parameter α

was included, which was multiplied by the sample size n and rounded up to an integer k. Then, the bandwidth $h(x)$ was defined as the distance from x to the kth closest x_i. In this study, measurements were taken at nonuniformly spaced time points, with more frequent measurements at the beginning, where a larger variation of the PK dynamic was expected among subjects. Under this condition, the nearest-neighbor bandwidth selection approach will provide better fit than the constant bandwidth. The nearest-neighbor bandwidth parameter α was determined by a leave-one-out (LOO) cross-validation due to the small sample size. Specifically, for a grid of possible α values, cross-validated mean square error (MSE) was calculated, and the α corresponding to the minimum MSE was selected.

2.3. Statistical Language.

The statistical analysis was carried out in the R language version 2.15 (http://www.r-project.org/). The R-code is available upon request.

3. Results and Discussion

3.1. Testing the Equivalence of Two Curves by Combining LPS and Bootstrapping: The Proposed Procedure.

Here, we propose to evaluate bioequivalence between drugs by directly comparing the plasma concentration curves PC(t). Suppose we have two treatments where n_1 subjects were randomly assigned to receive the test drug T and n_2 to the reference drug R. We proposed to reframe the hypothesis represented in (1) as

$$H_0 : d\left(g_T(t), g_R(t)\right) > \Delta, \tag{6}$$

$$\text{versus } H_1 : d\left(g_T(t), g_R(t)\right) \leq \Delta, \tag{7}$$

with d representing the difference between the PK curves for T and R drugs. For the study subjects, the plasma concentrations of these drugs were measured for those at a grid of time points. Using LPS (details in Section 2), concentration-versus-time curves for both treatments can be estimated as $g_R(x)$ and $g_T(x)$. To evaluate if the differences between two curves are within the tolerable range, we define the following estimator of the average difference between the two curves at a grid of prespecified time points, for example, the time points where the concentrations were taken in this study:

$$\ln(\hat{r}) = \frac{\sum_{k=1}^{K} \left|\ln\left(\hat{y}_{Tt_k}/\hat{y}_{Rt_k}\right)\right|}{K}, \tag{8}$$

where $\hat{y}_{Tt_k}$ and $\hat{y}_{Rt_k}$ are the fitted values of plasma concentration using LPS at time t_k for the test drug T and the reference drug R, respectively. Here, $k = 1, 2, \ldots, K$, and K is the number of time points at which the difference between two curves is evaluated. Note that $\ln(r)$ is a measure of the difference between the curves and when the two curves are identical, and they should be identical at all time points and thus the value of $\ln(r)$ should be 0.

The standard error for $\ln(\hat{r})$ can be estimated using bootstrapping [18]. For each bootstrapped replicate, LPS curve
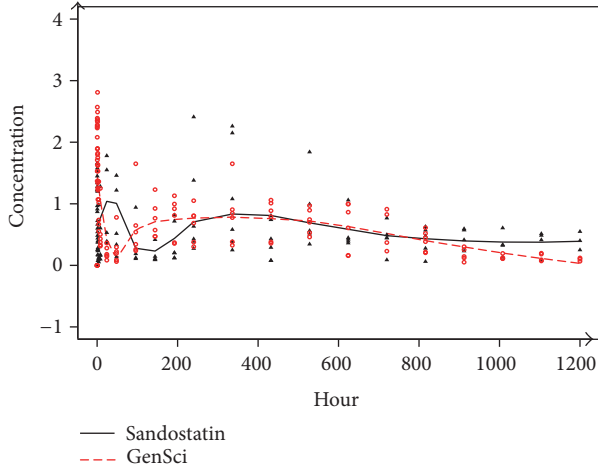
FIGURE 1: The fitted local polynomial smoother (LPS) curves for both GenSci and Sandostatin. Black triangle: the observed concentrations for dogs using Sandostatin; red circle: the observed concentrations for dogs using GenSci.

TABLE 1: The 90% CIs of PK parameters.

90% CI	Single dose $N = 6$	Combined data $N = 12$
$C_{\max}$	(72.23, 160.71)	(82.93, 160.88)
$T_{0.3\,\text{ng/nL}}$	(9.94, 382.63)	(9.51, 144.60)
AUC_{0-1200}	(71.98, 176.08)	—
AUC_{0-720}	(61.74, 169.85)	(91.44, 208.50)

was fitted for T and R and the difference between them is calculated using (8). The 90% CIs of r can be obtained as

$$\left(e^{\ln(\widehat{r})-1.645\times se(\ln(\widehat{r}))}, e^{\ln(\widehat{r})+1.645\times se(\ln(\widehat{r}))}\right). \tag{9}$$

If the 90% CIs is within 0.8~1.25, then the equivalence between two curves is claimed.

3.2. Case Study. Octreotide is an octapeptide that mimics natural somatostatin pharmacologically even though it is a more potent inhibitor of growth hormone, glucagon, and insulin than the natural hormone. Brand name Sandostatin (Novartis Pharmaceuticals) of octreotide had been approved for the treatment of several diseases, such as acromegaly and gigantism. To evaluate the BE between a generic octreotide developed by GenSci Pharmaceuticals (Changchun, China) and Brand name Sandostatin, a pilot study using 24 beagle dogs was conducted. The generic drug is referred to as GenSci hereafter. Using this study as an example, we showcase how to incorporate LPS seamlessly with other statistical methods to test BE. Once concentration-versus-time curves are obtained via LPS, testing for BE will be akin to testing if the two LPS-derived curves are the same.

Firstly, we conducted an equivalence test using the procedure described in Methods. The fitted plasma concentration PC(t) curves are shown in Figure 1. Interestingly, the curves estimated by LPS suggested some PK differences between the two drugs. Specifically, we observed a surge in plasma concentration and then a steep drop in the GenSci group. In addition, the GenSci group entered the plateau stage faster and remained in this stage longer than the Sandostatin group, dropping down at a sharper scope. Based on these observations, we suspected that GenSci might have a better PK behavior than Sandostatin.

The 90% CIs for the proposed metric of GenSci against Sandostatin were calculated as 119.02% ~166.99%. With its upper bound above 125%, the equivalence between the two

drugs cannot be established based on this method. As a comparison, we also estimated the 90% CI of the ratio of the traditionally considered PK parameters (i.e., (AUC_0^t), $C_{\max}$, and $T_{\max}$) based on the FDA guidelines [4]. The results summarized in Table 1 show that all 90% CIs of the ratio of (AUC_0^t), $C_{\max}$, and $T_{\max}$ of GenSci against Sandostatin were outside 80% to 125%, the FDA-sanctioned BE criteria. In summary, BE cannot be established using either the FDA-criteria or our proposed procedure. Nevertheless, the proposed procedure has the advantage of testing only a single statistic and hence avoiding multiple hypotheses issues.

The sample size of 6 dogs per group raises major concerns in light of the failure to establish BE between GenSci and Sandostatin. To explore the impact of the sample size in this pilot study, we combined the data from this single-dose cohort with additional data from a multiple-dose cohort, but including only data collected in the first 720 hours where no boosts were given and the experimental methods\conditions were exactly the same as in the single-dose cohorts. Interested readers may see Section 2 for more information on how the experiments were conducted. Analysis on the combined cohorts including 12 dogs per group led to similar results. Given that this application represents a pilot study, as preliminary work for larger studies, no confirmative conclusions were drawn here. By presenting this study, our intent is to illustrate the LPS-based BE test as an alternative to the classical univariate tests on PK parameters.

3.3. Simulation Study. We conducted a simulation study to characterize our BE-testing procedure. Here, we used the PK model presented in [11] where the concentration function is represented by

$$c(t) \propto \frac{K_a}{K_a - K_e}\left(e^{-k_e t} - e^{-k_a t}\right), \tag{10}$$

where k_a and k_e are the absorption and the elimination rate constants. The concentration-versus-time curves for the reference drug and test drug were simulated as

$$\log(Y_R) = \log(c(t)) + \varepsilon,$$
$$\log(Y_T) = \log(c(t)) + \nabla + \varepsilon, \tag{11}$$
$$\varepsilon \sim \text{MVN}\left(0_p, \Sigma_{p\times p}\right),$$

where the correlation between two measures, on the logarithm scale, from the same subject k ($k = 1, 2, \ldots, 2n$) was specified as a positive constant; that is, corr($\varepsilon_{ik}, \varepsilon_{jk}$) =

TABLE 2: The exact coverage of the 90% CIs for $\ln(r)$ (the proposed measure of the difference between two curves) using simulations.

	$n = 12$	$n = 18$	$n = 24$	$n = 30$	$n = 50$
	$\nabla = 0$, $k_{aT}/k_{aR} = 1$ (two curves are identical)				
$\rho = 0$	100	100	100	100	100
$\rho = 0.2$	100	100	100	100	100
$\rho = 0.5$	99.9	100	100	100	100
$\rho = 0.8$	97.8	99.6	100	100	100
	$\nabla = 0.223$, $k_{aT}/k_{aR} = 1$ (the difference is within the equivalence bound)				
$\rho = 0$	100	100	100	100	100
$\rho = 0.2$	100	100	100	100	100
$\rho = 0.5$	98.7	99.7	100	100	100
$\rho = 0.8$	95.5	97.2	99.1	99.7	100
	$\nabla = 1$, $k_{aT}/k_{aR} = 1$ (the difference is beyond the equivalence bound)				
$\rho = 0$	0	0	0	0	0
$\rho = 0.2$	3.3	0.7	0.3	0	0
$\rho = 0.5$	11.6	6.6	3.7	3.2	0.7
$\rho = 0.8$	18	11.2	7.9	5.8	2.6
	$k_{aT}/k_{aR} = 1.25$, $\nabla = 0$ (the difference is within the equivalence bound)				
$\rho = 0$	100	100	100	100	100
$\rho = 0.2$	99.9	100	100	100	100
$\rho = 0.5$	97.7	99.1	99.7	100	100
$\rho = 0.8$	95	97.5	98.8	99.3	100
	$k_{aT}/k_{aR} = 3$, $\nabla = 0$ (the difference is beyond the equivalence bound)				
$\rho = 0$	0	0	0	0	0
$\rho = 0.2$	5	1.5	0.6	0.4	0
$\rho = 0.5$	13.2	8.2	5.4	3.8	0.7
$\rho = 0.8$	17.7	15.8	10.2	6.5	2.7

Note: k_{aT} represents the absorption rate for the generic drug and k_{aR} represents the absorption rate for the reference drug; n represents the sample size of each group; ∇ represents a constant difference between the concentrations of two drugs on the logarithm scale; ρ is the correlation coefficient between measures over time from the same subject; its influence on the estimation is expected to be bigger when its value is larger.

ρ for $i \neq j$. The curves above differ only by a constant ∇ on the logarithm scale. In the second set of simulations, the two concentration curves for the test drug and the reference drug were set to have different absorption rates (i.e., k_{aT} and k_{aR}) instead of $\nabla \neq 0$.

By varying the values for ρ, ∇, and k_{aT}/k_{aR} and the sample size, we simulated 100 different scenarios in total. The empirical coverage of the 90% CIs for each scenario was calculated via simulating 1000 bootstrap replicates. The simulation results are presented in Table 2, from which we found that a sample size of 24 per group is large enough to provide an adequate statistical power for the proposed testing procedure. Furthermore, the empirical coverage of the 90% CIs when two curves are equivalent is approximately 95% or above, suggesting the 80/125% rule is applicable for the proposed metric.

3.4. Our Extension to the Superiority Test between Curves. It is worth pointing out that by virtue of being coupled with a permutation test [19, 20] to calculate an overall p value, LPS can also be used to evaluate if two curves are different. In this case, the hypothesis tested is

$$H_0 : d\left(g_A\left(x\right), g_B\left(x\right)\right) = 0,$$

$$\text{versus } H_1 : d\left(g_A\left(x\right), g_B\left(x\right)\right) \neq 0,$$

(12)

with d representing the difference between the two curves. As shown on the simulation studies (in Supplementary Material available online at http://dx.doi.org/10.1155/2016/4680642), the permutation-based test provides a valid means for testing the difference of two curves, achieving the desired type I error rate. We applied this procedure to longitudinal data collected by Orange et al. [21] and showed that the results obtained were consistent with those obtained by using a generalized estimating equation model [22, 23]. Therefore, the proposed procedure to test the difference between curves using LPS can be employed in a wide range of applications in longitudinal data analysis.

4. Conclusions

In this study, we presented a framework which directly compares the concentration-versus-time curves constructed using LPS to test BE. This approach avoids multiple testing by evaluating the equivalence between curves with one single metric, instead of multiple PK parameters. Furthermore, since LPS can be viewed as an extension of linear weighted regression and can be easily implemented in any standard statistical software, our proposed procedure can be easily utilized in other applications.

van der Meersch et al. [24] have argued that the objectives
and methodology of bioequivalence trials differ greatly from
those of noninferiority and equivalence trials due to the
decision criteria and study design. In our opinion, their
claims are not persuasive. In terms of decision criteria, it is
difficult to define an equivalence margin in a longitudinal
sense because the measures such as plasma concentrations are
collected over time. It might represent one major reason why
the FDA and other regulatory agencies have defined BE by
using confidence limits for a grid of PK parameters instead
of a single hypothesis testing. Regarding the study design,
van der Meersch et al. [24] deemed a crossover design to be
suitable and appropriate. In practice, however, there are many
BE studies using a parallel design that the FDA reports [4]
explicitly listed the cases where a parallel design is considered
more suitable than a crossover design. Given that a BE test
is a longitudinal equivalence test by nature, we believe that
the use of nonparametric smoothers such as LPS as a basis
to construct the concentration-versus-time curves for direct
comparisons warrants additional attention and research.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] B. Ewald, "Making sense of equivalence and non-inferiority
trials," *Australian Prescriber*, vol. 36, no. 5, pp. 170–173, 2013.

[2] J. Ocana, M. Sánchez, A. Sanchez, and J. L. Carrasco, "On
equivalence and bioequivalence testing," *SORT. Statistics and
Operations Research Transactions*, vol. 32, no. 2, pp. 151–176,
2008.

[3] A. S. Kesselheim, A. S. Misono, J. L. Lee et al., "Clinical equiva-
lence of generic and brand-name drugs used in cardiovascular
disease: a systematic review and meta-analysis," *Journal of the
American Medical Association*, vol. 300, no. 21, pp. 2514–2526,
2008.

[4] FDA, *Guidance for Industry: Bioequivalence Guidance*, 2006.

[5] L. E. Barker, E. T. Luman, M. M. McCauley, and S. Y. Chu,
"Assessing equivalence: an alternative to the use of difference
tests for measuring disparities in vaccination coverage," *Ameri-
can Journal of Epidemiology*, vol. 156, no. 11, pp. 1056–1061, 2002.

[6] E. Walker and A. S. Nowacki, "Understanding equivalence and
noninferiority testing," *Journal of General Internal Medicine*, vol.
26, no. 2, pp. 192–196, 2011.

[7] S. S. Jambhekar and P. J. Breen, *Basic Pharmacokinetics*, Phar-
maceutical Press, 2009.

[8] H. Quan, J. Bolognese, and W. Yuan, "Assessment of equivalence
on multiple endpoints," *Statistics in Medicine*, vol. 20, no. 21, pp.
3159–3173, 2001.

[9] W. Wang, J. T. G. Hwang, and A. Dasgupta, "Statistical tests for
multivariate bioequivalence," *Biometrika*, vol. 86, no. 2, pp. 395–
402, 1999.

[10] P. Ghosh and M. Gönen, "Bayesian modeling of multivariate
average bioequivalence," *Statistics in Medicine*, vol. 27, no. 13, pp.
2402–2419, 2008.

[11] H. A. Bayoud and A. M. Awad, "Performance of several
bioequivalence metrics for assessing the rate and extent of
absorption," *Journal of Bioequivalence and Bioavailability*, vol.
3, no. 7, pp. 174–177, 2011.

[12] W. S. Cleveland and C. Loader, "Smoothing by local regression:
principles and methods," in *Statistical Theory and Computa-
tional Aspects of Smoothing*, W. Härdle and M. G. Schimek, Eds.,
Contributions to Statistics, pp. 10–49, Springer, New York, NY,
USA, 1996.

[13] J. Quackenbush, "Microarray data normalization and transfor-
mation," *Nature Genetics*, vol. 32, no. 5, supplement, pp. 496–
501, 2002.

[14] S. Anders and W. Huber, "Differential expression analysis for
sequence count data," *Genome Biology*, vol. 11, no. 10, article
R106, 2010.

[15] W. S. Cleveland, "Robust locally weighted regression and
smoothing scatterplots," *Journal of the American Statistical
Association*, vol. 74, no. 368, pp. 829–836, 1979.

[16] J. Fan and I. Gijbels, "Variable bandwidth and local linear
regression smoothers," *The Annals of Statistics*, vol. 20, no. 4, pp.
2008–2036, 1992.

[17] C. R. Loader, "Bandwidth selection: classical or plug-in?" *The
Annals of Statistics*, vol. 27, no. 2, pp. 415–438, 1999.

[18] B. Efron, "Bootstrap methods: another look at the jackknife,"
The Annals of Statistics, vol. 7, no. 1, pp. 1–26, 1979.

[19] P. I. Good, *Permutation, Parametric and Bootstrap Tests of
Hypotheses*, Springer, Berlin, Germany, 3rd edition, 2005.

[20] W. J. Welch, "Construction of permutation tests," *Journal of the
American Statistical Association*, vol. 85, no. 411, pp. 693–698,
1990.

[21] D. E. Orange, N. E. Blachere, J. Fak et al., "Dendritic cells
loaded with FK506 kill T cells in an antigen-specific manner
and prevent autoimmunity in vivo," *eLife*, vol. 2013, no. 2, article
e00105, 2013.

[22] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using
generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22,
1986.

[23] S. L. Zeger and K. Y. Liang, "Longitudinal data analysis for
discrete and continuous outcomes," *Biometrics*, vol. 42, no. 1, pp.
121–130, 1986.

[24] A. van der Meersch, A. Dechartres, and P. Ravaud, "Quality of
reporting of bioequivalence trials comparing generic to brand
name drugs: a methodological systematic review," *PLoS ONE*,
vol. 6, no. 8, article e23611, 2011.