Research Article

Statistics
in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

Comparison of Drug Dissolution Profiles: A Proposal Based on Tolerance Limits

Shuyan Zhai,Thomas Mathew* and Yi Huang

Meaningful comparison of the dissolution profiles between the reference and test formulations of a drug is critical for assessing similarity between the two formulations, and for quality control purposes. Such a dissolution profile comparison is required by regulatory authorities, and the criteria used for this include the widely used difference factor f_1 and a similarity factor f_2, recommended by the FDA. In spite of their extensive use in practice, the two factors have been heavily criticized on various grounds; the criticisms include ignoring sampling variability and ignoring the correlations across time points while using the criteria in practice. The goal of this article is to put f_1 and f_2 on a firm statistical footing by developing tolerance limits for the distributions of f_1 and f_2, so that both the sampling variability and the correlations over time points are taken into account. Both parametric and nonparametric approaches are explored, and a bootstrap calibration is used to improve accuracy. In particular, the methodology in this article can be used to compute upper confidence limits for the medians of f_1 and f_2. Simulated coverage probabilities show that the method leads to accurate tolerance limits. Two examples are used to illustrate the methodology. The overall conclusion is that the tolerance limit based approach offers a statistically rigorous procedure for in vitro dissolution testing.

Keywords: Bootstrap calibration; Difference factor; Dissolution testing; Order statistics; Parametric bootstrap; Similarity factor.

1. Introduction

Dissolution profile comparison is critical for both drug development and quality control purposes. Both industry and regulatory authorities use in-vitro information provided by dissolution profiles to predict in-vivo performance, to establish the final dissolution specification for drug dosage, and to assess the similarity of drug formulations prior to and after moderate changes. The "moderate changes" mentioned in the U. S. FDA's guidance documents [1, 2, 3, 4] include scale-up, manufacturing changes, component and composition changes and equipment and process changes. To ensure the continued quality of the drug before and after such changes, without carrying out costly bioequivalence studies, similarity comparisons of dissolution profiles are required for the approval of such moderate changes, and are considered adequate for determining the similarity of drug formulations.

Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, U.S.A.
** Correspondence to: Department of Mathematics and Statistics, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, U.S.A. E-mail: mathew@umbc.edu*

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

A dissolution profile captures the percentage of the active drug ingredient dissolved (based on one dosage unit) at multiple pre-specified time points. A general dissolution comparison contains two or more drug formulations to be compared, and on each formulation at least six profiles are obtained from one or more lots in a batch. The number of sampling time points may vary from drug to drug, affected by the speed of dissolution of the active drug ingredient [1, 2, 3, 4]. Let $Y_{R,i} = (Y_{R,i}^1, ..., Y_{R,i}^K)'$, $i = 1, ..., n_R$, and $Y_{T,j} = (Y_{T,j}^1; ...; Y_{T,j}^K)'$, $j = 1, ..., n_T$, be the observed dissolution profiles for the ith and jth dosage units from the reference and test formulations, respectively, where K denotes the number of pre-specified time points. Let $\bar{Y}_R = (\bar{Y}_R^1, \bar{Y}_R^2, ..., \bar{Y}_R^K)'$ and $\bar{Y}_T = (\bar{Y}_T^1, \bar{Y}_T^2, ..., \bar{Y}_T^K)'$ denote the sample mean profiles for the reference and test drugs, respectively. The two criteria commonly used and recommended by the FDA for dissolution profile comparison [5] are:

$$\begin{aligned} \text{Difference factor: } f_1 &= \frac{\sum_{t=1}^{K} |\bar{Y}_R^t - \bar{Y}_T^t|}{\sum_{t=1}^{K} \bar{Y}_R^t} \times 100\% \\ \text{Similarity factor: } f_2 &= 50 \times \log_{10} \left(\left[1 + \frac{1}{K} \sum_{t=1}^{K} w_t (\bar{Y}_R^t - \bar{Y}_T^t)^2 \right]^{-0.5} \times 100 \right), \end{aligned} \tag{1}$$

where the w_t's are the pre-specified weights, often set to 1 (the weights are set to 1 throughout this paper). The FDA guidance document [1] indicates that f_1 values less than 15 (i.e., 0-15) and f_2 values greater than 50 (i.e., 50-100) maybe taken as evidence to conclude the equivalence of the dissolution profiles of the test and reference products. Notice that $f_1 = 0$ and $f_2 = 100$ for two identical dissolution profiles [1, 2, 3, 4].

In spite of their popularity, f_1 and f_2 have quite a few limitations [6]. First, f_1 is very sensitive to the choice of reference lots. Simply interchanging the roles of reference and test batches will change the value of f_1 in general, even though the similarity evaluation should not be affected. As a result, f_2 is more popular in practice. Secondly, both of them are sensitive to K - total numbers of time points, especially when both dissolution profiles level off. FDA sets clear guidance on the total number of time points for different types of drugs, in order to address such concerns. Third, f_1 and f_2 don't take into consideration the correlation among the repeated dissolution measures across times. Finally, the similarity evaluation using f_1 and f_2 ignores the sampling variability in the data. A critical evaluation of the factor f_2 is provided in the paper by [7].

Both model-independent methods and model-dependent methods have been developed for dissolution comparisons to address the last two concerns [6]. Here the term "model dependent" refers to the use of appropriate models for the mean vectors of the dissolution profiles, modeled as a function of time. Models used for this purpose include the exponential model, Probit model, Gompertz model, Logistic model and Weibull Model; the Weibull model has been noted to provide a good fit for the mean vectors [8, 9, 10]. A population version of f_2 is considered in [11]; the authors modified f_2 by replacing $\bar{Y}_R^t$ and $\bar{Y}_T^t$ in (1) with the corresponding population mean vectors, so that the criteria are unknown parametric functions, and then discussed hypothesis testing procedures for dissolution comparison [11, 12, 13, 14].

Our purpose is to develop procedures for dissolution comparisons based on the criteria f_1 and f_2 in (1) by taking into account simultaneously both the sampling variability and the correlations across multiple time points. Since drug responses from individual subjects are of interest in practice, we shall consider criteria similar to f_1 and f_2 by replacing $\bar{Y}_R$ and $\bar{Y}_T$ by the respective individual response vectors Y_R and Y_T, respectively. Indeed, [5] suggested such criteria based

on individual responses. We shall denote the resulting criteria by g_1 and g_2, defined as

$$\begin{aligned} g_1 &= \frac{\sum_{t=1}^{K} |Y_R^t - Y_T^t|}{\sum_{t=1}^{K} Y_R^t} \times 100\% \\ g_2 &= 50 \log_{10} \left(\left[1 + \frac{1}{K} \sum_{t=1}^{K} w_t (Y_R^t - Y_T^t)^2 \right]^{-0.5} \times 100 \right) \\ &= 50 \times \log_{10} \left([1 + X]^{-0.5} \times 100 \right), \text{where } X = \frac{1}{K} \sum_{t=1}^{K} w_t (Y_R^t - Y_T^t)^2. \end{aligned} \tag{2}$$

Note that both g_1 and g_2 are random variables. Furthermore, the similarity factor g_2 is large if the quantity X defined in (2) is small. Thus, in the context of the similarity factor g_2 defined above, estimating a cutoff point below which a specified percentage or more of the X distribution will fall (with a given confidence level) can be used to assess dissolution similarity. Such an upper cutoff value (to be estimated using a random sample) is referred to as an upper tolerance limit for the distribution of X. If $\hat{X}_U$ is an upper tolerance limit for X, then a lower tolerance limit, say $\hat{g}_{2L}$, for the distribution of g_2 is given by:

$$\hat{g}_{2L} = 50 \times \log_{10} \left([1 + \hat{X}_U]^{-0.5} \times 100 \right) \tag{3}$$

If $\hat{g}_{2L}$ is large (say, greater than 50 according to the FDA guideline), then the g_2 distribution is mostly above 50, with a certain confidence level. If so, we conclude that the dissolution profiles across the test and reference populations are similar. In Section 2, our tolerance limit method is described in the context of g_2. A similar approach can be adopted for the difference factor g_1 given in (2), and also for the factors f_1 and f_2 given in (1). It should be noted that we have developed our methodology in a parametric set up, assuming multivariate normality, and in a non-parametric set up, without making any distributional assumption. Section 3 presents simulation studies on the accuracy of our proposed approaches. Simulated coverage probabilities show that our methodology is accurate in the parametric as well as non-parametric set ups. Section 4 presents two real applications based on published dissolution profile data. Conclusions and discussions appear in Section 5. Since the computation of a tolerance limit uses the actual population distribution, the variability in the population distribution is taken into account, together with the correlations across different time points. Furthermore, the sampling variability is also taken into account through the use of an associated confidence level. In other words, our approach offers a rigorous method for assessing dissolution profile similarity, based on criteria currently in use. In particular, our methodology can be used to compute an upper confidence limit of the median of each of the random variables g_1, g_2, f_1 and f_2.

We conclude this introductory section with two observations. First of all, the methodologies used in our work are not new; we have used two existing methodologies, namely, non-parametric tolerance limit computation and bootstrap calibration, in order to develop a statistically valid approach for dissolution profile comparison based on criteria recommended by the FDA. As already noted, such an approach has been lacking, in spite of the widespread use of the FDA recommended criteria. Secondly, our methodology may appear somewhat cumbersome to understand and implement; see the computational steps in Algorithm 2 and Algorithm 4 in the next section. However, this should not be a hinderance in practical applications, since we have developed the necessary R code, available online as supporting material.

2. Tolerance limits for dissolution profile comparisons

By definition, an upper tolerance limit for the distribution of X defined in (2) is a limit computed from a random sample, so that a proportion p or more of the distribution of X is below the limit, with a given confidence level, say $1 - \alpha$. The

quantity p is referred to as the content of the one-sided tolerance interval, whose upper limit is the upper tolerance limit. Furthermore, the confidence level $1-\alpha$ reflects the sampling variability, since the tolerance limit is computed using a random sample. It is well known that an upper tolerance limit for X, having content p and confidence level $1-\alpha$, is simply a $100(1-\alpha)\%$ upper confidence limit for the pth percentile of X (Chapter 1, [15]). An upper tolerance limit can be computed parametrically or non-parametrically, and the latter is based on order statistics. Even though we are in a parametric set up, we face several difficulties when it comes to computing an upper tolerance limit for the distribution of X. First of all, neither the distribution of X, nor its percentile, is available in a closed form. Even if we are to ignore the parametric assumption, and decide to compute a non-parametric upper tolerance limit for X, we face the difficulty that a sample is not available from the distribution of X; samples are available from $Y_R \sim N(\mu_R, \Sigma_R)$ and $Y_T \sim N(\mu_T, \Sigma_T)$ and X is a function of Y_R and Y_T. In order to circumvent these difficulties, we proceed as follows. Based on samples Y_{Ri}, i = 1, 2, ..., n_R, and Y_{Ti}, i = 1, 2, ..., n_T from $N(\mu_R, \Sigma_R)$ and $N(\mu_T, \Sigma_T)$, respectively, obtain estimates of the unknown parameters μ_R, Σ_R, μ_T, and Σ_T, and denote the estimates by $\hat{\mu}_R$, $\hat{\Sigma}_R$, $\hat{\mu}_T$, and $\hat{\Sigma}_T$, respectively. Now generate B parametric bootstrap samples consisting of pairs (Y^*_{Rj}, Y^*_{Tj}) as $Y^*_{Rj} \sim N(\hat{\mu}_R, \hat{\Sigma}_R)$ and $Y^*_{Tj} \sim N(\hat{\mu}_T, \hat{\Sigma}_T)$, j = 1, 2, ..., B, where the Y^*_{Rj}s and the Y^*_{Tj}s are generated independently. However, note that we are pairing them. Now we let $X^*_j = \frac{1}{K}\sum_{t=1}^{K} w_t (Y^{*t}_{Rj} - Y^{*t}_{Tj})^2$, j = 1, 2, ..., B, where Y^{*t}_{Rj} and Y^{*t}_{Tj} are the tth components of the vectors Y^*_{Rj} and Y^*_{Tj}, respectively (t = 1, 2, ..., K). In order to compute a non-parametric upper confidence limit having content p and confidence level $1-\alpha$, we proceed using standard methodology as explained in Chapter 8 of [15]. Thus consider $W \sim \text{Binomial}(B, 1-p)$, and let k be the largest integer satisfying $P(W \geq k) \geq 1-\alpha$. We then select the $(B-k+1)$th order statistic among the X^*_j as our upper tolerance limit for the distribution of X. However, we don't expect the resulting upper tolerance limit to be accurate, since the sample used is a parametric bootstrap sample, and is not a sample from the distribution of X. In order to correct for this, we use a bootstrap calibration on the content p, and this finally provides the desired upper tolerance limit. The bootstrap calibration requires an estimate of the pth percentile of the distribution of X, which is not available in an analytic form. We shall however use an efficient approximation due to [16]; see the Appendix. Algorithm 1 and Algorithm 2 given below provide the steps necessary to implement the process just described for computing an upper tolerance limit. Algorithm 1 describes the computation of the non-parametric upper tolerance limit based on a parametric bootstrap sample, and Algorithm 2 explains the bootstrap calibration. We refer to [17], Chapter 18, for an explanation of the bootstrap calibration idea.

Algorithm 1 (Parametric bootstrap upper tolerance limit)

1. From the original samples Y_{Ri}, i = 1, 2, ..., n_R, and Y_{Ti}, i = 1, 2, ..., n_T, compute the unbiased estimates of the mean vectors μ_R and μ_T, and the covariance matrices Σ_R and Σ_T as

$$\hat{\mu}_R = \bar{Y}_R,\ \hat{\mu}_T = \bar{Y}_T,\ \hat{\Sigma}_R = \frac{1}{n_R - 1}\sum_{i=1}^{n_R}(Y_{Ri} - \bar{Y}_R)(Y_{Ri} - \bar{Y}_R)',$$

$$\text{and}\ \ \hat{\Sigma}_T = \frac{1}{n_T - 1}\sum_{i=1}^{n_T}(Y_{Ti} - \bar{Y}_T)(Y_{Ti} - \bar{Y}_T)',$$

   where $\bar{Y}_R$ and $\bar{Y}_T$ are the respective sample mean vectors. Then

$$\hat{\mu}_R \sim N(\mu_R, \frac{1}{n_R}\Sigma_R),\ \hat{\mu}_T \sim N(\mu_T, \frac{1}{n_T}\Sigma_T),$$
$$\hat{\Sigma}_R \sim W_K\left(n_R - 1, \frac{1}{n_R - 1}\Sigma_R\right),\ \text{and}\ \ \hat{\Sigma}_T \sim W_K\left(n_T - 1, \frac{1}{n_T - 1}\Sigma_T\right),$$

   where $W_r(m, \Sigma)$ denotes the $r-$dimensional Wishart distribution with df = m, and scale matrix equal to Σ.
2. Generate parametric bootstrap samples of size B each: $Y^*_{Rj} \sim N(\hat{\mu}_R, \hat{\Sigma}_R)$, and $Y^*_{Tj} \sim N(\hat{\mu}_T, \hat{\Sigma}_T)$, j = 1, 2, ..., B.

Write $Y^*_{Rj} = (Y^{1*}_{Rj}, Y^{2*}_{Rj}, ..., Y^{K*}_{Rj})'$, $Y^*_{Tj} = (Y^{1*}_{Tj}, Y^{2*}_{Tj}, ..., Y^{K*}_{Tj})'$, and compute $X^*_j = \frac{1}{K}\sum_{t=1}^{K}(Y^{t*}_{Rj} - Y^{t*}_{Tj})^2$, j = 1, 2, ..., B.

3. Let $W \sim \text{Binomial}(B, 1-p)$, and let k be the largest integer satisfying $P(W \geq k) \geq 1-\alpha$.
4. The $(B-k+1)$th order statistic among the X^*_js is then an upper tolerance limit for the distribution of $X = \frac{1}{K}\sum_{t=1}^{K}(Y^t_R - Y^t_T)^2$.

Algorithm 2 (Bootstrap calibration on the content p):

1. Let $\hat{X}_p$ denote an estimate of the pth percentile of X; see the Appendix for its computation.
2. Generate a bootstrap sample of size B_1 parametrically from the distributions of $\hat{\mu}_R$, $\hat{\Sigma}_R$, $\hat{\mu}_T$, $\hat{\Sigma}_T$:
   $\hat{\mu}^*_{Ri} \sim N(\hat{\mu}_R, \frac{1}{n_R}\hat{\Sigma}_R)$, $\hat{\mu}^*_{Ti} \sim N(\hat{\mu}_T, \frac{1}{n_T}\hat{\Sigma}_T)$,
   $\hat{\Sigma}^*_{Ri} \sim W_K\left(n_R - 1, \frac{1}{n_R-1}\hat{\Sigma}_R\right)$ and $\hat{\Sigma}^*_{Ti} \sim W_K\left(n_T - 1, \frac{1}{n_T-1}\hat{\Sigma}_T\right)$, i = 1, 2, ..., B_1.
3. For each i = 1, 2, ..., B_1, generate B_2 second level bootstrap samples as follows:

$$Y^{**}_{R,ij} \sim N(\hat{\mu}^*_{Ri}, \hat{\Sigma}^*_{Ri}), \text{ and } Y^{**}_{T,ij} \sim N(\hat{\mu}^*_{Ti}, \hat{\Sigma}^*_{Ti}),\ j = 1, ..., B_2.$$

   Write $Y^{**}_{R,ij} = (Y^{1**}_{R,ij}, Y^{2**}_{R,ij}, ..., Y^{K**}_{R,ij})'$, $Y^{**}_{T,ij} = (Y^{1**}_{T,ij}, Y^{2**}_{T,ij}, ..., Y^{K**}_{T,ij})'$ and compute

$$X^{**}_{ij} = \frac{1}{K}\sum_{t=1}^{K}(Y^{t**}_{R,ij} - Y^{t**}_{T,ij})^2,\ j = 1, ..., B_2, i = 1, ..., B_1.$$

4. Select s content values p_1, p_2, ..., p_s. For l = 1, 2, ..., s, let $W_l \sim \text{Binomial}(B_2, 1-p_l)$, and let k_l be the largest integer satisfying $P(W_l \geq k_l) \geq 1-\alpha$. For each i = 1, 2, ..., B_1, let $X^{**}_{i,(B_2-k_l+1)}$ denote the (B_2-k_l+1)th order statistic among the X^{**}_{ij} (j = 1 , 2, ..., B_2).
5. For each p_l, obtain the proportion of times (out of B_1) that $\hat{X}_p \leq X^{**}_{i,(B_2-k_l+1)}$.
6. Among all the p_l's, determine the value that makes the above proportion closest to $1-\alpha$; denote this value as $\hat{p}_0$.
7. Now implement Algorithm 1 using the content value $\hat{p}_0$.

Our method involves extensive use of the bootstrap, along with bootstrap calibration, and Algorithm 1 and Algorithm 2 provide a summary of the methodology under the multivariate normality assumption. However, the multivariate normality assumption of Y_R and Y_T may not always hold, in which case the parametric bootstrap algorithms are not appropriate. Instead, the bootstrap should be carried out non-parametrically for computing an upper tolerance limit for the distribution of the quantity X in (2). It should however be noted that for implementing the bootstrap calibration, it is necessary to have an estimate of the pth percentile of the distribution of X. Such an estimate can also be obtained using the bootstrap applied to the dissolution profile samples of sizes n_R and n_T, obtained for the reference drug and the test drug, respectively. For this, we proceed as follows. Let Y^*_R and Y^*_T represent observations selected with replacement from the dissolution profile samples of sizes n_R and n_T, and compute $X^* = \frac{1}{K}\sum_{t=1}^{K}(Y^{t*}_T - Y^{t*}_R)^2$. Repeat this many times, generating several values of X^*. The pth percentile of the X^*-values so obtained is an estimate of the pth percentile of the distribution of X. We shall once again use the notation $\hat{X}_p$ to denote the estimate so obtained. Here are the modified versions of Algorithm 1 and Algorithm 2, when the bootstrap is implemented non-parametrically:

Algorithm 3 (Non-parametric bootstrap upper tolerance limit):

1. Select B pairs of observations Y^*_{Rj} and Y^*_{Tj} randomly with replacement from the dissolution profile samples of sizes n_R and n_T for the reference drug and the test drug, respectively. Write $Y^*_{Rj} = (Y^{1*}_{Rj}, Y^{2*}_{Rj}, ..., Y^{K*}_{Rj})'$ and $Y^*_{Tj} = (Y^{1*}_{Tj}, Y^{2*}_{Tj}, ..., Y^{K*}_{Tj})'$, and compute $X^*_j = \frac{1}{K}\sum_{t=1}^{K}(Y^{t*}_{Tj} - Y^{t*}_{Rj})^2$, j = 1, 2, ..., B.
2. Let $W \sim \text{Binomial}(B, 1-p)$, and let k be the largest integer satisfying $P(W \geq k) \geq 1-\alpha$.

3. The $(B-k+1)$th order statistic among the X_j^*s is an upper tolerance limit for the distribution of $X = \frac{1}{K}\sum_{t=1}^{K}(Y_R^t - Y_T^t)^2$.

Algorithm 4 (Calibration on the content p):

1. Let $\hat{X}_p$ denote the non-parametric estimate of the pth percentile of $X = \frac{1}{K}\sum_{t=1}^{K}(Y_T^t - Y_R^t)^2$, computed as described earlier.
2. Non-parametrically generate B_1 bootstrap samples, each of size n_R drawn with replacement from the given dissolution profile sample of sizes n_R for the reference drug. Denote these B_1 samples by Y_{Ri1}^*, Y_{Ri2}^*, ..., $Y_{Rin_R}^*$, i = 1, 2, ..., B_1. Similarly generate B_1 bootstrap samples, each of size n_T drawn with replacement from the given dissolution profile sample of sizes n_T for the test drug. Denote these by Y_{Ti1}^*, Y_{Ti2}^*, ..., $Y_{Tin_T}^*$, i = 1, 2, ..., B_1.
3. For each i = 1, 2, ..., B_1, generate B_2 pairs of observations: $(Y_{R,ij}^{**}, Y_{T,ij}^{**})$, j =1, 2, ..., B_2, where the $Y_{R,ij}^{**}$'s are selected with replacement from Y_{Ri1}^*, Y_{Ri2}^*, ..., $Y_{Rin_R}^*$, and the $Y_{T,ij}^{**}$'s are selected with replacement from Y_{Ti1}^*, Y_{Ti2}^*, ..., $Y_{Tin_T}^*$. Write $Y_{R,ij}^{**} = (Y_{R,ij}^{1**}, Y_{R,ij}^{2**}, ..., Y_{R,ij}^{K**})'$, $Y_{T,ij}^{**} = (Y_{T,ij}^{1**}, Y_{T,ij}^{2**}, ..., Y_{T,ij}^{K**})'$ and compute

$$X_{ij}^{**} = \frac{1}{K}\sum_{t=1}^{K}(Y_{R,ij}^{t**} - Y_{T,ij}^{t**})^2,\ j = 1, ..., B_2,\ i = 1, ..., B_1.$$

4. Select s content values p_1, p_2, ..., p_s. For l = 1, 2, ..., s, let $W_l \sim \text{Binomial}(B_2, 1-p_l)$, let k_l be the largest integer satisfying $P(W_l \geq k_l) \geq 1-\alpha$. For each i = 1, 2, ..., B_1, let $X_{i,(B_2-k_l+1)}^{**}$ denote the (B_2-k_l+1)th order statistic among the X_{ij}^{**} (j = 1 , 2, ..., B_2).
5. For each p_l, obtain the proportion of times (out of B_1) that $\hat{X}_p \leq X_{i,(B_2-k_l+1)}^{**}$.
6. Among all the p_ls, determine the value that makes the above proportion closest to $1-\alpha$; denote this value as $\hat{p}_0$.
7. Now implement Algorithm 3 using the content value $\hat{p}_0$.

2.1. Models for the Mean Dissolution Profile

So far we have developed tolerance limits without assuming any structure for the mean dissolutions. The model-dependent methods investigated in the literature on dissolution profile comparisons assume models on the population mean dissolution profiles as an increasing function of time; in particular, the Weibull model is commonly used, [8, 9, 10] and the model is given by

$$\mu_R^t = 1 - \exp(-\alpha_R \times t^{\beta_R}),\ \mu_T^t = 1 - \exp(-\alpha_T \times t^{\beta_T}), t = 1, ..., K \tag{4}$$

where we write $\mu_R = (\mu_R^1, \mu_R^2, ..., \mu_R^K)'$ and $\mu_T = (\mu_T^1, \mu_T^2, ..., \mu_T^K)'$, and α_R, α_T, β_R and β_T are unknown parameters. The Weibull model could be incorporated into our parametric set up, where the unknown parameters (α_R, α_T, β_R, β_T, Σ_R, and Σ_T) can be estimated by maximum likelihood. The parametric bootstrap can then be implemented in a straightforward manner, under the multivariate normality assumption.

A constant mean difference model is sometimes assumed for the mean vectors, which assumes that the difference between the mean profiles μ_R and μ_T is a constant across time. That is

$$\mu_R - \mu_T = \delta \mathbf{1}_K, \tag{5}$$

where δ is an unknown scalar parameter, and $\mathbf{1}_K$ is a $K \times 1$ vector of ones. Under multivariate normality, the dissolution profile vectors are now distributed as

$$Y_R \sim N(\mu + \delta\mathbf{1}_K, \Sigma_R),\ Y_T \sim N(\mu, \Sigma_T),$$

where $\mu = \mu_T$. MLEs of the parameters can be numerically obtained and the parametric bootstrap can be implemented for computing lower tolerance limits. Assuming that $\Sigma_T = \Sigma_R$, [18] discussed testing interval hypothesis concerning the parameter δ under the above constant mean difference model. It should be noted that when $\Sigma_T = \Sigma_R$, it is possible to obtain explicit expressions for the maximum likelihood estimators of the parameters.

2.2. *Dissolution Comparisons Using the Factors f_2, g_1 and f_1*

Note that the factor f_2 defined in (1) is in terms of the difference between the sample means $\bar{Y}_R - \bar{Y}_T$, whereas the factor g_2 proposed in (2) is in terms of the difference $Y_R - Y_T$ between the individual dissolution profiles. Since f_2 appears to be a standard criterion for deciding the similarity between dissolution profiles, it may be of interest to compute a lower tolerance limit for f_2. This is equivalent to computing an upper tolerance limit for the distribution of $\frac{1}{K}\sum_{t=1}^{K}(\bar{Y}_R^t - \bar{Y}_T^t)^2$. This can be accomplished using a parametric bootstrap under the multivariate normality assumption, or it can be done non-parametrically. The algorithms given earlier can be modified in a straightforward manner to compute the required tolerance limits. In particular, under multivariate normality, we will be using the distributions

$$\bar{Y}_R \sim N\left(\mu_R, \frac{1}{n_R}\Sigma_R\right) \text{ and } \bar{Y}_T \sim N\left(\mu_T, \frac{1}{n_T}\Sigma_T\right). \tag{6}$$

In case some researchers prefer doing dissolution comparisons using difference factors g_1 and f_1 defined in (1) and (2), our proposed dissolution comparison approach for g_2 can be adopted to these criteria as well. Recall that the difference factor g_1 is an absolute scaled difference between the dissolution profiles for the reference drug and the test drug. An upper tolerance limit for g_1 is of obvious interest; if the upper tolerance limit is small (according to some regulatory guideline), we can conclude that the two dissolutions are similar with respect to the factor g_1. The parametric and nonparametric bootstrap approaches we have developed earlier can be applied for computing an upper (or lower) tolerance limit for any scalar valued function of the random variables Y_R and Y_T (or, the sample means $\bar{Y}_R$ and $\bar{Y}_T$). However, a difficulty while trying to implement the bootstrap calibration is that an estimate of the pth percentile of g_1 is not available, even as an approximation. Thus, this percentile has to be numerically obtained based on bootstrap samples, as noted while implementing the non-parametric bootstrap in Algorithm 4. Once such an estimate of the pth percentile is available, the bootstrap method (along with the bootstrap calibration) can be adapted for computing an upper tolerance limit for g_1, either parametrically (under multivariate normality) or non-parametrically. The same can also be done for the difference factor f_1.

An observation that may be of practical interest is that our methodology can be used to compute an upper confidence limit of the median of each of the random variables g_1, g_2, f_1 and f_2; simply choose the content p to be 0.50.

3. Simulation results

In order to evaluate the performance of our proposed approach, we shall now report numerical results on the estimated coverage probabilities associated with our tolerance limits. In our simulations, we have chosen content $p = 0.9$ and confidence level $1 - \alpha = 0.95$. The coverage probability calculation is quite time consuming since bootstrap calibration is also employed. Thus we have used only 1000 simulation runs in our estimation of the coverage probabilities.

For the simulations, we have chosen two sets of values for the population means and covariance matrices: those obtained from the data in [19], and from the data in [7]. The relevant data in [19] are given in Table 1 of their paper; the sample sizes are $n_R = 36$ and $n_T = 12$, and the number of time points for the data is seven, taken as 1, 2, 3, 4, 6, 8, 10 (here the time is in hours). The data set is reproduced in the online supporting material, along with the means and covariance matrices computed from the data. These computed values are used as the true parameter values for the purpose of simulation. In the first simulation set up, we shall assume the cases of both equal and unequal Σ_T and Σ_R. Also, we varied the sample sizes

Table 1. Coverage of one-sided tolerance limits based on Algorithm 2 and Algorithm 4 using 1000 simulation runs for the parameter choices given in Appendix A of the online supporting material when $\Sigma_T = \Sigma_R$, with $B = B_2 = 1000$; EqDiff denotes the equal difference model, Weibull denotes the Weibull model, PB denotes parametric bootstrap and NPB denotes non-parametric bootstrap.

Target Variable	B_1	Boot-strap	Mean Model	(n_R, n_T) (12, 12)	(36, 12)	(36, 36)
g_1	1000	PB	None	0.946	0.940	0.947
g_2	1000	PB	None	0.947	0.941	0.950
f_1	1000	PB	None	0.942	0.937	0.955
f_2	1000	PB	None	0.948	0.945	0.951
g_1	500	PB	EqDiff	0.958	0.959	0.963
g_2	500	PB	EqDiff	0.962	0.961	0.964
f_1	500	PB	EqDiff	0.955	0.958	0.958
f_2	500	PB	EqDiff	0.960	0.959	0.961
g_1	1000	PB	Weibull	0.960	0.961	0.961
g_2	1000	PB	Weibull	0.942	0.942	0.960
f_1	1000	PB	Weibull	0.956	0.958	0.960
f_2	1000	PB	Weibull	0.957	0.958	0.965
g_1	1000	NPB	None	0.942	0.940	0.955
g_2	1000	NPB	None	0.943	0.943	0.948
f_1	1000	NPB	None	0.948	0.942	0.951
f_2	1000	NPB	None	0.943	0.943	0.945

Table 2. Coverage of one-sided tolerance limits based on Algorithm 2 and Algorithm 4 using 1000 simulation runs for the parameter choices given in Appendix A of the online supporting material when $\Sigma_T \neq \Sigma_R$, with $B = B_2 = 1000$; EqDiff denotes the equal difference model, Weibull denotes the Weibull model, PB denotes parametric bootstrap and NPB denotes non-parametric bootstrap.

Target Variable	B_1	Boot-strap	Mean Model	(n_R, n_T) (12, 12)	(36, 12)	(36, 36)
g_1	1000	PB	None	0.935	0.945	0.937
g_2	1000	PB	None	0.937	0.946	0.938
f_1	1000	PB	None	0.943	0.940	0.939
f_2	1000	PB	None	0.945	0.945	0.946
g_1	500	PB	EqDiff	0.963	0.962	0.965
g_2	500	PB	EqDiff	0.959	0.960	0.960
f_1	500	PB	EqDiff	0.962	0.959	0.964
f_2	500	PB	EqDiff	0.964	0.963	0.965
g_1	1000	PB	Weibull	0.959	0.962	0.961
g_2	1000	PB	Weibull	0.961	0.963	0.964
f_1	1000	PB	Weibull	0.958	0.957	0.962
f_2	1000	PB	Weibull	0.957	0.960	0.965
g_1	1000	NPB	None	0.941	0.941	0.944
g_2	1000	NPB	None	0.942	0.944	0.943
f_1	1000	NPB	None	0.936	0.938	0.940
f_2	1000	NPB	None	0.942	0.945	0.947

n_R and n_T between 36 and 12. Unstructured and structured means were both considered; these are specified in Appendix A of the online supporting material. The estimated coverage probabilities for various scenarios are given in Table 1 and Table 2.

Our second choice of the parameter values is obtained from the data in [7]. Here the number of time points is 8, taken as 1, 2, 3, 4, 5, 6, 7, 8. The data, along with the means and covariance matrices computed from the data are given in Appendix

Table 3. Coverage of one-sided tolerance limits based on Algorithm 2 and Algorithm 4 using 1000 simulation runs for the parameter choices given in Appendix B of the online supporting material when Σ_T and Σ_R are unequal, with B = B_2 = 1000; EqDiff denotes the equal difference model, Weibull denotes the Weibull model, PB denotes parametric bootstrap and NPB denotes non-parametric bootstrap.

Target Variable	B_1	Boot-strap	Mean Model	(n_R, n_T) (12, 12)	(36, 12)	(36, 36)
g_1	1000	PB	None	0.944	0.940	0.944
g_2	1000	PB	None	0.951	0.937	0.947
f_1	1000	PB	None	0.945	0.939	0.943
f_2	1000	PB	None	0.950	0.940	0.948
g_1	500	PB	EqDiff	0.960	0.959	0.962
g_2	500	PB	EqDiff	0.957	0.956	0.957
f_1	500	PB	EqDiff	0.959	0.958	0.961
f_2	500	PB	EqDiff	0.960	0.958	0.959
g_1	1000	PB	Weibull	0.959	0.958	0.962
g_2	1000	PB	Weibull	0.961	0.955	0.960
f_1	1000	PB	Weibull	0.960	0.959	0.960
f_2	1000	PB	Weibull	0.960	0.957	0.962
g_1	1000	NPB	None	0.944	0.945	0.949
g_2	1000	NPB	None	0.950	0.947	0.952
f_1	1000	NPB	None	0.946	0.937	0.945
f_2	1000	NPB	None	0.942	0.945	0.950

B of the online supporting material. Again, we varied the sample sizes n_R and n_T between 36 and 12. Unstructured and structured means were both considered, as specified in Appendix B of the online supporting material. The coverage probabilities are reported in Table 3.

The numerical results in Table 1– Table 3 indicate that our proposed methodology does result in accurate tolerance limits for dissolution comparisons, simultaneously accounting for the sampling variability and the correlations across multiple time points. All the coverage probabilities are close to the assumed nominal level of 0.95, even when $n_T = n_R = 12$. In particular, we note that if a model is not available for the mean profile, and if the normality assumption is not tenable, our non-parametric bootstrap approach, without the assumption of a model for the mean, still provides an accurate upper tolerance limit.

4. Two Examples

Two real examples are presented here to illustrate the application of our tolerance interval methodology. The data sets used are taken from the articles by [19] and [7], and are reproduced in Appendix A and Appendix B, respectively, of the online supporting material.

4.1. Example 1: The Tsong et al. (1997) data set

The data set includes four dissolution batches in total. The first three batches are the reference batches, consisting of 12 tablets per batch; thus $n_R = 36$. A fourth batch with 12 tablets forms the test batch; thus $n_T = 12$. Furthermore, there are $K = 7$ times points: 1, 2, 3, 4, 6, 8, and 12 hours. We shall apply our methods to calculate upper tolerance limits for the distributions of $\frac{1}{K}\sum_{t=1}^{K}(Y_R^t - Y_T^t)^2$, $\frac{1}{K}\sum_{t=1}^{K}(\bar{Y}_R^t - \bar{Y}_T^t)^2$, $\frac{\sum_{t=1}^{K}|Y_R^t - Y_T^t|}{\sum_{t=1}^{K} Y_R^t}$ and $\frac{\sum_{t=1}^{K}|\bar{Y}_R^t - \bar{Y}_T^t|}{\sum_{t=1}^{K} \bar{Y}_R^t}$, leading to lower tolerance limits for g_2 and f_2 and upper tolerance limits for g_1 and f_1. The content and confidence level are chosen to be

Table 4. Lower tolerance limits for the distributions of g_2 and f_2 and upper tolerance limits for the distributions of g_1 and f_1 for Example 1; content $p = 0.90$ and confidence level $1 - \alpha = 0.95$.

Target variable	Parametric: No mean structure	Parametric: Weibull structure	Non-parametric	Values of f_1 and f_2
g_2	48.119	45.805	45.112	
f_2	54.567	56.121	59.083	$f_2 = 64.111$
g_1	17.197	17.616	19.652	
f_1	10.613	10.987	7.213	$f_1 = 6.479$

$p = 0.90$ and $1 - \alpha = 0.95$, respectively. We shall consider model-independent as well as model-dependent cases, where a Weibull model is assumed for the mean profile in the model dependent case, as done in [19]. Here we shall implement both parametric and nonparametric bootstrap methods. We used $B = 1000$ bootstrap samples for computing the upper tolerance limit, after estimating the content value by bootstrap calibration using $B_1 = 1000$ and $B_2 = 1000$ bootstrap samples (for implementing Algorithm 2 and Algorithm 4). The lower tolerance limits for g_2 and f_2 and upper tolerance limits for g_1 and f_1 under various scenarios are given in Table 4. One should certainly expect the lower tolerance limit for g_2 to be less than that of f_2, and the upper tolerance limit for g_1 to be higher than that for f_1 since $Var(\bar{Y}_R) < Var(Y_R)$ and $Var(\bar{Y}_T) < Var(Y_T)$; we note this to be the case in Table 4. The table also gives the numerical values of f_2 and f_1; these numerical values certainly meet the FDA specifications for concluding profile similarity, namely $f_2 > 50$ and $f_1 < 15$.

From the results in Table 4 we note that the lower tolerance limit for f_2 is greater than 50, when the methodology is implemented parametrically or non-parametrically. Thus we conclude with 95% confidence that 90% or more of the f_2 distribution is above 50. Also, the upper tolerance limits for f_1 under both parametric and non-parametric approaches are less than 15; we thus conclude with 95% confidence that 90% or more of the f_1 distribution is below 15. In another words, dissolution profile similarity can be concluded if we use the criteria f_2 and f_1. However, if we look at the corresponding tolerance limits for g_2 and g_1, and use the same threshold values 50 and 15, respectively, we cannot conclude profile similarity. This conclusion holds under the parametric and non-parametric scenarios.

4.2. Example 2: The Ocana (2009) data set

This data set consists of dissolution profiles coming from a batch of Metoclopramide Hydrochloride tablets with tensioactive, and a batch of tablets without tensioactive. Each batch includes 12 dissolution profiles and each profile consists of observations across 8 time points. In other words, $n_R = n_T = 12$ and $K = 8$ $(t = 1, 2, 3, 4, 5, 6, 7, 8)$. The data set is available in Table 3 of [7], and are reproduced in Appendix B of the suppoirting material. We shall continue to use $p = 0.90$ and $1 - \alpha = 0.95$ as the content and confidence level, respectively. Again, $B = 1000$ bootstrap samples were used to calculate the upper tolerance limit after performing bootstrap calibration using Algorithm 2 and Algorithm 4 with $B_1 = 1000$ and $B_2 = 1000$. The tolerance limits for the various scenarios are given in Table 5, along with the numerical values of f_2 and f_1. We note from Table 5 that the f_1 and f_2 values (given in the last column of the table) do meet the FDA requirements, so that we can conclude similarity of the two dissolution profiles. However, this is no longer the case if we use the tolerance limits; the tolerance limits for g_2 and g_1 are far removed from the FDA specified thresholds. The conclusion is the same based on the tolerance limits for f_1 and f_2, except for the limits obtained non-parametrically. The non-parametric lower tolerance limit for f_2 just crosses the 50 threshold, and the non-parametric upper tolerance tolerance limit for f_1 is less than 15%. The overall conclusion that emerges from Table 5 is that the similarity of the two dissolution profiles cannot be concluded.

A close examination of the original data in [7] reveals the following. Among the eight time points at which dissolution data have been obtained, significant differences exist among the dissolution values at the first three time points, and considerable similarity is noticeable among the last 5 time points; see Appendix B of the online supporting material. Perhaps the dissimilarity that is so noticeable among the first three time points accounts for the lack of dissolution profile

Table 5. Lower tolerance limits for the distributions of g_2 and f_2 and upper tolerance limits for the distributions of g_1 and f_1 for Example 2; content $p = 0.90$ and confidence level $1 - \alpha = 0.95$.

Target variable	Parametric: No mean structure	Parametric: Weibull structure	Non-parametric	Values of f_1 and f_2
g_2	41.200	39.077	41.416	
f_2	46.472	44.085	50.037	$f_2 = 51.704$
g_1	24.629	29.882	26.228	
f_1	18.098	22.770	13.196	$f_1 = 12.635$

Table 6. Example2: Lower tolerance limits for g_2 using data on all time points and on last five time points ($p = 0.80, 0.90$, $1 - \alpha = 0.95$), compared with FDA defined f_2.

	NoMeanstruc		Weibull		nonpara		percentile	f_2
	$p = 0.9$	$p = 0.8$	0.9	0.8	0.9	0.8	0.9	
All time points	41.20	43.53	39.08	41.43	41.42	43.79	41.21	51.71
Last five time pts	46.10	50.66	46.35	50.12	48.99	53.39	46.33	68.19

similarity that emerges from the numerical results reported in Table 5. In view of this, it may be of interest to check for dissolution profile similarity, concentrating only on the last 5 time points. Such an evaluation of dissolution profile comparison using the data of last few time points could be of interest for certain drug products. Table 6 gives the lower tolerance limits for g_2 using the last five time points and using all the time points using the content level $p = 0.80$ and 0.90. Table 6 shows that we can conclude similarity of the two dissolution profiles based on f_2 values shown in the last column using all time points or last five time points. Furthermore, using the lower tolerance limit for g_2, we can also conclude similarity based on the last 5 time points if the content is chosen to be $p = 0.80$, but not for $p = 0.90$. Subject-specific knowledge on the tablets and the treated disease will perhaps be helpful to determine whether using the last few time points and a lower content level are clinically meaningful or not.

5. Discussion

Dissolution testing is a critical component in the development of pharmaceutical dosage forms, since it can serve as a substitute for in vivo studies. Valid statistical analysis of the relevant data is clearly a crucial part of dissolution profile comparisons. The criteria that are currently in use, based on the factors f_1 and f_2, appear ad hoc, and lack statistical rigor, even though they are widely used and recommended by the FDA. A number of alternative criteria have been suggested in the literature, and new criteria continue to be introduced in the recent literature; see [20]. Rather than introducing new criteria, the present work takes up the existing criteria and develops statistically rigorous procedures based on them, using the concept of tolerance limits. We feel that tolerance limits are the right quantities in this context, since they are meant to provide bounds on the entire population, and can thus be used to draw conclusions regarding the similarity of the population dissolution profiles. Even though the computation of the required tolerance limits can be demanding, we could circumvent some of the difficulties using a non-parametric tolerance limit computation, coupled with a bootstrap calibration. Since an upper tolerance limit is an upper confidence limit for a population percentile, and since an approximation is available for the percentile (as noted in the Appendix), a natural question that comes up is whether the bootstrap can be directly applied to the approximate percentile. We did try a percentile bootstrap method for computing an upper confidence limit for the approximate percentile, but the coverage probability was not satisfactory. A bootstrap calibration was then tried, and this did improve the accuracy of the coverage probability. However, the resulting computation turned out to be more time consuming than the methodology we are proposing in this paper. The computations

outlined in Algorithm 2 and Algorithm 4 may look cumbersome, but are straightforward to carry out using the R codes available as supplementary material online.

It could be argued that the FDA recommended thresholds applied to upper tolerance limits result in requirements that are too strong for concluding dissolution profile similarity. If so, one could relax the requirement by simply lowering the value of the content p. In particular, by choosing $p = 0.50$, our methodology will provide upper confidence limits for the median of the relevant random variable. Clearly, regulatory input is necessary before deciding a value of p.

We want to conclude by highlighting a few other aspects of the dissolution testing problem and the available literature, from the perspective of our work. Most researchers assume a common covariance matrix for the test and reference dissolution profile distributions. This is especially the case for the problem of comparing the mean dissolution profiles. This assumption is likely to be unrealistic in applications. A formal test for the equality of the covariance matrices resulted in rejection of the equality hypothesis for both of the data sets used in this paper. In our work, we have not made this assumption. If the common covariance matrix assumption does hold, our methodologies can be modified in a straightforward manner so as to reflect this assumption. A second concern is the multivariate normality assumption. It should be clear that the dissolution profile of each tablet (reference as well as test) is an increasing function of time. In other words, the random vector for which the multivariate normality is assumed, always has its components ordered from smallest to largest. In view of this, the multivariate normality assumption is simply not appropriate. Thus the non-parametric tolerance limit that we have developed should be of considerable interest.

Note that for the data in the first example, the data are available in batches, which suggests the possibility of using a model where random batch effects are present. We did consider such a model for this example, for the data on the reference drug, but the batch effects turned out to be highly insignificant. However, the tolerance limit problem can certainly be addressed in the presence of random batch effects. This is currently under investigation.

6. Appendix: An Approximation for the Percentile of $X = \frac{1}{K}\sum_{t=1}^{K}(Y_R^t - Y_T^t)^2$

Here is a brief description of the method due to [16], for approximating the p^{th} percentile of $X = \frac{1}{K}\sum_{t=1}^{K}(Y_R^t - Y_T^t)^2$. In order to explain the approximation, let

$$Q = (Y_R - Y_T)'(Y_R - Y_T),$$

so that $X = \frac{Q}{K}$. Clearly, $Y_R - Y_T \sim N(\mu_R - \mu_T, \Sigma_R + \Sigma_T)$. For notational convenience, let $\mu = \mu_R - \mu_T$ and $\Sigma = \Sigma_R + \Sigma_T$. The approximation to the cdf of Q is obtained by noting that Q is a linear combination of independent non-central chisquare random variables, with non-centrality parameters depending on μ and Σ, and the dfs depending on the multiplicity of the eigenvalues of Σ. If the eigenvalues of Σ are distinct, then each non-central chisquare in the linear combination has one df, and the number of terms in the linear combination is K (the dimension of Y_R, as well as that of Y_T). In order to give the approximation, let

$$c_k = \text{trace}(\Sigma^k) + k\mu'\Sigma^{k-1}\mu \qquad k = 1, 2, 3, 4.$$

Then [16] provide the approximation

$$P(Q \le u) \simeq P(\chi_l^2(\delta) \le u'),$$

where $u' = \left[\frac{u - c_1}{\sqrt{2}c_2} \times \sqrt{2(l + 2\delta)}\right] + l + \delta$, where l and δ are given by (i) and (ii) below, and depend on $s_1 = \frac{c_3}{c_2^{3/2}}$ and $s_2 = \frac{c_4}{c_2^2}$.

(i) if $s_1^2 > s_2$,

$$\delta = \frac{s_1}{(s_1 - \sqrt{s_1^2 - s_2})^3} - \frac{1}{(s_1 - \sqrt{s_1^2 - s_2})^2}$$

$$\text{and } l = \frac{1}{(s_1 - \sqrt{s_1^2 - s_2})^2} - 2\delta;$$

(ii) if $s_1^2 \leq s_2$,

$$\delta = 0, \text{ and } l = c_2^3/c_3^2.$$

Furthermore, the p^{th} percentile of Q can be approximated as $Q_p = \left((\chi^2_{l,\delta;p} - l - \delta) \times \sqrt{\frac{c_2}{l+2\delta}}\right) + c_1$, where $\chi^2_{l,\delta;p}$ is the p^{th} percentile of the non-central chisquare distribution $\chi^2_l(\delta)$. For $X = \frac{Q}{K}$, the p^{th} percentile, say X_p, can thus be approximated as

$$X_p = \frac{1}{K}\left[\left((\chi^2_{l,\delta;p} - l - \delta) \times \sqrt{\frac{c_2}{l+2\delta}}\right) + c_1\right]. \tag{7}$$

Acknowledgements

We want to thank Dr. Yi Tsong for several constructive suggestions. The first and second authors are supported on grant R01OH003628 from the National Institute of Occupational Safety and Health (NIOSH). The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS–0821258 and CNS–1228778) and the SCREMS program (grant no. DMS–0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC). See `www.umbc.edu/hpcf` for more information on HPCF and the projects using its resources.

References

1. FDA. *Guidance for Industry: Dissolution Testing of Immediate Release Solid Oral Dosage Forms*. Rockville, MD 1997.
2. FDA. *Guidance for Industry: Extended Release Oral Dosage Forms: Development, Evaluation, and Application of In Vitro/In Vivo Correlations*. Rockville, MD 1997.
3. FDA. *Guidance for Industry: Modified Release Solid Oral Dosage Forms Scale-up and Post-Approval Changes: Chemistry, Manufacturing, and Controls; In Vitro Dissolution Testing, and In Vivo Bioequivalence Documentation (SUPAC-MR)*. Rockville, MD 1997.
4. FDA. *Guidance for Industry: Nonsterile Semisolid Dosage Forms Scale-up and Post-Approval Changes: Chemistry, Manufacturing, and Controls; In Vitro Release Testing and In Vivo Bioequivalence Documentation (SUPAC-SS)*. Rockville, MD 1997.
5. Moore JW, Flanner HH. Mathematical comparison of dissolution profiles. *Pharmaceutical Technology* 1996; **20**:64–74.
6. O'Hara T, Dunne A, Butler J, Devane J. A review of methods used to compare dissolution profile data. *Pharmaceutical Science and Technology Today* 1998; **1**(5):214–223.
7. Ocana J, Frutos G, Sanchez P. Using the similarity factor f_2 in practice: A critical revision and suggestions for its standard error estimation. *Chemometrics and Intelligent Laboratory Systems* 2009; **99**:49–56.
8. Polli JE, Rekhi GS, Augsburger LL, Shah VP. Methods to compare dissolution profiles and a rationale for wide dissolution specifications for metoprolol tartrate tablets. *Journal of Pharmaceutical Sciences* 1997; **86**(6):690–700.
9. Sathe PM, Tsong Y, Shah VP. In vitro dissolution profile comparison: Statistics and analysis, model dependent approach. *Pharmaceutical Research* 1996; **13**(12):1799–1803.
10. Yuksel N, Kanik AE, Baykara T. Comparison of in vitro dissolution profiles by anova-based, model-dependent, and -independent methods. *Biometrika* 1983; **70**(1):41–55.
11. Shah VP, Tsong Y, Sathe P, Liu JP. In vitro dissolution profile comparison - statistics and analysis of the similarity factor, f_2. *Pharmaceutical Research* 1998; **15**(6):889–896.
12. Ma MC, Lin RP, Liu JP. Statistical evaluations of dissolution similarity. *Statistica Sinica* 1999; **9**:1011–1027.

13. Ma MC, Wang BB, Liu JP, Tsong Y. Assessment of similarity between dissolution profiles. *Journal of Biopharmaceutical Statistics* 2000; **10**(2):229–249.
14. Liu JP, Ma MC, Chow SC. Statistical evaluation of similarity factor f_2 as a criterion for assessment of similarity between dissolution profiles. *Drug Information Journal* 1997; **31**:1255–1271.
15. Krishnamoorthy K, Mathew T. *Statistical Tolerance Regions: Theory, Applications, and Computation*. John Wiley and Sons, 2009.
16. Liu H, Tang Y, Zhang H. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics and Data Analyccsis* 2009; **53**:853–856.
17. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
18. Saranadasa H, Krishnamoorthy K. A multivariate test for similarity of two dissolution profiles. *Journal of Biopharmaceutical Statistics* 2005; **15**:265–278.
19. Tsong Y, Hammerstrom T, Chen JJ. Multipoint dissolution specification and acceptance sampling rule based on profile modeling and principal component analysis. *Journal of Biopharmaceutical Statistics* 1997; **7**(3):423–439.
20. Gomez-Mantilla JD, Casabo V, Schaefer U, Lehr CM. Permutation test (pt) and tolerated difference test (tdt): Two new, robust and powerful nonparametric tests for statistical comparison of dissolution profiles. *International Journal of Pharmaceutics* 2013; **441**:458–467.

Supporting information

Additional supporting information may be found in the online version of this article at the publishers web site.

