Pharmaceutical Research, Vol. 15, No. 6, 1998

In Vitro Dissolution Profile Comparison—Statistics and Analysis of the Similarity Factor, f_2

Vinod P. Shah,[1,4] Yi Tsong,[2] Pradeep Sathe,[1] and Jen-Pei Liu[3]

Received December 4, 1997; accepted February 26, 1998

Purpose. To describe the properties of the similarity factor (f_2) as a measure for assessing the similarity of two dissolution profiles. Discuss the statistical properties of the estimate based on sample means.

Methods. The f_2 metrics and the decision rule is evaluated using examples of dissolution profiles. The confidence interval is calculated using bootstrapping method. The bias of the estimate using sample mean dissolution is evaluated.

Results. 1. f_2 values were found to be sensitive to number of sample points, after the dissolution plateau has been reached. 2. The statistical evaluation of f_2 could be made using 90% confidence interval approach. 3. The statistical distribution of f_2 metrics could be simulated using 'Bootstrap' method. A relatively robust distribution could be obtained after more than 500 'Bootstraps'. 4. A statistical 'bias correction' was found to reduce the bias.

Conclusions. The similarity factor f_2 is a simple measure for the comparison of two dissolution profiles. But the commonly used similarity factor estimate $\hat{f}_2$ is a biased and conservative estimate of f_2. The bootstrap approach is a useful tool to simulate the confidence interval.

KEY WORDS: dissolution; similarity factor; estimation bias; bootstrap confidence interval.

INTRODUCTION

For immediate release solid oral drug products, a single time-point dissolution specification has been routinely employed as a quality control release test. In general, a single point dissolution test does not characterize the dosage form completely, and therefore the dissolution profile and dissolution profile comparison is recommended in recently released guidances by the Agency (1–4). For the post-approval changes such as (i) scale-up, (ii) manufacturing site, (iii) component and composition, (iv) equipment and process changes, a comparison of dissolution profiles between pre-change and post-change products is recommended in SUPAC-IR guidance (1) as it provides a more precise measurement of product similarity using dissolution characteristics. Dissolution profiles may be considered similar by virtue of (i) overall profile similarity and (ii) similarity at every dissolution sample time point. The dissolution profile comparison can be carried out using model independent or model dependent methods.

In the last decade, several methods for the comparison of dissolution profiles were proposed in the literature (5–9). However, a major problem has been the quantification of the comparison of dissolution profile. Shah et al. proposed a multivariate analysis of variance method to test for the difference between two dissolution profiles (5). Chow et al. proposed dissolution difference measurement and similarity testing based on parameters after fitting a one-degree autoregression time series model (6). Sathe et al. proposed dissolution difference measurement and similarity testing based on parameters of the profiles after fitting a selected mathematical model (7). Tsong et al. proposed dissolution difference measurement and similarity testing based on multivariate 'Mahalanobis' distance between two dissolution data sets (8). However, the statistical methods proposed in most of these examples involved the complicated estimation of covariance matrix.

Recently, Moore and Flanner proposed a simple model independent approach using mathematical indices to define difference factor, f_1, and similarity factor, f_2, to compare dissolution profiles (9). The f_1 and f_2 factors are derived from Minkowski difference (average absolute differences) and mean-squared difference respectively. The similarity factor f_2 and a similarity testing criteria based on f_2 were therefore recommended for dissolution profile comparison in the FDA's Guidances for Industry (1–4). The simplicity of similarity factor generated considerable interest. Subsequently, examples of the application of f_2 appeared in the literature (10–12), and some statistical properties of f_2 were also delineated in two papers (12,13).

The purpose of this work is to (i) describe f_2 as a population measure for assessing the similarity of two dissolution profiles (ii) describe how a similarity criteria for f_2 is defined for the two dissolution profiles (iii) discuss the statistical properties of $\hat{f}_2$, an estimate of population f_2 based on sample means, (iv) discuss the estimation of the confidence interval of f_2 based on $\hat{f}_2$ and calculation of the bias of $\hat{f}_2$, and (v) discuss the corresponding hypotheses for similarity testing based on f_2 and $\hat{f}_2$. These discussions will provide rational steps for the application of similarity factor f_2 in dissolution profile comparison.

SIMILARITY FACTOR

A. Theoretical Considerations

The profile comparison in general refers to the comparison of two dissolution profiles between (i) a reference batch and a test batch (ii) a pre-change batch and a post-change batch, and (iii) different strengths of products for biowaivers as discussed in various guidances. The principles can be applied at anytime when a profile comparison is needed.

To illustrate the applications of similarity factor, f_2, consider the dissolution profiles of the two batches generated using P number of sample points. For comparison of the dissolution profiles of two batches, the dissolution measurements should be made under the same test conditions and the dissolution time points for both the profiles should be the same, e.g., for

---

[1] Office of Pharmaceutical Science, Center for Drug Evaluation and Research, Food and Drug Administration, HFD-350,1451 Rockville Pike, WOC 2, Rockville, Maryland 20852.

[2] Division of Biometrics III, Office of Epidemiology and Biometrics, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland 20857.

[3] Department of Statistics, National Cheng-Kung University, Tainan, Taiwan, ROC.

[4] To whom correspondence should be addressed. (e-mail: Shahvi@CDER.FDA.GOV)

DISCLAIMER: The manuscript represents the personal opinions of the authors and does not necessarily represent the views or policies of FDA.

immediate release products, 15, 30, 45, 60 minutes and for controlled release products, 1, 3, 5 and 8 hours. Let (μ_{r1}, μ_{r2}, . . . , μ_{rP}) represent the dissolution measurements at P time points on the reference profile and (μ_{t1}, μ_{t2}, . . . , μ_{tP}) be the corresponding P measurements on the test profile. The distances between the two profiles at these P time points are ($|\mu_{r1} - \mu_{t1}|$, $|\mu_{r2} - \mu_{t2}|$, . . . , $|\mu_{rP} - \mu_{tP}|$). The distances at the P time points may be combined into one measure, by either the sum, $D_1 = |\mu_{r1} - \mu_{t1}| + |\mu_{r2} - \mu_{t2}| + \ldots + |\mu_{rP} - \mu_{tP}|$, or the square root of the sum of squares, $D_2 = \sqrt{[(\mu_{r1} - \mu_{t1})^2 + (\mu_{r2} - \mu_{t2})^2 + \ldots + (\mu_{rP} - \mu_{tP})^2]}$.

In 1996, Moore and Flanner proposed measurements of relative distance and similarity of two dissolution profiles as functions of D_1 and D_2, as follows:

$$f_1 = \left\{\left[\sum_{i=1}^{P} |\mu_{ti} - \mu_{ri}|\right] \Big/ \left[\sum_{i=1}^{P} \mu_{ri}\right]\right\} \bullet 100$$

and

$$f_2 = 50 \bullet \log\left\{\left[1 + (1/P) \sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2\right]^{-1/2} \bullet 100\right\}$$

$$= 50 \bullet \log\{[1 + (1/P)D_2^2]^{-1/2} \bullet 100\} \qquad (1)$$

where log is the logarithm based on 10. Note that f_1 reflects the cumulative difference between the two curves at all time points, and is a measure of relative error between the two curves. Conceptually, f_1 which is a function of the average absolute difference between the two dissolution curves could be referred as a 'difference' factor. On the other hand, f_2 metric is a function of the reciprocal of mean square-root transform of the sum of square distances at all points. Conceptually, f_2 which is a measure of the similarity in the percent dissolution between two curves, could be referred as a 'similarity' factor. When the two profiles are identical, $f_2 = 50 \bullet \log(100) = 100$, and when the dissolution of one batch is complete before the other begins, $f_2 = 50 \bullet \log\{[1 + (1/P)\Sigma_{i=1}^{P}(100)^2]^{-1/2} \bullet 100\} = -.001$, which can be rounded to 0. Thus the value of f_2 ranges between 0 to 100 with a higher f_2 value indicating more similarity between the two profiles.

In a real life situation, due to the batch-to-batch variation in dissolution profiles, it is not expected to have f_2 value be anywhere near 100 even when the two dissolution curves are generated from the same batch of tablets (or capsules). A test batch is therefore accepted as 'similar' to a reference batch if the dissolution profile difference between the two batches is no more than the dissolution profile difference between the two reference batches. Empirically, the experience in dissolution data analysis leads one to believe that an average difference of no more than 10% at any sample time point, of the batches of the same formulation may be acceptable. When this 10% average difference is substituted in the Equation 1, f_2 becomes:

$$f_{2,10} = 50 \bullet \log\left\{\left[1 + (1/P) \sum_{i=1}^{P} |10|^2\right]^{-1/2} \bullet 100\right\}$$

$$= 50 \bullet \log\{[101]^{-1/2} \bullet 100\}$$

$$= 50 \bullet \log(9.95037) = 49.89$$

which may be rounded to 50 for simplicity. A test batch dissolution is therefore considered similar to that of the reference batch if the f_2 value of the two true profiles is not less than 50. It is clear that once the average distance at any sample time point between any two reference batch is defined, the similarity limit based on f_2 can be defined independent to the test batch or the specific reference batch and independent to the number of sampling time points to be used in the assessment of dissolution similarity. Table 1 provides the f_2 similarity limits for different average distances at multiple time points by appropriate substitution in Equation 1.

Table 1. Average Difference Between Two Dissolution Profiles of Reference Batches

	2%	5%	10%	15%	20%
f_2 Limit	83	65	50	41	36

B. Results and Discussions

Example #1, One Reference Batch and Four Test Batches (Tables 2, 3 and Figures 1 and 2). To illustrate the concept of assessing similarity and dissolution profile comparison using

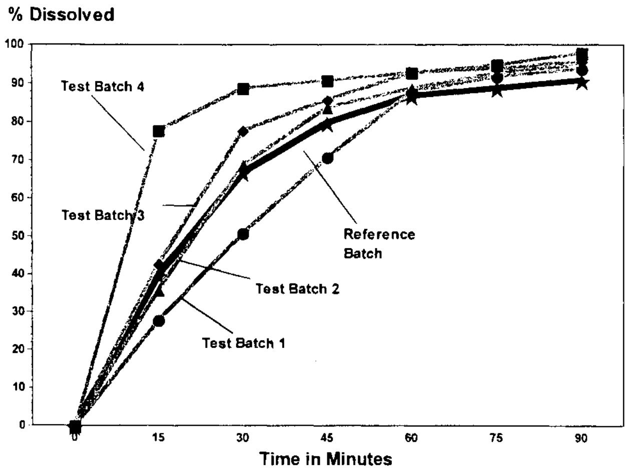

Fig. 1. Actual mean data.

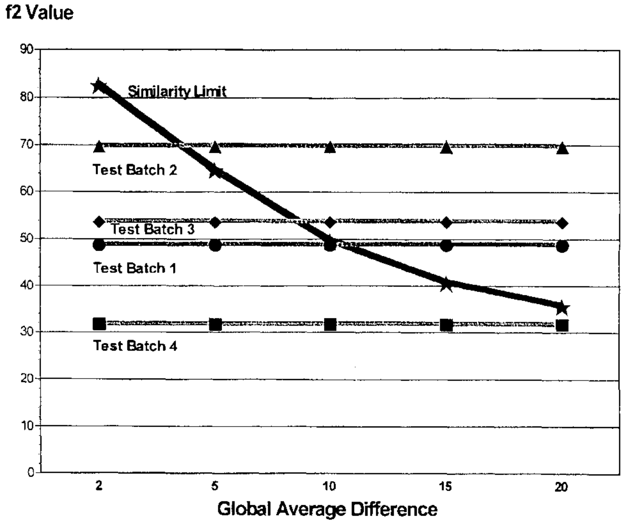

Fig. 2. Actual profile comparison with similarity limits.

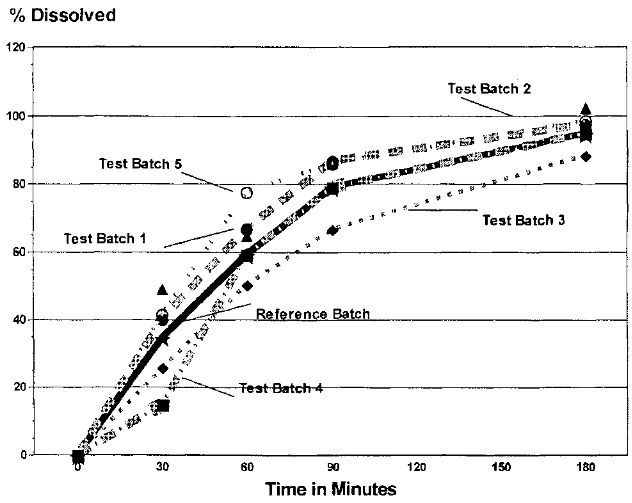

Fig. 3. Sample mean dissolution.

f_2, consider the following example. In Table 2 provides the actual cumulative dissolutions at 15, 30, 45, 60, 75 and 90 minutes of a reference batch and four test batches. The f_2 value for each of the four test batches compared to the reference batch is given in Table 3.

Based on measurements up to 60 minutes (the time when the reference product is dissolved up to 87%), it is clear that test batch #2 is similar to the reference batch with an average difference of 5% at the four time points. Test batch #3 can be claimed to be similar to the reference batch with an average difference of 10% at the four time points. Test batch #1 can only be considered to be similar to the reference batch if the average difference between any two reference batches is 15%. Test batch #4 is not similar to the reference batch even when one allows an average difference of 20% at all time points (Figure 1). Using more time points after more than 85% dissolution, will invariably increase the f_2 value leading to bias in the similarity assessment. For example, when using cumulative dissolutions up to 90 minutes, for the same four test batches, the f_2 values increases in almost all test batches (Table 3). It is therefore important to limit the number of sample points to no more than one, once any of the batch (product) reaches 85% dissolution.

Table 2. Example #1: Dissolution Profile of One Reference and Four Test Batches

	% Drug dissolved in					
Batch	15	30	45	60	75	90 minutes
Reference	40	67	80	87	89	91
Test batch #1	28	51	71	88	92	94
Test batch #2	36	69	84	89	93	95
Test batch #3	43	78	86	93	94	96
Test batch #4	78	89	91	93	95	98

Table 3

	f_2 Value for test batch			
	1	2	3	4
When calculated up to 60 minutes only	48	70	54	32
When calculated up to 90 minutes	52	71	57	36

Note: f_2 value calculated by using data presented for example #1, in Table 2.

C. Estimation of Similarity Factor

The properties illustrated in the last section are based on the f_2 of the actual (population) dissolution profiles of the reference and test batches. In practice, dissolution testing is often carried out with no more than 12 units and the dissolution profile of each batch is an estimate based on dissolutions of the 12 units. Hence $(\bar{x}_{r1.}, \bar{x}_{r2.}, \ldots, \bar{x}_{rP.})$ and $(\bar{x}_{t1.}, \bar{x}_{t2.}, \ldots, \bar{x}_{tP.})$ are used to estimate $(\mu_{r1}, \mu_{r2}, \ldots, \mu_{rP})$ and $(\mu_{t1}, \mu_{t2}, \ldots, \mu_{tP})$ respectively, where $\bar{x}_{ti.}$, $\bar{x}_{ri.}$ are the mean dissolution value of the twelve tablets measured at the i-th time point of the test and the reference batch respectively. With these estimates, f_2 is estimated as follows

$$\hat{f}_2 = 50\bullet\log\left\{\left[1 + (1/P)\sum_{i=1}^{P}(\bar{x}_{ti} - \bar{x}_{ri})^2\right]^{-1/2}\bullet 100\right\}$$

D. Confidence Interval of Similarity Factor

Because of the sampling variation for the estimate, dissolution similarity of the test and reference batches may not be assessed by direct comparison of $\hat{f}_2$ with the similarity limit, SL. The SL proposed in the guidances is 50 (1–4). Assuming the expected value of $\hat{f}_2$ equals f_2, i.e., $E(\hat{f}_2) = f_2$, for an assurance of 95% correct decision, one should compare the 90% lower confidence limit of $E(\hat{f}_2)$ with SL instead. In order to have a mathematical form of the confidence interval, one needs to derive the sampling distribution of $\hat{f}_2$. Each component of the mean vector $\bar{x}_r = (\bar{x}_{r1.}, \bar{x}_{r2.}, \ldots, \bar{x}_{rP.})$ and $\bar{x}_t = (\bar{x}_{t1.}, \bar{x}_{t2.}, \ldots, \bar{x}_{tP.})$ is a random variable with standard error $se(\bar{x}_{ki.})$, where k=r,t, and the elements in $\bar{x}_k$ are correlated. In order to have a standard (or asymptotically standard) distribution for $\hat{f}_2$, one needs to standardize $\hat{f}_2$ by its covariance matrix. If there is a known standardized form of $\hat{f}_2$, it would be a complicated function of the 'Mahalanobis' distance as described by Tsong et al. (7,8). Alternatively, the 90% confidence interval can be simulated through bootstrap method as given by Tsong et al. (14).

A bootstrap sample of f_2 can be generated by random sample with replacement twelve times from $x_{rj} = (x_{r1j.}, x_{r2j.}, \ldots, x_{rPj.})$ and $x_{tj} = (x_{t1j.}, x_{t2j.}, \ldots, x_{tPj.})$, where j=1 to 12. Let $x'_{rj} = (x_{r1j'.}, x_{r2j'.}, \ldots, x_{rPj'.})$ and $x'_{tj} = (x_{t1j.'}, x_{t2j.'}, \ldots, x_{tPj'.})$, $j'=1$ to 12, be the twelve dissolution vectors re-sampled from the 12 tablets of the test and reference batches respectively. Note that some of the vectors of dissolution values may be identical because of the replacement in the sampling. Let $\hat{f}'_2$ denote the estimated f_2 value of the bootstrap sample. Considering that M sets of sample are generated using the bootstrap method, the 90% percent confidence interval is defined by $[L\hat{f}'_2, U\hat{f}'_2]$, where $L\hat{f}'_2$ and $U\hat{f}'_2$ are the 5th and 95th percentiles of the $E\hat{f}'_2$ values. Since distribution of $\hat{f}'_2$ is skewed, an adjustment may be necessary. The adjusted confidence interval, BC_α of $E(\hat{f}_2)$ of bias correction $(\hat{f}_2'^{(\alpha 1)}, \hat{f}_2'^{(\alpha 2)})$ is defined with

$$\alpha 1 = \Phi(Z_0 + (Z_0 + Z^{(\alpha)})/[1 - a(Z_0 + Z^{(\alpha)})]$$

$$\alpha 2 = \Phi(Z_0 + (Z_0 + Z^{(1-\alpha)})/[1 - a(Z_0 + Z^{(1-\alpha)})$$

$$z_0 = \Phi^{-1}(\#(\hat{f}_2'(m) < \hat{f}_2)/M)$$

$$a = \sum (\hat{f}_{2(\cdot)} - \hat{f}_{2(i)})^3/\{6[\sum (\hat{f}_{2(\cdot)} - \hat{f}_{2(i)})^2]^{3/2}\}$$

where α is the level of type I error, $\hat{f}_{2(i)}$ is the i-th jackknife statistic, $\hat{f}_{2(\cdot)}$ is the mean of jackknife statistics, $\hat{f}_2'(m)$ is the bootstrap estimate of the m-th bootstrap sample, $\hat{f}_2$ is the original sample mean, $z^{(\alpha)}$ is the α-th percentile of standard normal distribution.

Example #2, One Reference Batch and Five Test Batches (Tables 4, 5 and Figure 2). To illustrate the application of bootstrap method in confidence interval estimation and assessment of dissolution similarity, consider the cumulative percent of dissolution at 30, 60, 90 and 180 minutes of five test batches and one reference batch with 12 tablets each as shown in Table 4. Table 4 also provides the sample means of each batch at every time point. The covariance and correlations among time points are given in Table 5. From Table 5, the correlation between two time points can be as high as 0.93 and some times, the cumulative percent dissolved at different time points may be negatively correlated. The mean dissolution values of test batch #1 differ from the reference batch by no more than 8%. Test batch #2 dissolved 15% more than the reference batch at 30 minutes, but the differences between the test and reference batches are less than 8% at any time point after 30 minutes. Test batch #3 is more than 12% different compared with the reference batch at 90 minutes and less than 10% at any other time points. Test batch #4 differs with the reference batch by more than 19% at 30 minutes and shows no difference at any other time point. Test batch #5 differs with the reference batch by more than 17% at 60 minutes, but less than 10% at any other time point. The $\hat{f}_2$ of the five test batches are 60.04 for test batch #1, 51.08 for test batch #2, 51.19 for test batch #3, 50.07 for test batch #4 and 48.05 for test batch #5. When taking $\hat{f}_2$ as f_2 for dissolution similarity assessment, one would consider that all test batches except batch #5 have dissolution profile similar to the reference batch, when the similarity criterion value, 50 (computed based on an 10% average distance at all key time points) is used. However, the bootstrap confidence intervals of $E(\hat{f}_2)$ give the lower 90% confidence limits lower than similarity criterion in this example, using either the percentage confidence interval (PI) or the BC_α confidence interval which are given in Table 6 with 100, 200, 400, 500 and 1,000 bootstrap samples. The 90% lower confidence limits BC_α based on the 500 samples are 52.79 for test batch #1, 48.39 for test batch #2, 48.59 for test batch #3, 48.38 for test batch #4 and 46.11 for test batch #5. It indicates that all test batches except test batch #1 fail to show dissolution similarity to the reference batch when the f_2 value of 50 is used as a cutoff point for accepting similarity between two dissolution profiles.

Table 4. Example #2 Dissolution Data of Reference and Five Test Batches

	Reference batch				Test batch 1				T batch 2			
Time	30	60	90	180	30	60	90	180	30	60	90	180
Tablet												
1	36.1	58.6	80	93.3	38.75	61.79	85.14	100.2	48	60	84	103
2	33	59.5	80.8	95.7	36.16	61.21	84.25	97.3	52	75	89	99
3	35.7	62.3	83	97.1	38.49	63.89	84.94	96.39	48	60	83	101
4	32.1	62.3	81.3	92.8	37.27	62.52	85.65	95.47	53	70	93	103
5	36.1	53.6	72.6	88.8	48.12	77.18	95.32	99.3	45	60	84	105
6	34.1	63.2	83	97.4	48.45	80.62	95.05	98.94	48	66	90	103
7	32.4	61.3	80	96.8	41.08	67.62	84.94	99.03	51	71	91	100
8	39.6	61.8	80.4	98.6	39.64	63.68	80.73	95.63	49	63	89	104
9	34.5	58	76.9	93.3	36.06	61.59	82.22	96.12	44	60	84	103
10	38	59.2	79.3	94	36.69	63.6	84.5	98.42	53	68	81	104
11	32.2	56.2	77.2	96.3	39.95	67.98	87.4	98.1	49	63	86	105
12	35.2	58	76.7	96.8	43.41	74.07	93.95	97.8	52	68	87	104
Mean	34.92	59.5	79.27	95.08	40.34	67.15	87.01	97.73	49.33	65.33	86.75	102.83
	Test batch 3				Test batch 4				Test batch 5			
1	28.7	48.2	63.8	85.6	17.1	58.6	80	93.3	41.5	78	86.4	98.3
2	26.4	53.1	68.3	90.6	16	59.5	80.8	95.7	43.7	78.3	85.9	102.9
3	25.4	52.4	70	89.5	12.7	62.3	83	97.1	46.3	78.3	86.9	96.4
4	23.2	49.5	65.5	92.2	15.1	62.3	81.3	92.8	44	79.9	88.6	96
5	25.1	50.7	68	87.6	14.1	53.6	72.6	88.8	42.6	73.2	81.4	95.5
6	28.7	54.1	70.8	93.6	12.1	63.2	83	97.4	44.4	78.4	86.2	98.4
7	23.5	50.3	66.1	85.1	14.4	61.3	80	96.8	43	79	87.5	99.5
8	26.2	50.6	67.7	88	19.6	61.8	80.4	98.6	44.4	79.6	87.3	99.9
9	25	49.1	63.6	85.8	14.5	58	76.9	93.3	44.8	78.7	86.9	97.8
10	24.9	49.5	66.7	86.6	14	59.2	79.3	94	41.7	76.9	84.5	100
11	30.4	53.9	70.4	89.9	18.2	56.2	77.2	96.3	42.3	77	81.9	97.9
12	22	46.3	63	88.7	13.2	58	76.7	96.8	42	78.2	92.4	100.3
Mean	25.80	50.64	67.00	88.6	15.08	59.5	79.27	95.08	43.39	77.96	86.33	98.58

Table 5. Covariance and Correlation Matrices of the Six Batches

Batch	Time	Std	Covariance				Correlation			
			D30	D60	D90	D180	D30	D60	D90	D180
1	D30	2.36	5.55	−0.21	−0.58	0.19	1.00	−0.03	−0.08	0.03
	D60	2.84	−0.21	8.09	7.88	4.92	−0.03	1.00	0.93	0.63
	D90	2.98	−0.57	7.89	8.88	5.05	−0.08	0.93	1.00	0.62
	D180	2.73	0.19	4.92	5.05	7.43	0.03	0.63	0.62	1.00
2	D30	4.28	18.30	27.39	18.73	3.28	1.00	0.97	0.88	0.50
	D60	6.62	27.39	43.86	30.53	4.67	0.97	1.00	0.93	0.46
	D90	4.97	18.73	30.52	24.75	3.85	0.88	0.93	1.00	0.50
	D180	1.55	3.29	4.67	3.85	2.40	0.50	0.46	0.50	1.00
3	D30	2.96	8.79	12.33	4.18	−1.48	1.00	0.81	0.38	−0.26
	D60	5.10	12.33	26.06	11.18	−5.21	0.81	1.00	0.60	−0.54
	D90	3.67	4.18	11.18	13.48	−2.23	0.38	0.60	1.00	−0.32
	D180	1.90	−1.48	−5.21	−2.23	3.61	−0.26	−0.54	−0.31	1.00
4	D30	2.47	6.10	3.74	3.70	1.40	1.00	0.64	0.56	0.21
	D60	2.37	3.74	5.60	5.95	3.51	0.64	1.00	0.94	0.55
	D80	2.68	3.70	5.95	7.19	4.03	0.56	0.94	1.00	0.56
	D180	2.68	1.40	3.51	4.03	7.17	0.21	0.55	0.56	1.00
5	D30	2.56	5.10	−0.98	−0.56	0.81	1.00	−0.15	−0.08	0.13
	D60	2.84	−0.98	8.09	7.89	4.92	−0.15	1.00	0.93	0.63
	D90	2.98	−0.56	7.89	8.88	5.05	−0.08	0.93	1.00	0.62
	D180	2.73	0.81	4.92	5.05	7.43	0.13	0.63	0.62	1.00
6	D30	5.81	2.18	1.02	0.67	−0.72	1.00	0.40	0.16	−0.24
	D60	1.74	1.02	3.02	3.52	1.13	0.40	1.00	0.70	0.31
	D90	2.90	0.67	3.52	8.40	1.73	0.16	0.70	1.00	0.28
	D180	2.10	−0.72	1.13	1.73	4.39	−0.24	0.31	0.28	1.00

Efron and Tibshrani (15) indicated that in general a bootstrap of 400 sample sets give precise estimate. However, the rate of convergence of the bootstrap confidence limits is data dependent, and it is recommended to calculate a few bootstrap estimates in order to make sure that the estimate is stable. Table 6 shows that the confidence intervals are quite stable with 500 sample sets for both Percent interval and BC_α estimate.

E. Bias of the Estimate of Similarity Factor

The confidence interval estimated using bootstrap method is for the expected value of $\hat{f}_2$, $E(\hat{f}_2)$. The assessment of dissolution similarity using the confidence interval as in the last section is unbiased only if $\hat{f}_2$ is an unbiased estimate of f_2, which means $E(\hat{f}_2) = f_2$. Assuming that there are n tablets in both the test and reference batches, consider the expected value of $[(1/P)\ \Sigma_{i=1}^{P}\{\Sigma_{j=1}^{n}\ (x_{tij} - x_{rij})/n\}^2]$,

$$E\left[(1/P)\sum_{i=1}^{P}\left\{\sum_{j=1}^{n}(x_{tij} - x_{rij})/n\right\}^2\right]$$

$$= E\Bigg((1/P)\Bigg\{\sum_{i=1}^{P}\Bigg[\Bigg(\sum_{j}(x_{tij} - x_{rij})/n - (\mu_{ti} - \mu_{ri})\Bigg)^2\Bigg]$$

Table 6. Bootstrap Confidence Intervals

Type of CI	Test batch	Sample mean	100 Bootstraps		200 Bootstraps		400 Bootstraps		500 Bootstraps		1,000 Bootstraps	
			Mean	CI	Mean	CI	Mean	CI	Mean	CI	Mean	CI
PI[a]	1	60.03	61.16	(54.26, 70.28)	60.57	(53.73, 70.13)	60.17	(52.84, 68.69)	60.11	(52.79, 68.15)	60.22	(53.01, 68.34)
Bcα[b]		60.08[c]		(54.18, 70.24)		(54.34, 70.73)		(54.19, 70.73)		(54.07, 70.35)		(53.89, 70.24)
PI	2	51.08	50.96	(48.23, 53.32)	51.03	(48.36, 53.63)	50.97	(48.25, 53.71)	50.98	(48.33, 53.68)	51.01	(48.25, 53.69)
Bcα		51.01		(48.37, 53.46)		(48.37, 53.68)		(48.35, 53.77)		(48.39, 53.74)		(48.37, 53.74)
PI	3	51.19	51.22	(48.47, 54.11)	51.16	(48.52, 54.05)	51.27	(48.59, 54.05)	51.29	(48.59, 54.10)	51.29	(48.54, 54.56)
Bcα		51.19		(48.14, 53.95)		(48.49, 53.94)		(48.47, 53.87)		(48.41, 53.91)		(48.41, 54.22)
PI	4	50.07	49.86	(48.51, 51.42)	49.93	(48.49, 51.50)	49.96	(48.41, 51.55)	49.96	(48.39, 51.55)	49.99	(48.38, 51.59)
Bcα		50.06		(48.96, 51.90)		(48.75, 51.69)		(48.63, 51.85)		(48.60, 51.74)		(48.47, 51.73)
PI	5	48.05	48.17	(46.52, 49.91)	48.14	(46.35, 49.89)	48.00	(46.08, 49.91)	48.01	(46.11, 50.09)	48.01	(46.05, 50.04)
Bcα		48.05		(46.41, 49.89)		(46.01, 49.78)		(46.32, 50.33)		(46.33, 50.33)		(46.15, 50.17)

[a] Percent confidence interval.
[b] Bcα adjusted confidence interval.
[c] Jackknife mean.

$$+ \sum_{i=1}^{P} \sum_{i} (\mu_{ti} - \mu_{ri})^2 \Big\}\Big)$$

$$= (1/P)\Big[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]$$

where σ_{ti}^2 and σ_{ri}^2 are the variances of percent dissolution measured at the i-th time point of the test and reference batches respectively.

When n becomes large, the expected value of the mean squared differences between sample means $E[(1/P)\Sigma_{i=1}^{P}\{\Sigma_{j=1}^{n}(x_{tij}-x_{rij})/n\}^2]$ becomes very close to $(1/P)[\Sigma_{i=1}^{P}(\mu_{ti}-\mu_{ri})^2]$, the mean squared difference between population means. Hence, $\hat{f}_2$ is an asymptotically unbiased estimator of f_2.

In contrast,

$$E(\hat{f}_2) = E\Big(50\bullet\log\Big\{\Big[1 + (1/P)\sum_{i=1}^{P} (\bar{x}_{ti} - \bar{x}_{ri})^2\Big]^{-1/2}\bullet 100\Big\}\Big)$$

$$= E\Big\{100 - 25\bullet\log\Big[1 + (1/P)\sum_{i=1}^{P} (\bar{x}_{ti} - \bar{x}_{ri})^2\Big]\Big\}$$

$$\approx 100 - 25\bullet\log\Big(1 + E\Big[(1/P)\sum_{i=1}^{P} (\bar{x}_{ti} - \bar{x}_{ri})^2\Big]\Big)$$

$$= 100 - 25\bullet\log\Big(1 + (1/P)\Big[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big) \qquad (2)$$

with Taylor's expansion

$$= 100 - 25\Big\{\log\Big(1 + (1/P)\Big[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2\Big]\Big) + \Big[(1/P)\sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big/\Big[1 + (1/P)\Big[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big]$$

$$- \Big(\Big[(1/P)\sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big/\Big[1 + (1/P)\Big[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big]\Big)^2\Big/2 + \ldots\Big\}$$

which in term

$$< 100 - 25\bullet\log\Big(1+(1/P)\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2\Big]\Big)$$

$$= 50\bullet\log\Big\{\Big[1 + (1/P)\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2\Big]^{-1/2}\bullet 100\Big\}$$

$$= f_2.$$

With bias

$$= -25\Big[(1/P)\sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big/\Big[1 + (1/P)\Big[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big]$$

$$- \Big(\Big[(1/P)\sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big/\Big[1 + (1/P)\Big[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n\Big]\Big]\Big)^2/2 + \ldots\Big\}$$

This implies that the use of $\hat{f}_2$ is conservative in assessment of dissolution similarity for the criterion defined for actual dissolution profiles.

F. Bias Correction

As shown in last equation, the estimation bias is contributed by the term $\Sigma_{i=1}^{P}(\sigma_{ti}^2 + \sigma_{ri}^2)/n$ within the log function. An intuitive bias correction would lead to subtracting the unbiased estimate of $\Sigma_{i=1}^{P}(\sigma_{ti}^2 + \sigma_{ri}^2)/n$ within f_2 and we have the following unbiased estimate

$$\hat{f}_2^* = 50\bullet\log\Big(\Big\{1 + (1/P)\Big[\sum_{i=1}^{P} (\bar{x}_{ti} - \bar{x}_{ri})^2 - \sum_{i=1}^{P} (s_{ti}^2 + s_{ri}^2)/n\Big]\Big\}^{-1/2}\bullet 100\Big)$$

where s_{ti}^2 and s_{ri}^2 are the unbiased estimates of variance at the i-th time point of the test and reference batches respectively. The confidence interval is then adjusted accordingly. For the five test batches in example #2 the bias adjusted estimates are given in Table 7. It is shown in Table 7 that for the five comparisons, the bias adjusted estimates of f_2 are not much different to the biased estimate $\hat{f}_2$ because of the small dissolution variance of the six (one reference and five test) batches. However, the adjustment is not valid when $\Sigma_{i=1}^{P}(\bar{x}_{ti} - \bar{x}_{ri})^2 < \Sigma_{i=1}^{P}(s_{ti}^2 + s_{ri}^2)/n$.

G. Corresponding Hypotheses for Similarity Testing

When $\hat{f}_2$ is used as an estimate of f_2, the application of $\hat{f}_2$ for the assessment of dissolution similarity, can be interpreted in two ways.

1. When the similarity limit SL_{f2} is set independent of the data of existing reference batches, for example a fixed SL_{f2} =

Table 7. Bias Adjusted Estimate of f_2

Testing batch	f_2	$\Sigma (S_{ti}^2 + S_{ri}^2)/12$[a]	$\Sigma (X_{ti} - X_{ri})^2$[b]	f_2*
1	60.03	9.94	154.81	60.73
2	51.08	6.82	358.96	51.29
3	51.19	4.67	354.34	51.34
4	50.07	4.95	393.36	50.21
5	48.05	3.99	474.61	48.14

[a] Average of sum of within batch variances.

[b] Sum of between batch mean squares.

50 for all products, the similarity limit is set for f_2 instead of $E(\hat{f}_2)$. The similarity comparison using the lower 90% limit of the bootstrap confidence interval of $E(\hat{f}_2)$ as proposed earlier is an approximation test for the following hypotheses,

$$H_0'\colon 100 - 25 \log\left(1 + (1/P)\left[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma^2{}_{ri})/n\right]\right) \le SL_{f2}$$

versus

$$H_a'\colon 100 - 25 \log\left(1 + (1/P)\left[\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma^2_{ri})/n\right]\right) > SL_{f2}$$

H_0' will be rejected only when both $\Sigma_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2$ and $\Sigma_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n$ are small. Hence it may be considered to be a conservative test for the following hypotheses regarding f_2,

$$H_0\colon 100-25 \log (1 + (1/P) [\Sigma_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2]) \le SL_{f2}$$

versus

$$H_a\colon 100-25 \log (1 + (1/P) [\Sigma_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2]) > SL_{f2}$$

In this case, the proposed procedure can be taken either as a conservative test for H_0 or it is considered as an approximate test for H_0' when $\Sigma_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/nP$ is very small (i.e., if n is large and within batch variances, σ_{ti}^2 and σ_{ri}^2 are small at every sampling time point). Correction for estimation bias may not exist when batches are similar but within-batch variances are large.

2. However, if SL_{f2} is product specified and determined by experience with the observed means of the reference batches, the similarity limit is actually set for $E(\hat{f}_2)$ instead of f_2. In another word, it is actually $SL_{E(\hat{f}2)}$ instead of SL_{f2}. Correspondingly, the hypotheses are defined instead as

$$H_0''\colon 100-25 \log(1 + (1/P)[\textstyle\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n]) \le SL_{E(f2)}$$

versus

$$H_a''\colon 100-25 \log(1 + (1/P)[\textstyle\sum_{i=1}^{P} (\mu_{ti} - \mu_{ri})^2 + \sum_{i=1}^{P} (\sigma_{ti}^2 + \sigma_{ri}^2)/n]) > SL_{E(f2)} \qquad (3)$$

and the test would be considered unbiased.

H. Limitations of Similarity Factor

Measurements of difference or similarity of the profiles are often based on combining the differences at all time points into one measurement. Such measurements are often estimated by substituting sample means for the actual means. However, with the dissolutions correlated at the sample time points, such estimates are often complicated in that the variation of the estimate is difficult to calculate and the estimate itself may be biased, and the statistical properties are difficult to derive. The similarity measurement f_2 is not an exception. Even the logarithm transformation of f_2 complicates the known sampling distribution of mean squared differences. As pointed out by Liu et al. (12) there is no mathematical formula for the sampling distribution either in exact or asymptotic form. It is therefore difficult to assess the type I (consumer's risk) and type II (manufacturer's risk) error rates. Without these error rates, it is difficult to evaluate the power, sample size, magnitude of the bias, validity of the approximation, and sensitivity of the $\hat{f}_2$ test.

The similarity factor f_2 is a function of the mean differences and does not take into account the differences in dissolutions within the test and reference batches. Hence careful interpretation is warranted when f_2 is used as a similarity factor for batches with large difference in variance.

CONCLUSIONS

Through mathematical scaling, the f_2 measurement takes the values ranging from 0 to 100. A convenient critical value of 50 is derived for similarity of dissolution profiles based on average difference of 10% at all sampling time points. Since the f_2 is sensitive to the measurements obtained after either test or reference batch has dissolved more than 85%, it leads to the recommendation of limiting to no more than one sampling time point after 85% dissolution.

In conclusion, similarity factor f_2 provides a simple measure for the comparison of two dissolution profiles. The analysis and discussion suggest that the commonly used estimate $\hat{f}_2$ has complicated statistical properties. Its true expected value is a function of both between batch differences and within batch variations. It is therefore a biased and conservative estimate of f_2. The bootstrap method is proposed as a tool to simulate the confidence interval. The number of bootstrap sample required depends on the convergence rate of the confidence interval.

REFERENCES

1. FDA Guidance for Industry: Immediate Release Solid Dosage Forms: Scale-up and Post Approval Changes (SUPAC-IR): Chemistry, Manufacturing and Controls, In Vitro Dissolution Testing and In Vivo Bioequivalence Documentation, November, 1995.
2. FDA Guidance for Industry: Dissolution Testing of Immediate Release Solid Oral Dosage Forms, August, 1997.
3. FDA Guidance for Industry: Extended Release Oral Dosage Forms: Development, Evaluation and Application of In Vitro/In Vivo Correlations, September, 1997.
4. FDA Guidance for Industry: Modified Release Solid Oral Dosage Forms: Scale-Up and Post Approval Changes (SUPAC-MR): Chemistry, Manufacturing and Controls; In Vitro Dissolution Testing and In Vivo Bioequivalence Documentation, September, 1997.
5. V. P. Shah, L. A. Yamamoto, D. Shuirmann, J. Elkins, and J. P. Skelly. Analysis of in vitro dissolution of whole versus half controlled release theophylline tablets. *Pharm. Res.* **4**:416–419 (1987).
6. S. C. Chow and F. Y. C. Ki. Statistical comparison between dissolution profiles of drug products. *J. Biopharm. Stat.* **7**:241–258 (1997).
7. P. Sathe, Y. Tsong, and V. P. Shah. In-vitro dissolution profile comparison: statistics and analysis, model dependent approach. *Pharm. Res.* **13**:1799–1803 (1996).
8. Y. Tsong, T. Hammerstrom, P. Sathe, and V. P. Shah. Statistical assessment of mean difference between two dissolution data sets. *Drug Inf. J.* **30**:1105–1112 (1996).

9. J. W. Moore and H. H. Flanner. Mathematical comparison of curves with an emphasis on in-vitro dissolution profiles. *Pharm. Tech.* **20**(6):64–74 (1996).
10. J. E. Polli, G. S. Rekhi, and V. P. Shah. Methods to compare dissolution profiles. *Drug Info. J.* **30**:1113–1120 (1996).
11. J. E. Polli, G. S. Rekhi, L. L. Augsburger, and V. P. Shah. Methods to compare dissolution profiles and a rationale for wide dissolution specifications for metoprolol tartrate tablets. *J. Pharm. Sci.* **86**:690–700 (1997).
12. J. P. Liu, M. C. Ma, and S. C. Chow. Statistical evaluation of similarity factor f_2 as a criterion for assessment of similarity between dissolution profiles. *Drug Info. J.* **31**:1255–1271 (1997).
13. H. L. Ju and Liaw S. On the assessment of similarity of drug dissolution profile—a simulation study. *Drug Info. J.* **31**:1273–1289 (1997).
14. R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis,* Prentice-Hall, Inc. Englewood Cliffs, NJ, 1982
15. Y. Tsong, P. Sathe, and V. P. Shah. Comparing two dissolution data sets for similarity, American Statistical Association. 1996 Proceedings of the Biopharmaceutical Section, 129–134, 1997.
16. B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap,* Chapman & Hall, New York, NY, 1993.