A new principal component analysis-based approach for testing “similarity” of drug dissolution profiles

Rubén M. Maggio, Patricia M. Castellano, Teodoro S. Kaufman

School of Biochemical and Pharmaceutical Sciences, National University of Rosario, Suipacha 531, S2002LRK Rosario, Argentina
Institute of Chemistry of Rosario (IQUIR, CONICET-UNR), Suipacha 531, S2002LRK Rosario, Argentina

Article info
Article history: Received 19 December 2007 Received in revised form 14 February 2008 Accepted 20 February 2008 Published on line 4 March 2008

Keywords: Principal component analysis Dissolution profiles Similarity test Acetaminophen Furosemide Multivariate method

Abstract
A new approach for testing batch “similarity” through comparison of drug dissolution profiles, based on principal component analysis with the establishment of a confidence region (PCA-CR), is presented. The dissolution curves corresponding to three brands each of Furosemide and Acetaminophen tablets, taken as model drugs, were prepared by dissolution measurements at multiple pre-specified time points. Reference and test data were simultaneously subjected to PCA and pairwise comparisons between the dissolution characteristics of lots of the same and different brands were carried out. The comparisons involved plotting the weighed scores of the first two principal components of reference and test lots, while decision about “similarity” was made by checking for inclusion of more than 80% of the tablets of the test lot in the 95% confidence ellipse of the reference samples. Two published datasets were also analyzed in the same fashion and all the results were compared with information provided by the difference ($f_1$) and similarity ($f_2$) factor tests. Unlike the $f_2$ criterion, the proposed method reflects variability within the individual dissolution curves, being also highly sensitive to profile (shape and size) variations. Comparison between the area enclosed by the confidence ellipses of the weighed scores plot and the region obtained from the bootstrap-calculated acceptable values of the corresponding $f_2$ tests suggested that PCA-CR represents, in general, a more discriminating standard.

1. Introduction

In vitro dissolution testing is an economic and useful quality control tool to effectively assure acceptable product quality during different stages of the development and production of tablets, capsules and other solid dosage forms (Dressman and Kramer, 2005). The test enables detection of the influence of key manufacturing factors including excipients, binder and mixing effects, as well as granulation procedure and coating parameters, providing better control of the production process and assuring consistent batch to batch quality of the product. The dissolution has also been employed in product development and during dosage form optimization to assist in proper formulation selection. In addition, it has served as a means to compare different formulations (Naylor et al., 1993) and determine final dissolution specifications for pharmaceutical dosage forms (Elkoshi, 1999).

The dissolution test has also been used during stability studies, helping establish shelf life, and it has been recognized as an important in vitro parameter of tablets’ quality because of its correlation with drug bioavailability (Williams et al., 1991; Fassihi and Ritschel, 1993; Munday and Fassihi, 1995; Grundy...
et al., 1997). As a result, under certain strictly defined conditions, the test can also be employed as a surrogate of in vivo studies for the assessment of product bioequivalence, helping to reduce costs by circumventing the need to perform human volunteer experiments (Leson, 1995; Yu et al., 1996).

Because it is essential to investigate the drug release characteristics of pharmaceutical preparations, dissolution has become highly significant and one of the primary pharmacopoeial tests that is performed to ensure that tablets, capsules and other drug products comply with pre-established quality standards.

For a drug product, the curve of the mean dissolution rate over time is referred to as its dissolution profile. There are several circumstances under which comparison of the dissolution profiles of two solid oral dosage forms is important. Among them, when an approved formulation is subjected to a post-approval change due to modifications of some critical parameters, including manufacturing site, composition, manufacturing process and batch size. In these cases, FDA guidances for scale-up and post-approval changes for solid oral dosage forms (FDA, 1995) require that the dissolution profiles of the pre-change and post-change products must be “similar”.

Another paradigmatic scenario is in the development of generic preparations. Here, a proprietary product, which has been available in the market for some time and has a clinically established efficacy, is selected as a reference against which to compare the new formulation. Because of the “similarity” requirement, the generic preparation should be formulated with its dissolution profile as closely similar as possible to that of the proprietary product.

In response to the need of assessing “similarity”, numerous strategies have been proposed for comparing dissolution profiles. These, which are divided in ANOVA-based, model-dependent and model-independent approaches, have been extensively reviewed (Polli et al., 1996, 1997; O’Hara et al., 1998; Costa and Sousa Lobo, 2001). The ANOVA-based methods (Mauger et al., 1986; Yuksel et al., 2000) assume the existence of underlying models, but do not require fitting of a curve. They test statistical differences of the dissolution profiles in terms of “shape” and “size” of the curves, providing probability values related more to statistical equivalence than to pharmaceutical similarity.

Model-dependent methods rely on curve-fitting procedures, which facilitate data analysis and interpretation because they describe the dissolution profiles as functions of a few model parameters that can be determined and statistically compared. In general, however, these are rather rigid representations, there is no universal model to fit all dissolution profiles and there are no established criteria to select the proper mathematical model.

Model-independent methods do not require a preconceived or fitted model. The difference ($f_1$) and similarity ($f_2$) factors introduced by Moore and Flanner (1996) as mathematical indices to compare dissolution profiles constitute the most widely known examples of the model-independent approach. This procedure, where the dissolution behaviour of a number of samples ($n$) of reference ($R$) and test ($T$) products are compared at $t$ time points (Eqs. (1) and (2)), is being recommended by the FDA Guidance for Industry (FDA, 1995), and has been accepted by European agencies and other regulatory bodies (Human Medicines Evaluation Unit, 1999). For testing purposes, a discriminatory medium can be identified by varying stirring rate and parameters of the dissolution medium, including pH, ionic strength, volume, etc.

$$f_1 = 100 \left( \frac{\sum_{t=1}^{n} |R_t - T_t|}{\sum_{t=1}^{n} R_t} \right)$$

$$f_2 = 50 \log \left\{ \left[ 1 + \left( \frac{1}{n} \sum_{t=1}^{n} (R_t - T_t)^2 \right)^{0.5} \right] \times 100 \right\}$$

Since drug release depends on many variables, such as the physicochemical properties of the drug, the excipients and the structural properties of the tablet matrix, an understanding of the complex causalities between different variables and responses becomes difficult. Therefore, for decision taking, it is useful to collapse this complex information into a minimum identifiable number of parameters. As a variable simplification approach, in many cases two batches are compared through the determination of their percentage of dissolved active component at a certain time point. However, this provides less meaningful conclusions than the independent comparison of specifications at each of multiple time points or the analysis of the entire dissolution profile. For such problems, multivariate data analysis is the tool of choice. Multivariate methods such as principal component analysis (PCA) have been suggested for the evaluation of dissolution profiles (Tsong et al., 1997; Adams et al., 2001, 2002), while other approaches including artificial neural networks with similarity factor (Peh et al., 2000; Goh et al., 2002, 2003) and Gaussian mixture models (Lim et al., 2005) as well as partial least squares (Korhonen et al., 2005), have been proposed as multivariate strategies for the prediction of dissolution profiles.

Here, we propose the application of PCA with confidence regions (PCA-CR) as a new and alternative method to compare solid dosage forms dissolution behaviour and decide about their “similarity”. The usefulness of the suggested strategy was demonstrated by comparing different brands and lots of tablet preparations containing either Furosemide or Acetaminophen, as models, and also two selected literature datasets. For assessing the scope and limitations of the proposed approach, the PCA-CR results were confronted in each case with the conclusions provided by the corresponding $f_1$ and $f_2$ factors, taken as reference.

2. Materials and methods

2.1. Equipment, software and reagents

Dissolution tests were performed with a Hanson SR8-Plus dissolution test station configured as USP-apparatus II (paddle). The amounts of drug dissolved were determined in 1-cm quartz cells, employing a Shimadzu UV-1601PC spectrophotometer interfaced to a computer running UV-Probe software.
v. 2.00. Determinations were carried out against a blank of dissolution medium, on filtered samples suitably diluted with dissolution medium, by comparison with standard solutions containing known concentrations of the corresponding analytes. All the reagents employed were of analytical grade; double distilled water was employed as solvent. All the computations were performed in Matlab v. 5.3 (Natick, MA); the Matlab scripts are freely available from the authors.

2.2. Table preparations and dissolution conditions

All the brands and lots of Furosemide and Acetaminophen drug products used met the pharmacopoeial specifications for weight variation, content uniformity and assay.

2.2.1. Furosemide

Eight lots corresponding to three different brands of tablet products (40 mg) were studied. Each product was randomly labelled with a specific letter for identification, designating with A1 the reference lot of the innovator product. The release characteristics were determined at 37 ± 0.5 °C, using the procedure of the “Dissolution Test 1” of USP 30 (USP Convention, 2007). The medium was 900 ml of Phosphate buffer (0.05 M, pH 5.8) and the stirring rate was 50 rpm. One tablet was used in each vessel, and each test comprised two runs of six tablets yielding a total of 12 tablets per lot (FDA, 1995). During each experiment, aliquots of 3 ml were removed at 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 26 and 30 min, filtered and suitably diluted with medium. The amount of drug dissolved was determined from the absorbances of the samples at 274 nm. Each dissolution curve contained a total of 17 time points.

2.2.2. Acetaminophen

Three different brands of tablet products (500 mg) were studied and brand A was used as the reference product (innovator). The other brands were each randomly designated with letters B and C for identification, designating with B1 the reference lot of the innovator product. Dissolutions were determined at 37 ± 0.5 °C in 900 ml of Phosphate buffer (0.05 M, pH 5.8), using a slight modification (stirring rate was 30 rpm) of the USP 30 procedure in order to increase selectivity. One tablet was used in each vessel, and each test comprised two runs of six tablets yielding a total of 12 tablets per lot (FDA, 1995). During each experiment, aliquots of 3 ml were removed at 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 26 and 30 min, filtered and suitably diluted with medium. The amount of drug dissolved was determined from the absorbances of the samples at 243 nm. Each dissolution curve contained a total of 10 time points.

2.2.3. Literature data

The other brands were each randomly designated with letters B and C for identification, designating with A1 the reference lot of the innovator product. Three different brands of tablet products (500 mg) were studied. Each product was randomly labelled with a specific letter for identification, designating with A1 the reference lot of the innovator product. The release characteristics were determined at 37 ± 0.5 °C, using the procedure of the “Dissolution Test 1” of USP 30 (USP Convention, 2007). The medium was 900 ml of Phosphate buffer (0.05 M, pH 5.8) and the stirring rate was 50 rpm. One tablet was used in each vessel, and each test comprised two runs of six tablets yielding a total of 12 tablets per lot (FDA, 1995). During each experiment, aliquots of 3 ml were removed at 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 26 and 30 min, filtered and suitably diluted with medium. The amount of drug dissolved was determined from the absorbances of the samples at 274 nm. Each dissolution curve contained a total of 17 time points.

2.2.4. Theoretical background of the \( f_1 \) and \( f_2 \) estimators and the PCA algorithm

2.2.4.1. Factors \( f_1 \) and \( f_2 \) as estimators of difference and similarity. Eqs. (1) and (2) correspond to the difference (\( f_1 \)) and similarity (\( f_2 \)) factors, respectively (Moore and Flanner, 1996).

The \( f_1 \) index computes the absolute cumulative differences between drug release in reference and test samples, relative to the drug dissolved in the reference sample. Therefore, the value of this parameter, which is proportional to the average difference between both profiles, depends on which sample is taken as reference. Acceptable values of \( f_1 \) are \( 0 \leq f_1 \leq 15 \).

On the other hand, \( f_2 \) is a logarithmic function of the reciprocal of the mean square-root transform of the sum of squared errors at all points, and is a measure of the degree of similarity in the percent rate of drug release between two dissolution profiles. The \( f_2 \) values are independent from the sample taken as reference, and they range between 0 and 100, with a higher number indicating better similarity between profiles. Acceptable values are \( 50 \leq f_2 \leq 100 \), which is considered equivalent to a difference in approximately 10% between the dissolution profiles being compared (Shah et al., 1998).

2.2.4.2. Principal component analysis. The principles underlying PCA have been extensively discussed elsewhere (Wold et al., 1987); the following is a brief description of this multivariate method.

Given matrix \( X_{(p \times q)} \), where each row contains \( t \) different pieces of information gathered from \( p \) objects, the column mean centred data matrix \( X_c \) can be obtained by subtracting the row vector containing the mean values of its columns \( (Xm) \), from each row of the original matrix \( (X) \).

In turn, \( X_c \) can be decomposed into the product of an orthogonal matrix \( U \), a diagonal matrix \( S \) and another orthogonal matrix \( V \) (Eq. (3)), where \( U \) \( S \) \( V \) represents the singular value decomposition (SVD) of \( X_c \) (Manly, 1986).

\[
X_{c(p \times t)} = U_{(p \times t)} S_{(t \times t)} V^{T}_{(t \times t)} (p > t)  \tag{3}
\]

The score matrix \( U_{(p \times t)} \) is the unweighted (normalized) score matrix and represents the projections of the data on the PCs; therefore, similar samples are represented by similar scores. On the other side, the diagonal matrix \( S_{t \times t} \) contains the singular values, which are the square roots of the eigenvalues associated to the corresponding PCs (eigenvectors). These diagonal terms reflect the dynamics of the dissolution; therefore, the largest eigenvalues correspond to the dimensions that explain larger amounts of variance of the dataset. Matrix \( T_{(t \times k)} \) known as the weighed (unnormalized) score matrix, is the product between \( U \) and \( S \) \((T = U S)\). Finally, the loadings matrix \( V_{(p \times t)} \) contains in its columns the weights contributed by the original variables (eigenvectors) to the PCs.

2.2.5. Detection of outliers

Outlier detection was performed by means of Hotelling’s test (Jackson, 1991). For that purpose, the test was implemented for each dataset, according to Eq. (4), where \( m_x \) is the mean of the data and \( S_{XX}^{-1} \) is the inverse of the data covariance matrix \( S_{XX} \) (Eq. (5)). The required Mahalanobis distance was calculated according to Eq. (6), where \( q \) is the number of dissolution curves in the reference and test lots (Section 2.2.9), and was compared with the corresponding Chi square value at a 99% confidence level and \( t \) (number of data points per
2.2.8. Bootstrapping procedure for finding the $f_2 \leq 50$ region

The mean vector of data $a_{(t \times q)}$ (dissolution profile) of the reference lot was successively transformed into a new vector $d_{(t \times q)}$ by replacing some of its items with artificial data containing deviations able to originate $f_2$ values around 50. This procedure was repeated a number of times, and in each case the values of $f_2$ and the PCs of the artificial dissolution curve were calculated (Efron and Tibshirani, 1986, 1993; Shah et al., 1998; Adams et al., 2001). Plots of the $f_2 = 50$ ellipses (enclosing the $f_2 \leq 50$ region) are shown in the graphics.

2.2.9. Procedure for the comparison of dissolution profiles

Given the data matrices $A_{(q \times t)}$ and $B_{(q \times t)}$, containing the dissolution curves of $q$ tablets each corresponding to the lots of dosage forms to be compared, taken at the same time points, the following five steps are proposed to be sequentially carried out:

(a) Detect outliers in the individual datasets, employing Hotelling’s test (Section 2.2.5).
(b) Construct the matrix $X_{(2 q \times t)}$, which contains the data of $A$ and $B$ ($2q = p$, see Section 2.2.4.2); mean-center (columnwise) this matrix and carry out the SVD operation on the resulting matrix $X_{c(2 q \times t)}$, which contains matrices $U$, $S$ and $V$.
(c) Select the number of PCs to be retained (Section 2.2.6) and compute matrix $T_{(q \times t)}$.
(d) Draw the 95% confidence region (Section 2.2.7), in order to test the hypothesis of similarity.
(e) Decide about “similarity”, based on the inclusion of the test data (>80%) in the confidence ellipse of the reference.

3. Results and discussion

3.1. Characteristic features of the proposed PCA-CR approach

The proposed approach for the assessment of “similarity” through the PCA-CR analysis of dissolution curves entails five steps, including (a) detection of outliers in reference and test data matrices; (b) construction and column mean centering of a single data matrix containing both data of reference and test samples, which is submitted to a SVD operation; (c) selection of the number of PCs to be retained; (d) plotting of the weighted scores of reference and test lots, and drawing of the 95% confidence region based on scores plot of the reference lot; (e) “similarity” decision making based on the percentage of test samples included in the above confidence region.

These sequential steps constitute the appropriate means for pre-processing, analyzing and visualizing the data, also establishing a convenient approach for final decision taking. Hotelling’s test represents a useful strategy for outlier detection, helping to avoid inclusion of dissolution curves with exceptionally high variability. On the other hand, PCA is a mathematical procedure that allows the representation of a complex set of multivariate data with a reduced number of new and uncorrelated variables (PCs), which are linear combinations of the original data. In the proposed method, joining reference and test datasets and carrying out the SVD on a single matrix allows the optimization of system parameters leading to an improved projection of the test data in the reference–test joint data space; therefore, misadjustments resulting form fitting test data into a pre-established reference model, are avoided.

By discarding feature elements with low variability, PCA allows data visualization and the discovery of hidden trends.
in fewer dimensions. The PCs are ordered according to their ability to explain data variability, and in the selected examples discussed below, plot of the weighed scores of the first two PCs is proposed, as these allow satisfactory reconstruction of the original data matrix. This approach is much simpler than that proposed by Tsong et al. (1997), which employs all the PCs for comparison.

After some trial and error experiments, and taking into account that “similarity” is a property of the lot and not of the individual tablets, for decision taking, the following criterion was adopted “test lots are considered to be ‘similar’ if they contain more than (the arbitrarily chosen value of) 80% of their tablets (Chen and Tsong, 1997) inside of the 95% confidence region of the reference lot”. The “>80%” requirement takes into account test data variability, while the 95% confidence ellipse considers variability of the weighed scores in the reference lot.

In order to assess the usefulness of the proposed method, the dissolution curves of Acetaminophen and Furosemide tablets, and two datasets selected from the literature, were individually analyzed by the PCA-CR methodology and compared with the information provided by the $f_1/f_2$ criteria. Results for each set of data are presented below and discussed separately.

### 3.2. Dissolution of Furosemide tablets

Eight lots of Furosemide tablets, corresponding to three different brands, A (lots A1, A2 and A3), B (lots B1, B2 and B3) and C (lots C1 and C2), were studied, with brand A being the innovator. The mean percentages of drug released over a 30-min period are depicted in Fig. 1.

The individual profiles of the eight lots complied with the FDA requirements for the evaluation of similarity and difference; i.e., the tablets dissolved less than 85% of their active principle in the first 15 min, the data coefficient of variation (CV%) was less than 20% for the first time points, being less than 10% for the remaining time points (≥6 min) and the overall CV% was less than 15%. The corresponding $f_1$ and $f_2$ values were calculated, employing data acquired at 2, 7, 12, 18 and 26 min, taking care that no more than one time point (26 min) corresponded to more than 85% of dissolved drug. For the sake of the analysis, all of the possible pairwise lot comparisons were carried out, with the results consigned in Table 1.

From the data of Table 1, it follows that when compared against A1, both additional lots of tablets of brand A (A2 and

<table>
<thead>
<tr>
<th>Lot</th>
<th>$f$-Criterion</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>$f_1$</td>
<td>2.7</td>
<td>2.8</td>
<td>4.0</td>
<td>3.7</td>
<td>6.5</td>
<td>11.3</td>
<td>8.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>78.0</td>
<td>78.4</td>
<td>77.0</td>
<td>70.7</td>
<td>63.5</td>
<td>53.7</td>
<td>59.2</td>
<td></td>
</tr>
<tr>
<td>A2</td>
<td>$f_1$</td>
<td>2.8</td>
<td>1.0</td>
<td>5.0</td>
<td>6.6</td>
<td>9.4</td>
<td>8.8</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>78.0</td>
<td>94.5</td>
<td>67.4</td>
<td>64.6</td>
<td>58.2</td>
<td>57.7</td>
<td>64.6</td>
<td></td>
</tr>
<tr>
<td>A3</td>
<td>$f_1$</td>
<td>2.9</td>
<td>1.0</td>
<td>4.3</td>
<td>5.9</td>
<td>8.7</td>
<td>9.4</td>
<td>7.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>78.4</td>
<td>94.5</td>
<td>68.5</td>
<td>67.1</td>
<td>60.1</td>
<td>55.8</td>
<td>61.9</td>
<td></td>
</tr>
<tr>
<td>B1</td>
<td>$f_1$</td>
<td>4.0</td>
<td>4.8</td>
<td>4.2</td>
<td>4.8</td>
<td>6.1</td>
<td>11.6</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>77.0</td>
<td>67.4</td>
<td>68.5</td>
<td>74.1</td>
<td>67.4</td>
<td>51.3</td>
<td>56.2</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>$f_1$</td>
<td>3.6</td>
<td>6.2</td>
<td>5.6</td>
<td>4.7</td>
<td>2.6</td>
<td>14.5</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>70.7</td>
<td>64.6</td>
<td>67.1</td>
<td>74.1</td>
<td>84.1</td>
<td>46.9</td>
<td>51.1</td>
<td></td>
</tr>
<tr>
<td>B3</td>
<td>$f_1$</td>
<td>6.1</td>
<td>8.6</td>
<td>8.0</td>
<td>5.7</td>
<td>2.6</td>
<td>16.7</td>
<td>14.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>63.5</td>
<td>58.2</td>
<td>60.1</td>
<td>67.4</td>
<td>84.1</td>
<td>43.9</td>
<td>47.4</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>$f_1$</td>
<td>12.7</td>
<td>9.7</td>
<td>10.4</td>
<td>13.1</td>
<td>16.9</td>
<td>20.0</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>53.7</td>
<td>57.7</td>
<td>55.8</td>
<td>51.3</td>
<td>46.9</td>
<td>43.9</td>
<td>81.1</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>$f_1$</td>
<td>9.7</td>
<td>6.8</td>
<td>7.5</td>
<td>10.1</td>
<td>13.8</td>
<td>16.8</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$f_2$</td>
<td>59.2</td>
<td>64.6</td>
<td>61.9</td>
<td>56.2</td>
<td>51.1</td>
<td>47.4</td>
<td>81.1</td>
<td></td>
</tr>
</tbody>
</table>

*Letters designate different brands; numbers differentiate between different lots of the same brand. Non-complying figures are shown in italics.*
$A_3$ exhibited acceptable difference (2.7 and 2.8) and similarity (78.0 and 78.4) results; analogously, the parameters for the three lots of brand B ($f_1 = 3.7–6.5$ and $f_2 = 63.5–77.0$) indicated that they should be considered similar to $A_1$. The $f_1/f_2$-test also suggested that tablets of brand B were similar among them and with brand A. On the other hand, although complying with the requirements for similarity when compared against the tablets of brand A, brand C exhibited a different range of $f_1/f_2$ values ($f_1 = 6.4–11.3$ and $f_2 = 53.7–64.6$), which shifted more towards non-compliance when analyzed against the data of lot B1 ($f_1 = 9.2–11.2$ and $f_2 = 51.3–56.2$). This trend was more clearly evidenced when tested against lots B2 and B3, furnishing in some cases non-complying values. Interestingly, both lots of brand C demonstrated to be similar to each other ($f_1 = 2.6$, 2.7 and $f_2 = 81.1$). Among the tested lots, both $f$-factors allowed to arrive at the same conclusion, except in the case of the $B_2$–$C_1$ comparison, where the $f_1$ estimator suggested “similarity”, while its $f_2$ counterpart indicated that the lots were not similar.

To evaluate the performance of the PCA-CR method, Hotelling’s test was run and, since no outliers were detected, the weighed scores of the first two PCs of pairs of Furosemide lots were plotted, with selected results shown in Fig. 2. Each plot displays the corresponding 95% confidence ellipse and the coordinates of the mean values of the weighed scores of the reference and test lots. The regions where most of the samples would exhibit $f_2 = 50$, calculated employing the bootstrapping technique, are also included (Shah et al., 1998). The images clearly show that lots $A_2$, $B_1$ and $B_2$ can be considered similar to $A_1$ (Fig. 2a–c), despite that one of the samples of lot $B_2$ falls out of the 95% confidence region (Fig. 2c). On the other hand, and contradicting $f_1/f_2$ predictions, non-similarity between $A_1$ and both batches of brand C tablets is evident, despite of the fact that $C_1$ and $C_2$ exhibit similarity to each other (Fig. 2f). However, since in the $A_1$–$C_1$ comparison only four tablets of the test lot fall outside of the 95% confidence ellipse, should a multiple stage acceptance rule be in practice (Tsong et al., 1995; USP Convention, 2007), lot $C_1$ could perhaps be considered for a second stage. This instance, while representing a less demanding standard than the single stage PCA-CR method, may still be more discriminant than the $f_1/f_2$-criteria. As expected, comparison between lots B and C clearly evidenced non-similarity despite that some of the observed $f_2$ values (>45) were relatively close to the lower acceptable limit. The similarity and difference factors appear to be simple and easy to be calculated; perhaps this is the key for their
adoption by the industry, despite that they impose restrictions to the quality of the data to be used (not more than one point above 85% dissolution and specific limits to the CV% at different points of the dissolution profile), and their outcome exhibit some dependence upon the number and position of time points employed (Polli et al., 1997). The similarity function \( f_2 \) has been severely criticized by several authors (Eaton et al., 2003) arguing that it also lacks statistical justification (Tsong et al., 1996; Shah et al., 1998; Ma et al., 1999; Chow and Shao, 2002), and performs unnecessary use of the logarithmic reciprocal square root transformation, which makes its statistical distribution very complicated and almost intractable (Liu et al., 1997).

Contrarily, the proposed PCA-CR method offers a simple graphical and analytical way to decide about similarity, employing sound mathematical and statistically based procedures. In addition, it is able to make use of all the available data points, regardless the amount of drug dissolved and data variability. This is highly advantageous, since it provides a better appreciation of the dissolution behaviour of the lots being compared.

The PCA-CR results for lots A and B showed good agreement with the outcome of the corresponding determinations of \( f_1 \) and \( f_2 \); however, both methods provided different conclusions for the comparison between lots A and C. In the \( f_1 \)-test, brand C exhibited compliance but \( f_1 >6.0 \) and \( f_2 <60 \) values were observed to fall in a different range than those of brands A and B. This borderline compliance of both lots of brand C in the \( f_2 \)-test and non-compliance with the PCA-CR method reflects the fact that the latter method represents a slightly rigorous standard than the \( f_1 \)-based approach, being perhaps anticipating non-similarity, as detected when brands B and C were compared.

Both the \( f_1 \)-based and the PCA-CR methods revealed a closer likelihood of brand C towards brand A than with regards to brand B. In fact, the \( f_1 \)-based A–C comparison suggested “similarity”, while the B–C comparison indicated “non-similarity” in the \( f_2 \)-test cases; analogously, in the case of the PCA-CR counterpart, while concluding for “non-similarity” in every case, fewer points remained outside the 95% confidence region in the A–C comparisons (Fig. 2c and d) than in the B–C comparisons (Fig. 2g–i). On the other hand, as in the \( f_1 \)-based comparison, both lots C – of analogous shape and size – were considered similar to each other, despite not being able to achieve “similarity” with the A1 reference lot.

Interestingly, significant correlations were obtained when the number of data points left out of the confidence ellipses were plotted against \( f_1 \) or \( f_2 \) values. However, despite being correlated to the \( f \)-factors, the PCA-CR represents a more rigorous standard, being devoid of some of their major drawbacks.

### 3.3. Dissolution of Acetaminophen tablets

Fig. 3a displays the dissolution profiles of three different brands of Acetaminophen tablets, and Table 2 contains the \( f_1/f_2 \) values of all possible brand-to-brand comparisons, prepared with data taken at 6, 10, 14, 30 and 45 min.

According to the \( f \)-criteria, only brands A and B should be considered “similar”. After running the outlier detection test and demonstrating the suitability of all the dissolution curves, it was observed that this result was in perfect agreement with the conclusions emerging from application of the proposed PCA-CR method (Fig. 4a).

Regarding brand C, it exhibited non-complying \( f_1 \) and \( f_2 \) values, with the data suggesting a possible borderline situation for the B vs. C comparisons \( (f_1 = 15.2 \text{ and } 17.7; f_2 = 47.4) \). Although the latter seemed amenable for a second stage testing, in case of employing a multiple stage acceptance rule (Tsong et al., 1995; USP Convention, 2007), the PCA-CR weighed scores plot demonstrated beyond doubt that brand C was unable to achieve the “similarity” requirements at this stage (Fig. 4c), not qualifying for further “similarity” testing. The same “non-similarity” conclusion was obtained after comparing brands A and C (Fig. 4b). For the sake of discussion, the B–A comparison is also shown (Fig. 4d); despite that the conclusion about “similarity” agrees with the \( f_1 \)-based prediction, owing to different data variability within the reference brand (B), the test brand (A) exhibits two borderline dissolution curves.

Interestingly, areas of the 95% confidence ellipses not always were smaller in size than areas enclosed by the \( f_2 = 50 \) ellipses, calculated under bootstrap assistance. This is because, unlike the \( f_2 \) procedure, the size of the confidence ellipses is related to the variability of the data in the reference samples as well as to size and shape of the dissolution curves.

<table>
<thead>
<tr>
<th>Brand</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>7.4</td>
<td>65.8</td>
<td>20.0</td>
<td>43.2</td>
</tr>
<tr>
<td>B</td>
<td>7.9</td>
<td>65.8</td>
<td>15.2</td>
<td>47.4</td>
</tr>
<tr>
<td>C</td>
<td>25.0</td>
<td>43.2</td>
<td>17.7</td>
<td>47.4</td>
</tr>
</tbody>
</table>

* Non-complying figures are shown in italics.
3.4 Dissolution data of Tsong and Hammerstrom

These dissolution data (Fig. 3b) have been previously employed in different cases for dissolution profile comparisons (Chen and Tsong, 1997; Tsong et al., 1997). All of the $f_1/f_2$ possible pairwise comparisons between the three standard lots (A1–A3) and a fourth lot (B) are consigned in Table 3. The results indicate that the lots have “similar” dissolution characteristics, whichever of them is taken as reference. Noticeably, however, the $f_2$ values for test lot B against the lots A are in a markedly different range (63.5–64.7) from those of the lots A, when tested against each other (90.3–94.6).

No outlier curves were found in the dataset. Plots of the weighted scores of the first and second PCs of the three preapproved batches (A1, A2 and A3) and test batch B are depicted in Fig. 5. Here, all the PCA-based comparisons of the former demonstrated their similarity, despite that the A1–A2 comparison plot exhibited two tablets of the test lot out of the 95% confidence ellipse and those of A2–A3 and A1–A3 showed one tablet each, out of the confidence region.

On the other hand, pairwise comparison of batches A1 and A2 with test batch B, revealed that the latter could not be considered “similar” to any of the former two, surprisingly complying with similarity requirements only with batch A3, mainly due to its particular data variability. Despite that the $f_1/f_2$ criteria suggest similarity between all of the dissolution profiles, this is somehow in agreement with conclusions reached by Tsong and co-workers on the grounds of Mahalanobis distance-based multivariate region specification criteria (Chen and Tsong, 1997), and on the basis of confidence intervals of the characteristic parameters ($\alpha$ and $\beta$) of a Weibull curve fit (Tsong et al., 1997). The bootstrap-calculated acceptable values of the corresponding $f_2$ test shown in Fig. 5 reveals that, being of a more permissive nature, the $f_2$ factor estimation also supports the conclusion of lot similarity.

The graphical result of the A1–A2 comparison (and those of A2–A3 and A1–A3 to a minor extent) can be attributed to higher tablet data variability within the latter lot, compared with the reference. Since the lengths of the axes of the ellipse are related to the eigenvalues of the covariance matrix, the confidence region is sensitive to variability of the reference data; therefore, it should be made possible for some samples of the test lot to remain outside the confidence region due to their own (and sometimes higher) variability, as proposed.
To take into account data variability is another important feature of the PCA-CR method, in sharp contrast with the $f_2$ criterion, which is a function of mean differences, and has been criticized for not computing variability within the test and reference data. Not without reason, authors have recommended careful interpretation of $f_2$ results when the variances of the individual profiles are very different (Saranadasa and Krishnamoorthy, 2005).

Considering that the confidence region in the proposed method represents a tighter standard than the $f_1/f_2$ indicators, the allowance of up to 20% of the samples to fall outside of the 95% confidence ellipse represents a compromise which transforms the proposed PCA-CR method into a less restrictive tool and a test procedure with pharmaceutical significance, still remaining diagnostic of “similarity”.

3.5 Pre- and post-change dissolution data of Shah et al. (1998)

The dissolution data of one pre-change and five post-change batches are shown in Fig. 6a, with the $f_1/f_2$ results of the post-change batches against the pre-change sample consigned in Table 4. Many results (three out of five for $f_1$ and four out of five for $f_2$) seem to be borderline ($f_1 > 13$ or $f_2 < 60$). However, despite the differences among the curves and according to the $f$-criteria, all of the post-change batches comply with the requirements for “similarity”.

In their study of this dataset employing bootstrap techniques, Ma et al concluded that, depending on the estimators employed, only batch B or batches B and E could be considered “similar” to the pre-change batch A (Ma et al., 2000). The PCA-CR analysis of the data was run after assuring absence of outliers. Interestingly, however, this revealed that even batch B does not meet the “similarity” requirements, displaying five out of its 12 data points out of the confidence ellipse.

Closer inspection of the dissolution curves of batches A and B indicated that the CV% of the curves of batch B at the different time points (10.6, 9.9, 5.7 and 1.6%) were different from those of the reference batch, being data dispersion of the latter comparatively smaller (6.7, 4.8, 3.8 and 2.9%). Tentatively, this can provide an explanation to the PCA-CR

<table>
<thead>
<tr>
<th>Table 4 – Comparison of the dissolution profiles of pre-change batch A with five post-change lots, according to the $f_1/f_2$ criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>f-Criterion/lots</td>
</tr>
<tr>
<td>$f_1$</td>
</tr>
<tr>
<td>$f_2$</td>
</tr>
</tbody>
</table>
non-similarity result, on account of the sensitivity of the method to data variability, particularly in the reference lot. Indeed, the multivariate test allows to conclude that despite the seemingly analogous shapes of the reference and test profiles, the individual dissolution curves in both batches being compared behave different, hence, “similarity” criteria could not be reached.

When dissolution profiles of pre-change batch A and post-change batch E were compared, it was evident that except for the first time point (where batch E also exhibits considerably less drug dissolved than batch A), both have similar CV% (15.0, 4.8, 3.8 and 2.8% for batch E); however, since PCA-CR is sensitive to shape and size of the dissolution curves (Adams et al., 2001), batch E was also correctly interpreted by the multivariate method as possessing a “non-similar” profile.

3.6. Method flexibility

The need for counting at least 80% of the test tablets (Chen and Tsong, 1997) inside of the 95% confidence ellipse of the reference lot constitute arbitrary criteria for assessing “similarity”, which relate to the strictness of the proposed method with regards to decision taking. In this sense, PCA-CR represents a more rigorous standard than the $f$-based comparison. However, the proposed approach is flexible enough, so these proposed specifications do not rule out alternative combinations of confidence levels for the ellipses and number of test tablets allowed to remain outside of the confidence region, which might be set according to the experience or specific needs. Bootstrap studies are suitable means to provide evidence for this fine-tuning of the method.

4. Conclusions

In summary, the use of the weighed scores plot of the relevant principal components of the dissolution curves with 95% confidence regions (PCA-CR) has been proposed as a new and alternative strategy for the comparison of in vitro dissolution profiles of tablet preparations. The results observed with this multivariate approach exhibited good qualitative correlation with the $f_1$ and $f_2$ values computed from the dissolution profiles; however, conclusions regarding profile similarity were not always coincident.

This was mainly due to the facts that the proposed method is more discriminating, taking into account data variability within the reference lot in order to build the confidence ellipses. Variations within the test lot, as well as shape and size of the dissolution curves have also influence on the final result.

Unlike the $f_1/f_2$ methods, based on comparison of data means, the use of confidence ellipses built upon PC values of the individual tablet dissolution curves of the reference set allows a simple and rapid graphic assessment of data distribution. In addition, the proposed approach does not impose
restrictions to useful data in terms of their variability and number of allowed time points above a given degree of dissolution; making use of the all the available information, avoids data-dependent outcomes, a characteristic feature of the f-based methods.

Compared to previously reported PCA-based methodologies, the SVD operation carried out on a single matrix containing test and reference data allows the optimization of system parameters in such a way that an improved projection of the test data in the joint reference–test data space is achieved.

The use of bootstrapping techniques for the representation of the $f_2 = 50$ frontiers in the PCA scores’ space, and their comparison with the region enclosed by the 95% confidence ellipses clearly demonstrated the relationship between the official and the proposed methodologies, revealing the potential of PCA-CR under the proposed conditions, as a stricter but also adaptable to a multiple stage acceptance rule, as given in the USP.

**Acknowledgements**

The authors gratefully acknowledge CONICET, ANPCyT and UNR. R.M.M. is also thankful to CONICET for his fellowship. Provision of key literature references by Dr. Y. Tsong (FDA, USA) is also acknowledged.

**REFERENCES**


