Biocybernetics and Biomedical Engineering
2010, Volume 30, Number 4, pp. 17–27

Exploratory Data Analysis Methods for Comparison of Drug Dissolution Profiles

DENIS ENĂCHESCU*

University of Bucharest, Faculty of Mathematics and Computer Science, Bucharest Romania
Institute for Mathematical Statistics and Applied Mathematics, Bucharest, Romania

A new approach for "similarity" testing through comparison of drug products dissolution profiles, based on multivariate data analysis is presented. The dissolution curves corresponding to three products containing oxicams (piroxicam, meloxicam and tenoxicam) as oral solid dosage forms were obtained by dissolution tests at multiple pre-specified time points and in different compendial media. Dissolution data was simultaneously subjected to principal component and cluster analysis and comparisons between the dissolution characteristics of different products were carried out. All the results were compared with information provided by the difference ($f1$) and similarity ($f2$) factor tests. Unlike the $f2$ criterion, the proposed methods reflect variability within the individual dissolution curves, being also highly sensitive to profile variations.

K e y w o r d s: multivariate data analysis, drug products dissolution profiles, principal component analysis, cluster analysis, oxicams

1. Introduction

The *in vitro* dissolution test has been recognized as an important step in assessing drug products' quality and, because its foreseen correlation with drug bioavailability, under strictly defined conditions, as a surrogate of *in vivo* studies for the assessment of product bioequivalence. Because it is essential to investigate the drug release characteristics of pharmaceutical formulation, dissolution has become highly significant and one of the primary pharmacopoeial tests that is performed to ensure that tablets, capsule and other drug products comply with the pre-established quality standards.

* Correspondence to: Denis Enăchescu, University of Bucharest, Faculty of Mathematics and Computer Science, str. Academiei 14, Bucharest 010014, Romania, e-mail: denaches@fmi.unibuc.ro
Received 23 February 2010; accepted 07 July 2010

In response to the need of "similarity" assessing, numerous strategies have been proposed for comparing of dissolution profiles (see, for example, [1] for a review of them). FDA Guidance for Industry and the European regulatory bodies recommend the difference (f_1) and similarity (f_2) factors introduced in [2] as mathematical indices to compare the dissolution profiles. In this approach, the dissolution behavior of samples (n) of reference (R) and test (T) products are compared at t time points (see the below equations).

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. As a variable simplification approach, in many cases two pharmaceutical equivalent drug products, two batches or two strengths of the same product are compared in respect their 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 the dissolution profiles (see [3]).

Here, we propose application of PCA with unit variance (PCA-UV), K-means and tree-clustering as new and alternative methods to compare the dissolution behavior of oral solid dosage forms (i.e. tablets) and decide about their "similarity". Usefulness of the suggested strategies was demonstrated by comparing different solid oral formulations of three highly representative oxicams. Although the similarity is usually assessed for different products containing the same active pharmaceutical ingredient (API) in the same strength and conditioned in the same dosage form, the present analysis hinted to reveal a certain group-behavior, highly representative for the Biopharmaceutical Classification System (BCS) Class II drugs (low solubility, high permeability; see [4]). The highly restrictive criteria of classification lead frequently to very heterogeneous groups, both from the point of view of *in vitro* and *in vivo* dissolution profiles. Therefore, analysis of different API dissolution within immediate-release drug products, including specific process such as disintegration, was imposed. For assessing the scope and limitations of the proposed approaches, the results were confronted in each case with the conclusions provided by the corresponding f_1 and f_2 factors considered as reference based on regulatory authorities and pharmacopoeial recommendations.

2. Methods

Denote by $\mathbf{X}_{2n\times p}$ the data matrix (each row contains the p time points of a dissolution curve and we consider that we have $2n$ distinct dissolution curves, n for the reference (R) and n for the test (T) products). Then

$$f_1 = 100 \times \left(\frac{\sum_{i=1}^{n} \left| \bar{x}_{R,i} - \bar{x}_{T,i} \right|}{\sum_{i=1}^{n} \bar{x}_{R,i}} \right);$$

$$f_2 = 50 \times \log \left\{ \left[\left(1 + \left(\tfrac{1}{n} \right) \sum_{i=1}^{n} \left(\bar{x}_{R,i} - \bar{x}_{T,i} \right)^2 \right) \right]^{-0.5} \times 100 \right\}$$

where $\bar{x}_{R,i}$ and $\bar{x}_{T,i}$ are the mean values of the row i corresponding to a (R) and (T) curves, respectively [2].

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* $0 \le f_1 \le 15$ [5].

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. *Acceptable values of f_2 are* $50 \le f_2 \le 100$*, which is considered equivalent to a difference in approximately 10% between the dissolution profiles being compared* [5].

The principles underlying PCA and cluster analysis have been extensively discussed elsewhere (see, for example [6]); the following is a brief description of the PCA-UV.

Denote by

Xs the standardized data matrix having zero means and unit variances;

R = corr(**X**) the correlation matrix derived from

$\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_p > 0$ the eigenvalues of **R**;

$\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p$ the corresponding eigenvectors.

Then, the Principal Components (called sometimes Principal Coordinates) are defined as new variables $\mathbf{c}_i$

$\mathbf{c}_i = \mathbf{Xs} * \mathbf{u}_i$, $i = 1,...,p$

Obviously:

mean($\mathbf{c}_i$) = 0 and var($\mathbf{c}_i$) = λ_i, $i = 1,...,p$.

If we consider the space generated by the orthogonal vectors $\{\mathbf{u}_i\}_{i=1}^{p}$, (called Factorial Space, FS) the variables $\{\mathbf{c}_i\}_{i=1}^{p}$ can be re-scaled to the variables $\{\mathbf{z}_i\}_{i=1}^{p}$ with unit variance:

$$\mathbf{z}_i = \frac{\mathbf{c}_i}{\sqrt{\lambda_i}}, \quad i = 1, \ldots, p \, .$$

Obviously:

$\text{mean}(\mathbf{z}_i) = 0$ and $\text{var}(\mathbf{z}_i) = 1$, $i = 1,...,p$.

Suppose now, that $\mathbf{c}_1,...,\mathbf{c}_p$ are i.i.d. normal variables. Then $\mathbf{z}_1,...,\mathbf{z}_p$ are also i.i.d. normal variables. It follows that the squared $\mathbf{z}_1^2,\ldots,\mathbf{z}_p^2$, which represent squared distances from the origin, are each distributed as χ_1^2 and their sum, D_p^2 is distributed as χ_p^2.

We draw a hyper-sphere in the factorial space with center *O* and radius the square root of the distance D_p^2:

$$D_{p;\alpha} = \sqrt{\chi^2_{p;\alpha}}$$

where $\chi^2_{p;\alpha}$ is the α-quantiles of χ_p^2.

Hence, we propose that: *dissolution profiles are similar with a probability of 95%* (*or 99%*) *if the points* $\left\{\left(z_{1j},\ldots,z_{pj}\right)^T\right\}_{j=1}^{2n}$ *lie in the hyper-sphere of radius* $D_{p;0.05}$ (*or* $D_{p;0.01}$).

K-means clustering is a method for finding clusters and cluster centers in a set of unlabeled data. One chooses the desired number of cluster centers, say *K*, and the *K*-means procedure iteratively moves the centers to minimize the total within cluster variance.

If $K = 2$ we can consider the following "similarity" measure: *all points in the cluster of the reference product* (*i.e. the cluster containing all, or the majority, reference points*) *are similar with this product.*

The results of applying the *K*-means clustering algorithms depend on choice for the number of clusters to be searched and a starting configuration assignment. In contrast, hierarchical clustering methods do not require such specifications. Instead, they require the user to specify a measure of dissimilarity between (disjoint) groups of observations and a strategy of amalgamation (linkage) of the groups.

Each level of the hierarchy represents a particular grouping of the data into disjoint clusters of observations. The entire hierarchy represents an ordered sequence of such groupings. It is up to the user to decide which level (if any) actually represents a "natural" clustering in the sense that observations within each of its groups are sufficiently more similar to each other than to the observations assigned to different groups at that level.

Recursive binary splitting/agglomeration can be represented by a rooted binary tree. The nodes of the trees represent groups. The root node represents the entire data set. The *n* terminal nodes each represent one of the individual observations.

All agglomerative methods possess a monotonicity property. That is, the dissimilarity between merged clusters is monotone increasing with the level of merger. Thus the binary tree can be plotted so that the height of each node is proportional to the value of the intergroup dissimilarity between its two daughters. The terminal

nodes representing individual observations are all plotted at zero height. This type of graphical display is called *dendrogram*.

One of the very efficient agglomerative methods is the Ward rule. It attempts to minimize the Sum of Squares (SS) of any two (hypothetical) clusters that can be formed at each step.

If the distance between groups is the Euclidean one, the linkage method is the Ward rule and the binary tree is scaled to a standardized scale (i.e., linkage distance / max(linkage distance)*100) then a "similarity" measure can be the following: *all points in the group of the reference product (i.e. the cluster containing all, or the majority, reference points) obtained by 'cutting' the tree at level 50 are similar with this product.*

The working hypothesis was that all drugs within the products have a similar dissolution behavior.

The proposed approach for the assessment of "similarity" through the multivariate data analysis of dissolution curves entails the following steps:

a) computing the indexes for f_1 and f_2;
b) plotting the dissolution profiles for each tablet from the compared brands;
c) computing the D_p^2 distances, construction of the Principal Factorial plane, PF, (i.e. the plane generated by the orthogonal vectors $\mathbf{u}_1$ and $\mathbf{u}_2$) and representing the projection of the dissolution profiles;
d) obtaining of two clusters in the set of the compared brands using the *K*-means method;
e) plotting the dendrogram of the dissolution curves.

All the computations were performed in Statistica v.8.

3. Results and Discussion

Three different immediate release oral solid dosage forms containing the highest available strength (i.e. Piroxicam 20mg –pi, Meloxicam 15mg –me, Tenoxicam 20mg –te) were studied. For each product, the *in vitro* dissolution characteristics were determined using 900 mL of three 100mM buffer systems (i.e. acetate buffer, pH = 4.5, phosphate buffer pH = 6.8, and 7.2). Each dissolution test was performed on six drug products units, using a Vankel 7000 system – USP Apparatus 2 (paddle), with manual sampling. The dissolution media was degassed and heated to 37°C prior to the test debut.

Hence we obtained 54 dissolution curves (labeled each as bbp_*n* with bb = pi/me/te, *p* = 4/6/7 – the first digit of the pH and *n* = 1,...,6 the tablet). Each dissolution curve contained a total of 7 time points (corresponding to the sampling schedule: 0, 5, 10, 15, 20, 30, 45 and 60 minutes).

From the data of Table 1, it follows that the f_1 and f_2 criteria confirm only in half of comparisons the working hypothesis of the pharmacists that the profiles are similar.

Table 1. Results of the pairwise comparison of nine dissolution profiles of the oxicams tablets employing the difference (f_1) and similarity (f_2) criteria. In this comparison the Piroxicam tablets are the reference products

		me_4.5	**me_6.8**	**me_7.2**	**te_4.5**	**te_6.8**	**te_7.2**
pi_4.5	f_1	79.5587			8.1872		
	f_2	38.2854			80.8564		
pi_6.8	f_1		18.2869			7.5646	
	f_2		44.5323			61.9470	
pi_7.2	f_1			28.4190			7.2182
	f_2			27.1717			55.0922

pi_4.5 = Piroxicam product, Acetate buffer pH = 4.5, 100mM; **pi_6.8** = Piroxicam product, Phosphate buffer pH = 6.8, 100mM; **pi_7.2** = Piroxicam product, Phosphate buffer pH = 7.2, 100mM.
me_4.5 = Meloxicam product, Acetate buffer pH = 4.5, 100mM; **me_6.8** = Meloxicam product, Phosphate buffer pH = 6.8, 100mM; **me_7.2** = Meloxicam product, Phosphate buffer pH = 7.2, 100mM.
te_4.5 = Tenoxicam product, Acetate buffer pH = 4.5, 100mM; **te_6.8** = Tenoxicam product, Phosphate buffer pH = 6.8, 100mM; **te_7.2** = Tenoxicam product, Phosphate buffer pH = 7.2, 100mM.

The plots of dissolution curves (see, for example, Fig. 1) confirm the working hypothesis only in the case (pi_6.8; me_6.8; te_6.8).

The PCA-UV method confirms in all cases the working hypothesis with the significance of 95% and the explained inertia greater than 87% (see Table 2 and, as example, Fig 2).

The results of both clustering methods (i.e. *K*-means and hierarchical clustering) are concordant with the f_1 and f_2 criteria. It confirms the working hypothesis in the case (pi; me) (see Tables 3, 4, 5 and, as example, Fig 3).

From biopharmaceutical perspective, the dissolution test results underline the weak acidic character, leading to a high percentage of API dissolved as the pH of the medium was increased to 6.8 and 7.2, respectively. Despite this group physico-chemical characteristic, Meloxicam seems to have a different behavior, dominated by lower solubility at pH=4.5, as well as different dose: solubility number [7].

4. Conclusions

In summary, the use of the multivariate data analysis techniques 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 approach exhibited good qualitative correlation with 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 methods are more discriminating, taking into account data variability within the reference lot. Variations within the test lot, as well as shape of the dissolution curves have also influence on the final result.

Table 2. The $D_{7;\alpha}$ and $D_{2;\alpha}$ distances of the dissolution profiles of the Oxicame tablets

pH = 4.5			pH = 6.8			pH = 7.2		
	$D_{7;\alpha}$	$D_{2;\alpha}$		$D_{7;\alpha}$	$D_{2;\alpha}$		$D_{7;\alpha}$	$D_{2;\alpha}$
me4_1	1.5234	1.3929	me6_1	2.5371	1.7311	me7_1	2.0282	1.9875
me4_2	2.3026	1.3928	me6_2	2.5778	1.4479	me7_2	3.7143	2.1460
me4_3	1.4492	1.3325	me6_3	2.6989	1.6330	me7_3	3.4009	1.4705
me4_4	1.7856	1.4295	me6_4	2.6587	1.6033	me7_4	2.2322	1.5340
me4_5	1.4110	1.3523	me6_5	1.9824	1.5490	me7_5	2.6454	1.6702
me4_6	1.6273	1.3247	me6_6	1.9769	1.8363	me7_6	2.6620	1.8348
pi4_1	3.6173	1.2900	pi6_1	2.7778	1.5634	pi7_1	2.2208	0.8917
pi4_2	3.0706	1.2982	pi6_2	2.2210	0.6415	pi7_2	2.4951	0.7859
pi4_3	1.8691	1.0742	pi6_3	2.8449	1.6235	pi7_3	2.7713	
pi4_4	2.4594	1.5759	pi6_4	3.2599	1.8497	pi7_4	2.3787	1.9331
pi4_5	2.9108	1.3579	pi6_5	3.6406	1.6059	pi7_5	1.8255	1.5435
pi4_6	2.2335	1.6102	pi6_6	2.7643	1.4782	pi7_6	2.0872	1.8303
te4_1	3.1035	1.4191	te6_1	2.6610	1.4539	te7_1	3.0245	2.2318
te4_2	1.9329	1.3096	te6_2	1.7850	0.9733	te7_2	1.9331	1.3347
te4_3	2.8073	1.4307	te6_3	1.6451	0.4642	te7_3	2.8775	1.8077
te4_4	2.6805	1.4028	te6_4	3.0550	0.3134	te7_4	2.2074	1.2885
te4_5		1.3750	te6_5	2.0359	0.5376	te7_5	1.8111	1.4599
te4_6	3.4904	1.2871	te6_6	2.2376	0.7455	te7_6	2.9918	0.8672

$D_{7;0.05} = 3.7506$; $D_{7;0.01} = 4.2982$; $D_{2;0.05} = 2.4472$; $D_{2;0.01} = 3.0300$

Unlike the f_1 and f_2 methods, based on comparison of data means, the use of the individual dissolution curves allows a simple and rapid graphic assessment of data distribution. In addition, the proposed approach does not impose restrictions on the number of lots to be compared and on useful data in terms of their variability and number of allowed time points above a given degree of dissolution; making use of all available information, avoiding of data-dependent outcomes, a characteristic feature of the *f*-based methods.

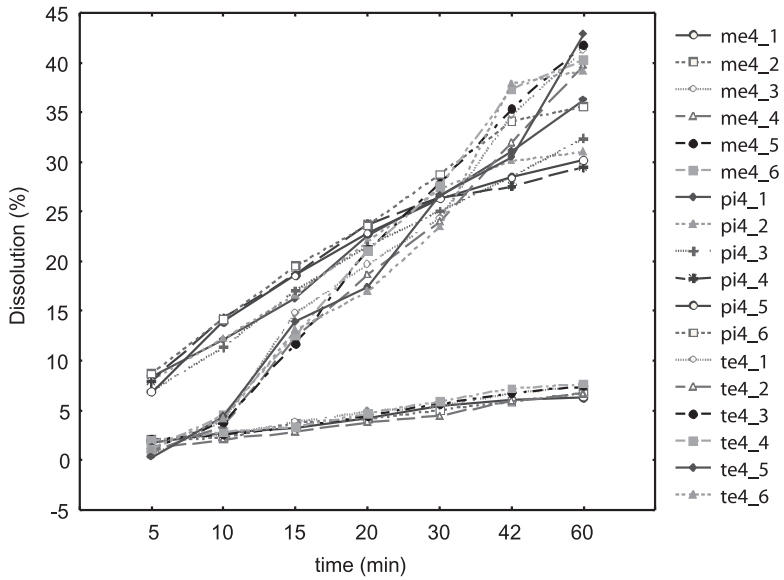

Fig. 1. Dissolution profiles of 18 tablets of oxicams, corresponding to three different products (Piroxicam –pi4_n, Meloxicam –me4_n, Tenoxicam –te4_n with n = 1,...,6 the tablet). The profiles were assessed in a medium of Acetate buffer 100mM pH=4.5

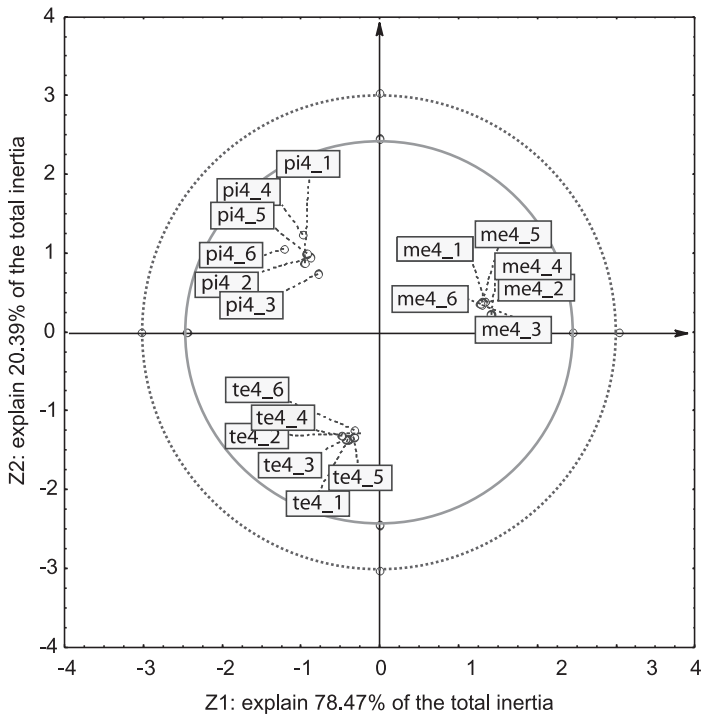

Fig. 2. Projection of the dissolution profiles in the PF plane explaining 98.86% of the total inertia. Confidence regions of 95% (solid line) and 99% (dashed line)

Frequently, the drug product manufacturers are lacking the official, compendial monographs, in order to accurately check quality of the certain drug product. Specific drugs or drug formulations are available only in a given area, either under FDA or EMEA authority. The easiest pathway is to implement an available pharmacopoeial recommendation for a physico-chemically related API.

Nevertheless, the validation of a specific dissolution methodology frequently requests the selection between various parameters, in order to generate the most discriminatory or robust conditions. Therefore, the use of multivariate data analysis techniques could be implemented for individual rather than mean dissolution profile evaluation (as a part of new dissolution test development procedure), as well as for drug products' dissolution profile comparison (as alternative to the pharmacopoeial metrics).

Table 3. The results of clustering the dissolution profiles into two groups using the K-means method. The members of the clusters are sorted in an ascending order with the distance to the respective cluster center

Cluster 1 contains 6 cases		Cluster 2 contains 12 cases	
	Distance		Distance
me4_5	0.2521	**pi4_1**	2.2108
me4_3	0.3205	**pi4_3**	2.7266
me4_1	0.3371	**te4_2**	3.1016
me4_2	0.3406	**pi4_2**	3.1089
me4_6	0.4911	**te4_1**	3.1127
me4_4	0.5631	**te4_4**	3.2665
		te4_3	3.4734
		pi4_6	3.5223
		te4_5	3.5970
		te4_6	3.7035
		pi4_5	3.8923
		pi4_4	4.4717

Table 4. The results of clustering the dissolution profiles into two groups using the K-means method. The members of the clusters are sorted in an ascending order with the distance to the respective cluster center

Cluster 1contains 11 cases		Cluster 2contains 7 cases	
	Distance		Distance
te6_6	3.427786	**me6_5**	3.6355
te6_5	3.777554	**me6_4**	4.3915
te6_4	4.244407	**te6_3**	4.9205
pi6_1	4.371218	**me6_1**	5.4445
pi6_2	4.456459	**me6_2**	5.5480
pi6_3	4.599834	**me6_6**	5.9484
te6_1	6.451660	**me6_3**	6.3534
te6_2	6.452507		
pi6_6	6.609734		
pi6_5	7.193142		
pi6_4	8.798158		

Table 5. The results of clustering the dissolution profiles into two groups using the K-means method. The members of the clusters are sorted in an ascending order with the distance to the respective cluster center

Cluster 1contains 6 cases		Cluster 2contains 12 cases	
	Distance		Distance
me7_4	1.5756	**te7_5**	1.1791
me7_6	2.0227	**te7_6**	1.8246
me7_1	2.4427	**pi7_1**	3.6716
me7_5	2.6789	**pi7_2**	3.7862
me7_2	4.5631	**pi7_5**	3.8445
me7_3	4.8823	**te7_2**	4.1331
		pi7_3	4.7356
		te7_4	4.8606
		pi7_6	6.0018
		te7_3	6.4072
		te7_1	7.5842
		pi7_4	8.5016

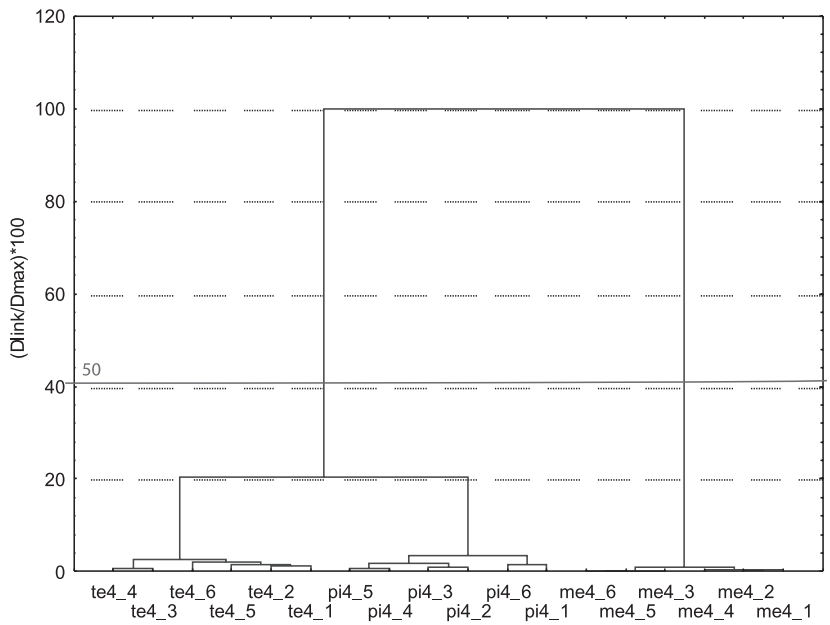

Fig. 3. Tree Diagram for 18 cases representing the dissolution profiles determined in a medium with pH = 4.5. The linkage method is the Ward rule and the distance between cases is the Euclidian one

Acknowledgment

This research was supported by Romanian PNCDI-2007-2013, 4th Domain-Health, and grant nr. 42–135.

Reference

1. O'Hara T., Dunne A., Butler J., Devane J.: A review of methods used to compare dissolution profile date. Pharm. Sci. Technol. Today 1998, 1, 214–223.
2. Moore J.W., Flanner H.H.: Mathematical comparision of curves with an emphasis on in-vitro dissolution profiles. Pharm. Technol. 1996, 20, 64–74.
3. Maggio R., Castellano P.M., Kaufman T.: A new principal component analysis-based approach for testing "similarity" of drug dissolution profiles. Eur. J. Pharm. Sci. 2008, 34, 66–67.
4. Amidon G.L., Lennernäs H., Shah V.P., Crison J.R. : A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. Pharm. Res. 1995, 12, 413–420.
5. FDA 2000: Guidance for Industry, Waiver of in Vivo Bioavailability and Bioequivalence Studies for Immediate Release Solid Oral Dosage Forms Based on a Biopharmaceutics Classification System. FDA/CDER,Washington, DC.
6. Enachescu D.: Data mining. Methods and Applications (in romanian). Ed. Academiei, Bucureşti 2009.
7. Rinaki E., Dokoumetzidis A., Macheras P.: The mean dissolution time depends on dose/solubility ratio. Pharm. Res. 2003, 20, 406–408.