DISCRIMINANT ANALYSIS AND MAHALANOBIS DISTANCE (NIR DIFFUSE REFLECTANCE SPECTRA) IN THE ASSESSMENT OF DRUG'S BATCH-TO-BATCH DISPERSION AND QUALITY THRESHOLD ESTABLISHMENT

Maria Morozova, PhD
Peoples' Friendship University of Russia, Russian Federation
Tatyana Elizarova, PhD
Control analytical lab "Farmanaliz" LLC, Russian Federation
Tatyana Pleteneva, PhD, professor
Peoples' Friendship University of Russia, Russian Federation

Abstract

Distance-based supervised pattern recognition method (PRM) – discriminant analysis (DA) in principal component (PC) space described by Mahalanobis distance (MD) - was suggested for definition of batch-to-batch dispersion of 22 drugs, produced by 14 manufacturers. The range of under test medicines numbered 18 pharmacological classes, including biologics, microbiological drugs and combined drugs, containing several active ingredients. Thermo Scientific ANTARIS II FT-NIR Analyzer collected spectra in diffuse reflectance mode, data pre-processing and analysis were performed using the TQ Analyst software. The value of spectral dispersion for the most samples in the research except capsules did not exceed 10 MD, what correlated with the values met in literature (Mark & Tunnell, 1985). Mahalanobis distance threshold, which was calculated in leave-one-out-cross validation (LOOCV), proved equal 3 for the majority of drugs, what was in a good agreement with the theoretically determined limit (Mark, 2001, p. 321).

Keywords: Near infrared spectroscopy, discriminant analysis, Mahalanobis distance, batch-to-batch dispersion, pharmaceuticals

Introduction

Since near infrared spectroscopy has been introduced as a general analytical method in the official pharmacopoeias of Europe and the United

States, it becomes a frequently used technique within pharmaceutical industry and drug quality control system (European Pharmacopoeia, 2011; The United States Pharmacopoeia, 2011). Together with multivariate statistics, NIR spectroscopy is used in a broad range of applications for precise material characterization: identification of active pharmaceutical ingredients (API), raw materials, products authentication, homogeneity analysis, control of polymorphs, optical isomers and counterfeit products screening (Blanco & Romero, 2001; Frasson Scafi & Pasquini, 2001; Plugge & van der Vlies, 1996; Rodionova & Pomerantsev, 2010; Sarraguc & Lopes, 2009). As for quantitative analysis, examples include the determination of physical parameters, water content and dose uniformity (Elizarova et al., 2011; Reich, 2005). More than that, NIR technique, due to its possibilities of direct and fast measurements, is actively implemented to the manufacturing application in concept of process analytical technology (Ritchie et al., 2003; Märk et al., 2010). Variety of NIRS utilization indicates the techniques obvious advantages, however still considerable attention should be paid to the existence of certain methodological difficulties (Blanco et al., 1998; Blanco & Alcala, 2005). In the framework of this research, those dealt mainly with the qualitative analysis and drug authentication are considered.

A proved matter is that drug identification from the mere visual inspection of its NIR spectrum requires much effort; rather, a pattern recognition method must be applied to a library of spectra for this purpose. An ideal pattern recognition method should be simple to use, easily understood by non-mathematician, should correctly identify unknown and reject unacceptable samples (Gemperline & Boyer, 1995). Supervised modeling methods are of a best choice for sample qualifying procedure, because they rely on some prior training of the system with objects known to belong to the determining class. Such methods can be of the discriminant or modeling types.

Discriminant methods split the pattern space into as many regions as the classes encompassed by the training set and establish bounds that are shared by the spaces. Discriminant analysis is one of the most common discriminant methods, allows attributing an unknown sample to a certain class, what is especially appreciable in pharmaceutical quality control. The bounds, established during the analysis, can be based on correlation coefficients and distances.

Distance-based methods possess a superior discriminating power and allow highly similar compounds (e.g. several production batches of one drug) to be distinguished (Blanco & Alcala, 2005). The most commonly used distance measures are the Euclidean distance (ED) and the Mahalanobis distance. Both distances can be calculated in the original variable space and in the principal component (PC) space. The ED is easy to compute and

interpret, but this is less the case for the MD (De Maesschalck et al., 2000). The discussion of the pros and cons of using ED or MD in the field of multivariate calibration is outside the scope of this paper (Brereton, 2007; De Maesschalck et al., 2000; Mark, 2001, p. 310). However, a review of number on-topic scientific papers demonstrated successful experience of Mahalanobis distance use in PRM techniques for pharmaceuticals classification (Andre, 2003; Gemperline & Boyer, 1995; Märk et al., 2010; Mark & Tunnell, 1985; Plugge & van der Vlies, 1996; Sarraguc & Lopes, 2009; Ritchie et al., 2003; Whitfield et al., 1987). Already in 1985 it was proposed: "Using MD for classification allows computer programs to be written that contain a large measure of protection against marginal conditions, in a form that is easily understood and evaluated by the non-mathematician" (Mark & Tunnell, 1985). Taking into account all of the above, near-IR spectroscopy together with DA and MD appears to be a good instrument for drug authentication. It is just as well, but for any quality control laboratory there arises a question - what distance (MD) is considered a threshold for the class of authentic drugs? There is no one certain answer. Most of the revised articles discussed the problem of proper MD control limits determination, as well as reported similar results and conclusions. The earliest paper suggested that MD represented a measure of standard deviation, therefore the theoretically boundary of a group was thought to be three standard deviations away from the groupmean (Mark & Tunnell, 1985). Nevertheless, the authors' practical experience indicated, that a more reasonable cutoff point due to the presence of variations in the size, shape, or orientation between the different groups, small changes in the physical nature of the samples, small drifts in the instrument, etc. lied at Mahalanobis distance between 10 and 15 (Mark & Tunnell, 1985). Whitefield et al. (1987) also considered a MD criterion of 3 for deciding whether the sample belongs to the specified population unjustifiably restrictive. They calculated theoretical values for confidence limits for MD, depending on the sample size and number of wavelengths used. Ritchie et al. (2003) showed, the values from the Whitefield's tables were in good agreement with the empirical MD values, obtained for all their calibration samples. According to their results, based on combined criteria, the threshold for MD was set at 4. At last, Märk et al. (2010) said, the confidence limit for quality control, depending on the definition of the model number degrees of freedom had to be set empirically, based on an extensive testing effort.

That is why our research is focused on the issue of determination the qualification thresholds for the classes of authentic medicines and their comparison with the theoretical and experimental limits. In a certain way drugs are produced with some tolerance in API concentration, possible variations in excipients, impurities, moisture, density, viscosity or particle

size, while their NIR spectra reflect the whole quality of the drug and any deviations in the manufacturing process (Rodionova & Pomerantsev, 2010; Blanco & Alcala, 2005). Therefore, it can be proposed to use the meaning of acceptable dispersion between production batches, expressed in MD, as the MD quality thresholds for identical drugs. In other words, the objective of this paper is to assess the batch-to-batch dispersion of 22 different pharmaceuticals using NIR reflectance spectroscopy technique and highly accurate, easy understandable supervised pattern recognition method, provided by commercially available software.

Test samples

Table 1 presents the test sample set. Pharmaceuticals for the research were selected from the archival samples of the quality control laboratory, involved in the procedure of drug quality declaration. The choice of samples in the archive was in line with the presence of the highest number of series (at least 6), manufactured nearly at one period of time. The medicines quality was preliminary guaranteed by successful passing of conventional tests: description, packing, labeling, identification, assay, purity, disintegration test, average tablet weight, uniformity of content. Analysis of each drug was held in accordance with the regulatory requirements using different chemical and physicochemical methods: identification reactions, thin-layer chromatography, HPLC, ultraviolet spectroscopy, titration analysis. All quality indexes fitted into the normative limits, proving the high level of samples quality.

Table 1. Test set of pharmaceuticals

№	**SAMPLE NAME, TYPE**	**ACTIVE INGREDIENT**	**DOSE, mg**	**MAIN EXCIPIENT**	**DOSAGE FORM**	**MANUFAC-TURER COUNTRY**	**BATCHES NUMBER /TABLETS (capsules) NUMBER**	**SPEC-TRA NUM-BER**
1	A1 original	chondroitin sulfate	500	talk	capsules	A/ France	16/48	144
2	B1 original	pipemidic acid	200	magnesium stearate, silicon dioxide colloidal, corn starch	capsules	B/ Slovenia	25/75	225
3	C1 generic	fluconazole	150	lactose monohydrate, starch, silicon dioxide colloidal anhydrous,	capsules	C/ Macedonia	18/54	162

				magnesium stearate				
4	C2 generic	clopidogrel hydrosulfate	75	mannitol, MCC, silicon dioxide, talc, stearic acid	film-coated tablets	C/ Macedonia	10/30	90
5	C3 generic	carvedilol	12,5	lactose, sucrose, povidone K25, methyl cellulose	tablets	C/ Macedonia	26/78	234
6	C4 generic	carvedilol	25		tablets	C/ Macedonia	26/78	234
7	C5 generic	trimetazidine dihydrochlo-ride	35	hypromellose, MCC, silicon dioxide colloidal, magnesium stearate	film-coated tablets	C/ Macedonia	19/57	171
8	C6 generic	amlodipine besylate	10	lactose monohydrate, povidone, crospovidone, calcium stearate	tablets	C/ Macedonia	36/108	324
9	D1 original	dried powder of the bacteria strain Bacillus cereus IP 5832	35	calcium carbonate, kaolin	capsules	D/ France	25/75	225
10	E1 generic	diclofenac sodium	200	lactose monohydrate, corn starch, povidone K30, sodium lauryl sulfate	coated tablets	E/ Serbia	8/24	72
11	F1 original	amoxicillin, clavulanic acid	875/ 125	Magnesium stearate, silicon dioxide colloidal	coated tablets	F/ UK	6/18	54
12	G1 generic	bisacodyl	5	lactose monohydrate, wheat starch, silicon dioxide colloidal	coated tablets	G/ Bulgaria	6/18	54

				anhydrous, talc				
13	G2 original	metamizol, pitofenon, fenpiverin	500/ 2/ 0,02	lactose monohydrate, wheat starch, talc, magnesium stearate	tablets	G/ Bulgaria	17/51	153
14	H1 generic	ferrous sulfate, ascorbic acid	360/ 60	magnesium stearate, povidone, polyethylene powder	coated tablets	H/ Sweden, UK	6/18	54
15	I1 original	chlorpyra-mine hydrochlo-ride	25	stearic acid, gelatin, sodium, talc, potato starch, lactose monohydrate	tablets	I/ Hungary	9/27	81
16	J1 original	lornoxicam	4	calcium stearate, sodium bicarbonate, MCC, calcium phosphate	coated tablets	J/ Austria	7/21	63
17	J2 original	lornoxicam	8		coated tablets	J/ Austria	8/24	72
18	J3 original	dry deproteinized derivative from calf blood	200	magnesium stearate, povidone K90, talc, cellulose	coated tablets	J/ Austria	18/54	162
19	K1 original	risperidone	2	lactose, corn starch, microcrystalline cellulose, hypromellose, magnesium stearate, silicon dioxide colloidal anhydrous	coated tablets	K/ Italy	10/30	90
20	L1 original	propyphena-zone, paracetamol, caffeine,	210/ 250/ 50/10	magnesium stearate, sodium lauryl	tablets	L/ Macedonia	23/69	207

		codeine		sulfate, silicon dioxide colloidal anhydrous, microcrystal line cellulose				
21	M1 original	methyl-prednisolone	4	lactose monohydrat e, corn starch, magnesium stearate, gelatin, talc, purified water	tablets	M/ Finland	19/57	171
22	N1 generic	metamizol	500	magnesium stearate, polyethylene glycol 4000.	tablets	N/ India	30/90	270

The manufacturer companies' names and drug trade names were intentionally assigned the codes. In total the survey analyzed 22 products of 13 European and one Indian manufacturer. The capital letter in the first column of table 1 indicated drugs' appliance to the manufacturer. Most of the manufacturers were represented by one medicine, which was given the name like –"letter"1. Only three companies - C, G and J - were represented by several samples with the names like - "letter"1, "letter"2 etc. The first column of table 1 also specified which of the pharmaceutical samples were original and which were generic. In whole we analyzed 12 original and 10 generic medicines.

Among the test set there were drugs with the same API, but in different dose – sample C3 and C4 contained 12,5 and 25 mg of carvedilol; sample J1 and J2 – 4 and 8 mg of lornoxicam correspondingly. The most of the tested pharmaceuticals were coated, film-coated or uncoated tablets; only 4 samples - A1, B1, C1 and D1 – were represented in encapsulated dosage form.

As it follows from the Table 1, test samples (22 pharmaceuticals) had various chemical structures and belonged to 19 pharmacological classes including quinolones, probiotics, antiaggregants, beta-blockers, antianginals, calcium channel inhibitors, nonsteroidal anti-inflammatory drugs, antibiotics, laxatives, antihistamines, neuroleptics, glucocorticoids and others (see table 1). The range of under test medicines included biologics, microbiological

drugs and combined drugs, containing several active ingredients (samples F1, G2, H1, L1).

Table 1 also indicated the API dose, main drug excipients and the number of series, the number of tablets / capsules, the number of recorded spectra.

NIR spectra acquisition

The spectra were collected by Thermo Scientific ANTARIS II FT-NIR Analyzer. All measurements were performed in diffuse reflectance operation mode using an integrating sphere module. Collecting the radiation from all dihedral angles, integrating sphere offered much higher reproducibility, minor spectral offset variability and less noise possibility caused by fiber optics 12. Each spectrum was the average of 16 scans performed in the spectral range of 4000-9000 cm^{-1}at 4 cm^{-1} interval. An internal gold reference was used for automatic background collections, ensuring consistent sampling and repeatable results. To avoid the influence of blisters and glass bottles all samples were removed from their packaging. The tablets were measured by directly placing on the window surface of the integrating sphere; spectra were acquired from the both sides of the tablet, which was overturned before every measurement. The contents of capsules, emptied from opaque capsule shells, were put in a special glass cup with tight-fitting cap. To minimize scattering and reflection effects NIR spectra of capsules content were acquired after a standardized compaction procedure. Manufacturing variability was considered by including samples spectra from different production batches. NIR spectra were obtained 3 times per tablet/capsule for any 3 tablets/capsules from each batch, thus turned to 9 spectra for each batch. Thereby the measurement procedure fully satisfied the demands of the European Medicines Agency (EMA), which indicated the formation of the spectral library using three or more NIR spectra for at least three batches of pharmaceutical 22.

Data pre- processing

To decrease the influence of various sources that are not related to the chemical or physical information carried by raw spectra we used usual pre-processing technique, such as multivariate scatter correction (MSC). The MSC pathlength treatment was effective to compensate variations in sample thickness caused by particle size and scattering. Operation of MSC was performed both for the spectra of tablets and capsules content.

Software

The instrument was governed via a PC, using the RESULT Operation software from Thermo Electron Scientific Instruments Corporation both to acquire data and to construct the spectral library. Data pre-processing, analysis and results evaluation were performed using the TQ Analyst software, also supplied by Thermo Scientific.

Modelling

We determined manufacturing variability by estimating the spectral differences between the batches of a certain sample with one of the supervised pattern recognition methods and various options of spectra grouping. As the discriminant analysis (DA) classification technique had been proven to determine the classes of known materials that are most similar to the tested sample, this technique was considered to be suitable for the solution of designed problem. For the numerical assessment of manufacturing variability the Mahalanobis distances measurement between batches was used. The research covered all the spectral range, so the number of original variables had to be reduced by the principal component analysis. In all cases the extraction of first 10 PCs provided a sufficient degree of cumulative explanation of 99.9% of the calibration set variables. Thus the obtained values of MD, calculated in the PC space, allowed ranking test classes and determining the threshold of quality differences between manufacturing batches.

Spectral libraries and calibrations were formed separately for each sample drug. The first calibration set was constructed by the following principle: all spectra of every batch were included in one class, thus calibration set for every pharmaceutical contained the number of classes equal to the quantity of production batches. As the discriminant analysis was applied to the constructed data set it resulted into computation of pairwise distances between every class, measured in MD. This approach made it possible to reveal the highest spectral differences between separate lots in the recall of one type medicine.

The second grouping was a kind of LOOCV: all spectra of a single batch were used as the validation data, while the remaining spectra of the rest series as the training data (library). This was repeated such that spectra of each batch were used once as the validation data. LOOCV access enabled to: 1) evaluate the range of maximum spectral dispersion values in MD units; 2) set the value of quality threshold (MD), indicating the absence of reliable differences between the samples.

Detailed results of Discriminant analysis, obtained using different approaches to grouping of original data are represented further.

Batch-to-batch dispersion of capsules: pairwise distances between batches in simultaneous comparing

In the course of discriminant analysis all recorded capsules' spectra were included in the table of standards within TQ Analyst program: 144 spectra for sample A1, 225 spectra for sample B1, 162 spectra for sample C1 and 225 spectra for sample D1. Chemometric processing of NIR spectra allowed converting the spectral data and projecting it onto a subspace of smaller size, which was presented in the form of two-dimensional graphics in

Mahalanobis coordinates. For example, Figure 1 shows the result of discriminant analysis for sample A1 - pairwise distance graph between the separate batches of drug A1.

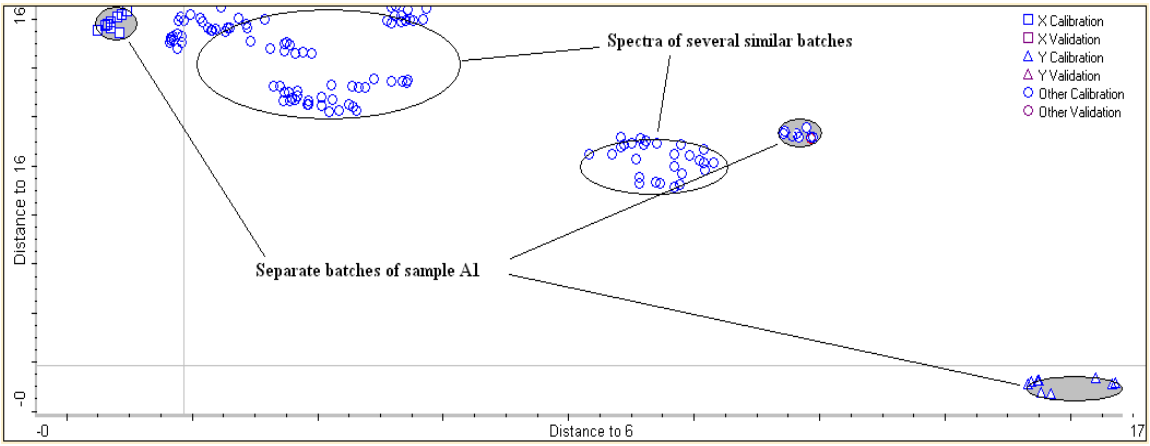

Figure 1. Pairwise distances for sample A1: squares - batch 8, triangles - batch 16, circles - the rest of the series

Each point of the resulting plane on Figure 1 corresponds to a particular spectrum of sample A1, at that triangles indicate the spectra of batch, selected by the vertical axis; squares denote the spectra of class selected on the horizontal axis. As can be seen from Figure 1, spectra of other batches, indicated as circles, fall in the resulting range between two axial classes. Gray lines, perpendicular to the axis, are the borderlines showing 95% confidence intervals for two selected classes. Points' mixing in the central part of the plane characterizes NIR spectra similarity, and consequently similarity of manufacturing batches. In some cases reliable discrimination of production batches is almost impossible. At the same time, the graph shows that the spectra, corresponding to the axial series, form two distinct clusters of points at a maximum distance from each other. Changing the batches on X and Y-axis, we determined the maximum pairwise distances between spectra of all tested batches for sample A1. Obtained values are presented in Table 2.

Table 2. Maximum spectral distances between individual batches of sample A1

Batch number	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	8.79	10.87	13.38	8.17	12.47	12.73	11.04	11.75	9.98	11.04	13.76	13.76	13.76	13.95	12.31
2		4.11	7.13	2.83	5.80	5.85	5.05	9.99	5.55	6.26	9.51	9.76	10.80	10.80	14.40
3			4.94	3.97	4.41	4.68	4.53	10.69	5.18	5.82	7.33	7.93	11.11	9.74	15.98
4				5.68	3.24	3.63	3.94	11.20	4.31	4.97	5.44	5.52	11.20	10.98	17.03
5					5.29	5.44	4.34	9.44	3.37	4.78	7.94	8.29	9.98	9.94	13.46
6						2.92	1.99	10.71	4.14	4.43	5.94	6.08	10.70	10.71	17.12
7							3.01	11.43	4.91	5.59	6.79	6.85	11.43	11.36	17.21
8								10.68	3.56	4.52	6.32	6.32	10.68	10.68	16.39
9									8.91	7.93	12.95	12.41	3.73	3.68	10.62
10										2.57	5.59	5.59	9.25	9.25	14.13
11											6.49	6.25	7.96	7.96	13.82
12												2.95	13.00	13.00	**17.38**
13													12.26	12.35	17.01
14														3.64	11.97
15															11.80

Table 2 indicates that the largest extension of spectral distance for sample A1 was detected between the series 12 and 16 - 17.38 MD. None of these batches were considered an outlier, since the distance between the pair of them and all the other classes didn’t exceed the average meaning of dispersion in the table.

For discriminant analysis of other encapsulated samples, the same procedures were conducted: we converted the spectral data and projected it onto a subspace of smaller size, determined the largest extension of spectral distance between the batches and excluded the existence of outliers. The results are performed on Figure 2 (black color); batch-to-batch dispersion of different capsules turned to be in the range from 10.14 to 24.32, what exceeded the limit of 10-15 MD, mentioned in Mark & Tunnell work (1985).

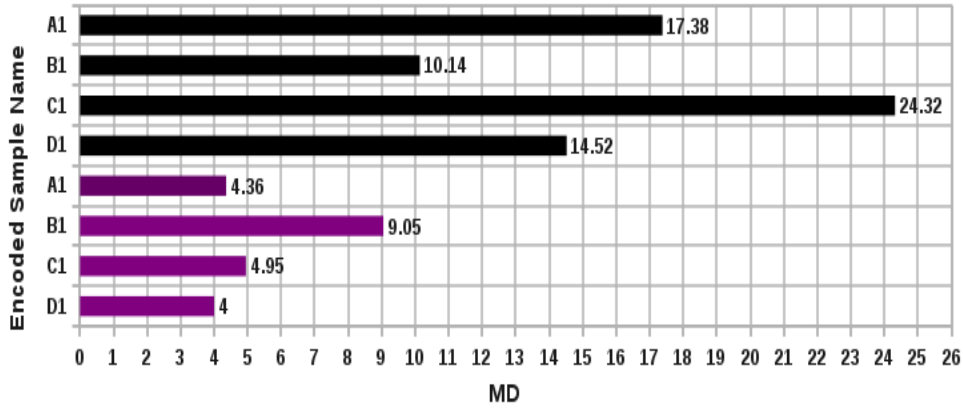

Figure 2. Maximum spectral distances between individual batches of capsules: black – NIR spectra and MSC, purple – NIR spectra first derivative and MSC

The obtained meanings were considered rather high and potentially risky as they could lead to false-negative results in the drug identification. When we measured capsules content, emptied from the coatings, the influence of variations in sample's thickness, size, shape, orientation between different batches was greater than in tablet's case. Tablets were more homogeneous samples, as their characteristics, such as size, thickness, surface type, uniformity of content were certified by manufacturer. In order to minimize physical differences in capsule samples, fist-derivative pre-processing with Savitsky-Golay filter (7 points, polynomial order 3) was additionally applied to the spectra of capsules. As a result of this procedure, the values of batch-to-batch dispersion decreased (Figure 2, purple color) and fitted to the indicated in research of Mark & Tunnell (1985) threshold of 10-15 MD.

Batch-to-batch dispersion of tablets: pairwise distances between drug batches in simultaneous comparing

The procedure of DA calibration for tablet samples was just the same as for capsules. But the amount of test drugs was higher, there were several samples from one manufacturer, medicines containing the same API with different dose, single-component and combined preparations (see Table 1). Figure 3 displays the results of discriminant analysis for the tablet calibration set, where each batch of a drug was considered as a separate class.

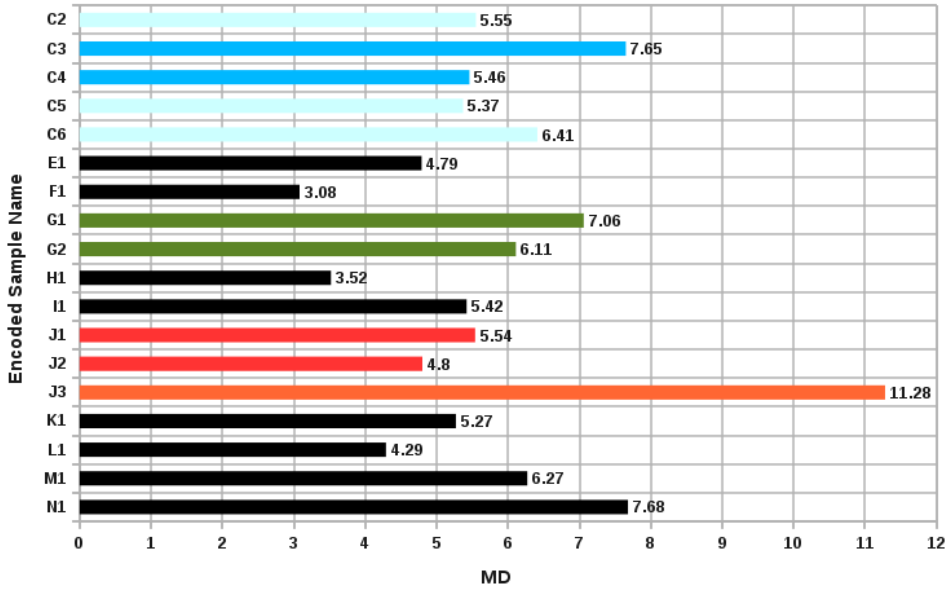

Figure 3. Maximum spectral distances between individual batches of tablets

Performed MD computations allowed the determination of maximum spectral distance between separate batches; these values were represented in the graph (Fig.3). Different colors indicated drugs of one manufacturer (samples with letters C, G, J). Integrally the range of spectral distances between the individual batches appeared from 3.08 (sample F1) to 11.28 (sample J3), that correlated with the results reported by Mark & Tunnell (1985). The value of spectral dispersion for most of the samples did not exceed 10 MD, except the sample J3. We supposed the reason was in the difficulties, met during the process of biological raw material standardization. Sample J3 is a high-purity deproteinized calf blood derivative, obtained by dialysis and ultrafiltration. The drug contains a set of micro-and macronutrients, amino acids, oligopeptides, nucleosides, fatty acids, oligosaccharides, which could vary resulting from deviations in quantitative composition of raw calf blood.

It seemed interesting to evaluate the batch-to-batch dispersion of pharmaceuticals, made by one producer and containing the same active ingredient but in different doses. This was achieved in course of discriminant analysis of spectral data, provided by samples C3, C4 and J1, J2. Analysis showed that in whole batch-to-batch variability in all mentioned samples (C3 and C4, J1 and J2) was close in value – nearly 6 MD (Fig.3). Although it became evident, that pharmaceuticals with lower doses had tendency to higher values of dispersion. Table 3 provided some data about samples, containing carvedilol and lornoxicam.

Table 3. API concentration and batch-to-batch dispersion (MD) for the samples C3, C4, J1, J2

API	**carvedilol**		**lornoxicam**	
Encoded sample name	**C3**	**C4**	**J1**	**J2**
API dose, mg	12.5	25	4	8
Tablet average weight, mg	180	180	215	211
API concentration, %	6.94	13.9	1.86	3.79
Batch-to-batch dispersion, MD	7.65	5.46	5.54	4.80

One can notice that with the increasing of API doses the tablet weight remained practically constant. That leaded to the growth of API concentration in tablet. Therefore, fluctuations of API distribution among excipients in the tablet with smaller dose could be the reason for the larger manufacturing variability of the former.

While completing the resulting tables with the meanings of maximal pairwise distances between the batches of tableted samples, we faced one observation that was more interesting. Discriminant analysis of tablets G1, G2, I1, and K1 allowed getting some latent information, characterizing batch-to-batch and in-batch dispersion of sample's spectra in the two-dimensional plane of principal components space. As shown in Figure 4 spectra of sample G1 were arranged in groups of three.

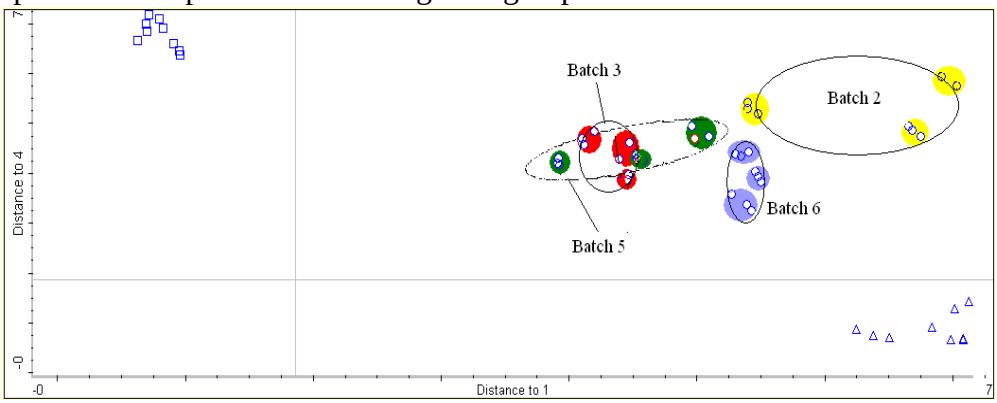

Figure 4. Results of discriminant analysis for the sample G1: comparison of batch-to-batch and in-batch dispersion

Each group of three points indicated the NIR spectra of sample G1, recorded for a single tablet from production batch. Figure 4 demonstrated that the difference between spectra in a single class was higher than dispersion between classes, some of which even overlapped with each other (batch 3 and 5). The same findings, when in-batch spectra dispersion exceeded the batch-to-batch dispersion, were also obtained for samples G2, I1, and K1. We considered these results were due to such technological

features of tablets manufacturing process as the mixing quality and flowability parameters of granulated material. Thus, while the drug technology from batch-to-batch was reproduced at a high level, the quality indexes of individual tablets within the same batch varied greatly, but stayed within normative demands.

To summarize all of the above, the general parameters of batch-to-batch dispersion despite some differences were comparable with each other for all of the investigated drugs. The absence of reliable qualitative differences between the series of tested drugs proved the high level of quality and reproducibility of tested products. The value of spectral dispersion between the batches of the samples did not exceed 10 MD, except the encapsulated samples and tablets with blood derivative, what correlated with the values met in literature [19]. The obtained results of spectral range magnitude between batches (MD) were thought the unique characteristics of each product. This measure specified the batch-to-batch reproducibility for the drug - the smaller distance between the spectral classes (series), the homogeneous product. In prospect these established in discriminant analysis numeric indexes of spectral dispersion could be used as an additional quality indicators. The latter could be incorporated into industry standards and regulations, where the manufacturer, depending on the possibility to reproduce productive conditions, would set the optimal value of the permissible batch-to-batch dispersion. This will allow avoiding the labor-consuming operations for the construction and constant updating of spectral library with a large sampling.

Leave-one-out-cross-validation (LOOCV) results: the range of maximum spectral dispersion and the meanings of quality thresholds (MD)

In the practice of pharmaceutical analysis, NIR spectroscopy is often used to confirm the authenticity of the drug. Herewith it is necessary to compare the spectra of under tested drug with a library of standards spectra and draw a conclusion about the sample quality. As it was mentioned before discriminant analysis together with Mahalanobis distance is suitable for that. The only difficulty is that proper control limits (MD), under which the sample meets qualification criterion, must be pre-set. So that was the purpose of our second type calibration set construction – LOOCV, which provided more stable MD results, almost uniform for all the tested drugs. The new values of dispersion were considerably less than the values, obtained by simultaneous comparison of separate product batches (Fig. 2, 3). Mathematical features of discriminant analysis can explain this: during the calibration, software calculated the average spectrum for each comparison class and created a model of distribution by measuring the deviations at each frequency range. The averaged spectrum was then subtracted from each spectrum providing the range of dispersion. For this reason, the association

of all batches' spectra in one common class significantly changed the averaged spectrum, and consequently the results of final comparison between tested drug and spectral library.

During the process of calibration, the result tables with the meanings of pairwise distances were constructed for all the samples. They included the information about the maximum value of dispersion in MD between every batch and the library class. The minimum and maximum values for each sample were extracted from these result tables to obtain the range of peak dispersion (Fig. 5).

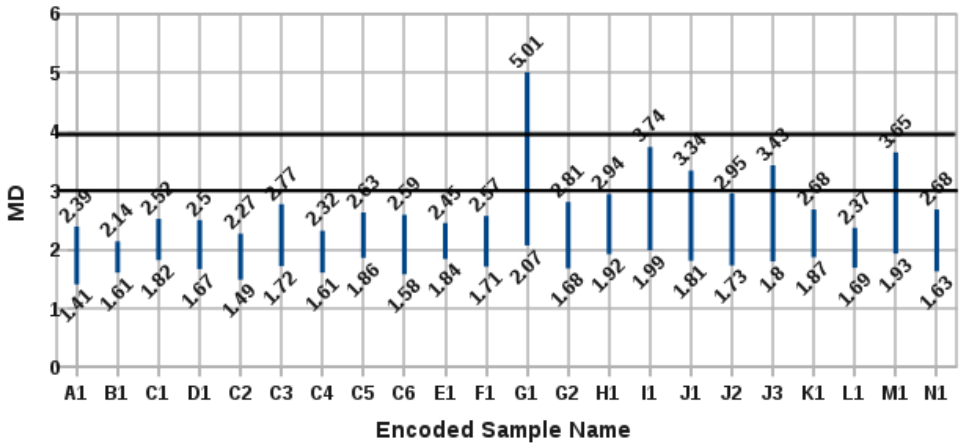

Figure 5. LOOCV results: the range of maximum spectral dispersion values (MD)

Thereby, in such a comparison for the majority of the being tested drugs (77.3 %) the magnitude of spectral differences fitted into 3 MD (Fig. 6). For the 18.2 % of the investigated drugs the range of peak dispersion did not uphold 4 MD, and for the 4.5 % - was more than 4 MD (Fig. 6).

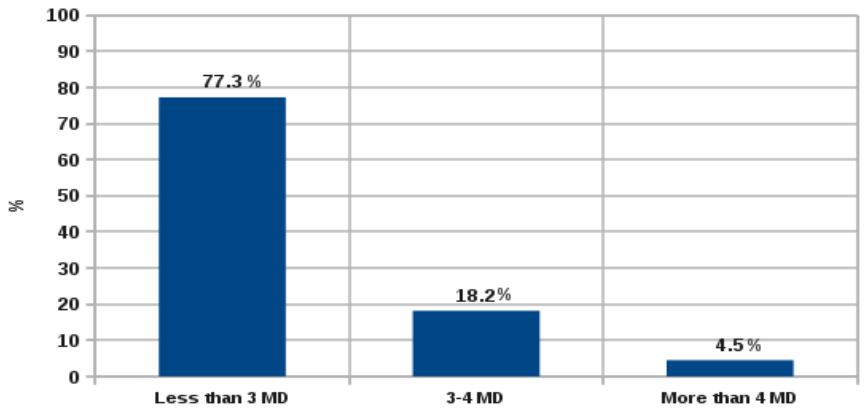

Figure 6. Distribution diagram: investigated drugs in accordance with different ranges of maximum spectral distance between batches

Therefore, the method enabled to determine the proper control limits on the MD, under which the sample meets qualification criterion. The results were largely coincided with the theoretical – 3 MD (Mark, 2001, p. 321) and experimental data – 4 MD (Ritchie et al., 2003).

Conclusion

The results of the investigation displayed the use of discriminant analysis together with the MD is an accurate mean for the determination of batch-to-batch drug dispersion. The value of spectral dispersion for the most of samples in the research did not exceed 10 MD, except the encapsulated samples and tablets with blood derivative, what correlated with the values met in literature (Mark & Tunnell, 1985). For the medicines authentication we recommended the introduction of an additional indicator of spectral variation, which was determined empirically in the result of LOOCV. Evaluated so Mahalanobis distance threshold, indicating the absence of reliable differences in qualitative characteristics, proved to be equal 3 for the majority of drugs, what was in a good agreement with the theoretically determined limit (Mark, 2001, p. 321).

References:

Andre M. “Multivariate Analysis and Classification of the Chemical Quality of 7-aminocephalosporanic Acid Using Near-Infrared Reflectance Spectroscopy”. Anal. Chem. 75 (2003).

Blanco M., Alcala M. “Use of near-infrared spectroscopy for off-line measurements in the pharmaceutical industry” in Process analytical technology, Katherine A. Bakeev ed. Oxford: Blackwell Publishing Ltd, 2005.

Blanco M., Coello J., Iturriaga H., Maspoch S., de la Pezuela C. “Near-infrared spectroscopy in the pharmaceutical industry”. Analyst 123 (1998).

Blanco M., Romero M. A. “Near-infrared libraries in the pharmaceutical industry: a solution for identity confirmation”. Analyst 126 (2001).

De Maesschalck R., Jouan-Rimbaud D., Massart D. L., “The Mahalanobis distance”. Chemom. Intellig. Lab. Syst. 50 (2000).

Elizarova T. E., Morozova M. A., Pleteneva T. V. “Possibility of using near-IR spectroscopy for drug quality control with respect to dose uniformity”. Pharmaceutical chemistry journal 45 (5) (2011).

European Pharmacopoeia, 7 ed. Strasburg: Cedex, 2011.

Frasson Scafi S. H., Pasquini C. “Identification of counterfeit drugs using near-infrared Spectroscopy”. Analyst 126 (2001).

Gemperline P. J., Boyer N. R., “Classification of Near-Infrared Spectra Using Wavelength Distances: Comparison to the Mahalanobis Distance and Residual Variance Methods”. Anal. Chem. 67 (1995).

Mark H. “Qualitative Discriminant Analysis” in Handbook of Near-Infrared Analysis, Burns D. A., Ciurczak E. W. eds. New York: Marcel Decker, 2001.
Mark H. L., Tunnell D. “Qualitative near-infrared reflectance analysis using Mahalanobis distances”. Anal. Chem. 57 (1985).
Märk J., Andre M., Karner M., Huck C. W. “Prospects for multivariate classification of a pharmaceutical intermediate with near-infrared spectroscopy as a process analytical technology (PAT) production control supplement”. Eur. J. Pharm. Biopharm. 76 (2010).
Plugge W., van der Vlies C. “Near-infrared spectroscopy as a tool to improve quality”. J. Pharm. Biomed. Anal. 14 (1996).
Reich G. “Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications”. Adv. Drug Deliv. Rev. 57 (2005).
Richard G. Brereton. Applied Chemometrics for scientists. Chichester: John Wiley & Sons Ltd, 2007.
Ritchie G. E., Mark H., Ciurczak E. W. “Evaluation of the Conformity Index and the Mahalanobis Distance as a Tool for Process Analysis: A Technical Note”. AAPS Pharm. Sci. Tech. 4 (2) (2003).
Rodionova O. Ye., Pomerantsev A. L., Houmøller L., Shpak A., Shpigun O. “Noninvasive detection of counterfeited ampoules of dexamethasone using NIR with confirmation by HPLC-DAD-MS and CE-UV methods”. Anal Bioanal Chem 397 (2010).
Rodionova O.Ye., Pomerantsev A.L. “NIR-based approach to counterfeit-drug detection”, Tr. Anal. Chem. 29(8) (2010).
Sarraguc M. C., Lopes J. A. “Quality control of pharmaceuticals with NIR: From lab to process line”. J. Vib. Spec. 49 (2009).
The European Medicines Agency. Note for guidance on the use of near infrared spectroscopy by the pharmaceutical industry and the date requirements for new submissions and variations. //http://www.emea. europa.eu/pdfs/human/qwp/330901en.pdf
The United States Pharmacopoeia 34-th ed. Toronto, 2011.
Whitfield R. G., Gerger M. E., Sharp R. L. “Near-infrared spectrum qualification via Mahalanobis distance determination”. J. Appl. Spectrosc. 41 (1987).