Journal of Pharmaceutical and Biomedical Analysis 54 (2011) 141–147

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Effect of the statin therapy on biochemical laboratory tests—A chemometrics study

Tatiana Ďurčeková[a,*], Ján Mocák[a], Katarína Boronová[a], Ján Balla[b]

[a] Department of Chemistry, University of SS Cyril and Methodius, J. Herdu 2, Trnava, SK-917 01, Slovak Republic
[b] Analytical-Diagnostic Laboratory, Kováčska 15, Prešov, SK-080 01, Slovak Republic

ARTICLE INFO

Article history:
Received 3 June 2010
Received in revised form 21 July 2010
Accepted 22 July 2010
Available online 6 August 2010

Keywords:
Cardiovascular risk
Statins
Biochemical tests
Multivariate data analysis
ROC analysis

ABSTRACT

Statins are the first-line choice for lowering total and LDL cholesterol levels and very important medicaments for reducing the risk of coronary artery disease. The aim of this study is therefore assessment of the results of biochemical tests characterizing the condition of 172 patients before and after administration of statins. For this purpose, several chemometric tools, namely principal component analysis, cluster analysis, discriminant analysis, logistic regression, KNN classification, ROC analysis, descriptive statistics and ANOVA were used. Mutual relations of 11 biochemical laboratory tests, the patient's age and gender were investigated in detail. Achieved results enable to evaluate the extent of the statin treatment in each individual case. They may also help in monitoring the dynamic progression of the disease.

1. Introduction

Statins are selective inhibitors of the 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, the rate-limiting enzyme in the biochemical cascade of cholesterol biosynthesis [1,2]. Their predominant action is to reduce circulating levels of low-density lipoprotein (LDL) cholesterol; to a smaller degree, they also increase high-density lipoprotein (HDL) cholesterol and reduce triglyceride concentrations [3,4]. Statins have been demonstrated to significantly affect the prognosis and outcome of patients with risk factors to atherosclerosis. Several studies have suggested an extra-beneficial effect of the statins in the prevention of atherosclerosis and coronary artery disease [5].

Three kinds of statins were used in the drugs dosed during the therapy: Simvastatin, Atorvastatin and Rosuvastatin. Atorvastatin and Simvastatin [6] have similar effects on serum triglyceride, total cholesterol, and LDL cholesterol levels. Both drugs increase HDL cholesterol levels, but the effect of Simvastatin is considered significantly greater than that of Atorvastatin [7]. Rosuvastatin has been shown to produce large, dose-dependent reductions in LDL cholesterol and have beneficial effects on other lipid variables in hypercholesterolemic patients [8]. Although all statins share a common mechanism of action, they differ in terms of their chemical structures, pharmacokinetic profiles, and lipid-modifying efficacy [9].

At present it is well known that the treatment by statins lowers the levels of total cholesterol and its LDL fraction. However, it is not sufficiently known what changes may be expected in the level of other lipid markers, which may provide a complex information about the effect of the statin treatment. It is also not clear enough whether some side effect may be coupled with the statin treatment during a longer time. Due to the risk of unintended adverse side effects the patients who are prescribed statins should be closely monitored. A number of statins may raise the risk of liver dysfunction, acute renal failure, myopathy (diseases of muscle), and cataracts. Therefore regular blood checks are necessary to ensure the value of cholesterol is (and further lipid markers are) at a satisfactory level and the statins medication is not affecting the vital function [10].

With regard to the above-mentioned facts the aim of this work is (1) to investigate the changes in the concentration levels of all frequently monitored lipid markers (tCHOL, HDLc, LDLc, TG) and their combination, like aterogenity index, (2) to show the effect of statins upon the selected standard biochemical tests, which monitor especially the function of liver (ALT, AST, ALP, GMT), kidneys (CREA), and are related to the heart activity or may indicate the muscle dystrophy (CK) (all medical abbreviations are explained in Section 2.1).

* Corresponding author.
E-mail address: tatjana@centrum.sk (T. Ďurčeková).

0731-7085/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2010.07.047

2. Material and methods

2.1. Description of laboratory data

Patient data were obtained in collaboration with out-patient doctor – internist who is prescribing statins for his/her patients. The same individuals were evaluated before and after the treatment; the after treatment evaluation was 12 months since the time when the treatment started. Fulfilling these conditions was possible due to *a posteriori* selection of the patients results from a larger documentation in such a way that all selected results were complete. The medical laboratory providing the laboratory testing has been accredited according to the ISO 17025 and strictly followed particular requirements for quality and competence.

This study includes serum levels of the selected biochemical tests of 172 patients with lipoprotein metabolism failure, other kind of lipidaemia and with further diseases: essential hypertension (54% of men, 22% of women), ischemic heart disease (35% of men, 52% of women), pancreas disease (28% of men, 26% of women), atherosclerosis (13% of men, 16% of women), hypertension (11% of men, 14% of women), hepatopathy (22% of men), asthma (11% of men), acute myocardial infarction (one men, one women), angina pectoris (one men), heart failure (one men), varicose vein (12% of women).

The results for 84 men samples and 88 women samples were transferred into two basic tables, one for men another for women These tables contained the sample origin, i.e. the patients, in the rows and the determined biochemical tests and age were in columns. The measured column values are: (a) concentration of five lipid parameters – total cholesterol (designed in italics as *tCHOL* when used as the variable in calculations in the following text), high-density lipoprotein cholesterol (*HDLc*), low-density lipoprotein cholesterol (*LDLc*), triacylglycerols (*TG*), and aterogenity index (*AI*) given by the ratio (*tCHOL* − *HDLc*)/*HDLc*, (b) concentration of six standard biochemical parameters – creatinine (*CREA*), aspartate aminotransferase (*AST*), alanine aminotransferase (*ALT*), alkaline phosphatase (*ALP*), creatine kinase (*CK*), and gamma-glutamyl transferase (*GMT*), and, finally, (c) the age of the patient. Blood samples were drawn before statin treatment and 1 year after duration of the statin therapy; the patient's gender represents another difference among the samples.

The patients were treated with the following drugs: (a) Simvor (containing Simvastatin) – 54% of men, 48% of women, (b) Torvacard (Atorvastatin) – 28% of men, 26% of women, (c) Tulip (Atorvastatin) – 9% of men, 14% of women, (d) Simvacard (Simvastatin) – 4% of men, 12% of women 20 mg daily. One patient was treated with Crestor (Rosuvastatin) – 10 mg daily and one with Sortis (Atorvastatin) – 20 mg daily. In conjunction to statins, additional drugs were administered to the patients if necessary – hepatoprotectives, antihypertensives, cardiotonics, beta-blockers, ACE inhibitors, vasodilators, digestives, antiasthmatics, antidepressants, bronchodilators, antiuretics, analgesics and cytostatics.

2.2. Statistical data analysis and description of multidimensional methods

Four software packages were employed for statistical calculations: (1) SPSS 15.0 to perform PCA, discriminant analysis, logistic regression, ROC analysis, ANOVA and correlation analysis; (2) SAS Enterprise Guide 3.0 to carry out discriminant analysis and KNN classification; (3) Statgraphics Plus 5.1 for cluster analysis. Microsoft Excel 2003 was used for preparation of the data in the appropriate form and their subsequent processing.

Principal component analysis, PCA, [11] is a basic way of characterizing multidimensional data, providing a satisfactory representation of the studied objects (blood samples in this work) by projecting the original data set from the high dimensional space of variables (investigated laboratory tests) onto the lower dimension space. Often only two or three most important principal components, calculated by the linear combination of original variables, sufficiently represent the total variability of the original data [12].

Cluster analysis, CA, is the term applied to a group of techniques that seek how to divide a set of objects into a number of homogeneous groups or clusters when there no *a priori* information about the group structure of the data [13]. In CA, also variables may be grouped instead of objects and in this way the similarities among the variables can be demonstrated.

The goal of multivariate classification is to classify the investigated objects characterized by the selected attributes or variables; that is, to determine which class every object belongs to. Based on the set of data whose class is *a priori* known a set of rules are designed and generalized in order to classify the objects with the greatest precision possible [14].

Most known classification techniques are linear discriminant analysis [15–17] and quadratic discriminant analysis [17–19] as well as logistic regression [17,20], considered also as a discriminant analysis technique in a broader sense [17]. The *K*-th nearest neighbour discrimination technique [17], KNN, is also useful for classification method since it does not need any assumption on the error distribution, unless the previously mentioned ways of classification. The main variant of this technique is based on the majority vote rule, which means that *K* neighbour objects, nearest to the classified object, are searched and then the classification of the given object is made according to which class the neighbour objects are predominantly classified. The success of classification techniques is given by the ratio of the correctly categorized objects (patient samples) over the number of all objects. This is calculated for the training set as well as the validation set of objects; the validation results are much more important since they estimate the prediction power of the given classification.

The data preprocessing in PCA, CA and KNN was made by standardization (autoscaling).

Analysis of variance, ANOVA, although being not a multidimensional technique, is a multiple comparison procedure, which reveals whether several sample means can be considered to be equal [16].

3. Results and discussion

3.1. Principal component analysis

The obtained PCA results are visualized in the biplot form in Fig. 1, where 12 biochemical tests applied to 84 men samples are linearly combined in the form of principal components. This biplot simultaneously represents the samples as markers together with twelve selected variables, depicted by the rays stretched from the origin to the point determining the variable position in the plane of the principal components. A close position of *tCHOL*, *AI*, *LDLc* and *TG* at the shown PCA biplot confirms their strong mutual dependence, independently confirmed also by performed correlation analysis. The mentioned variable position helps to understand the PC1 axis as expressing the cardiovascular risk. This statement is confirmed by the almost opposite position of *HDLc* at the low PC1 value. The position of *CREA* (and also *CK* in some extent) indicates its partial relation to the cardiovascular risk, represented by the PC1 coordinate. Further variables, mainly *AST*, *ALT* and *GMT* and their opponents *ALP* and *Age*, located along the PC2 axis, are perpendicular to the cholesterol variables, which indicate that they are independent on them and on the cardiovascular risk. The same output provided the results of correlation analysis.

The biplot obtained for women exhibited *tCHOL* and *LDLc* as the strongest agents indicating the cardiovascular risk. *CREA*, *ALP* and

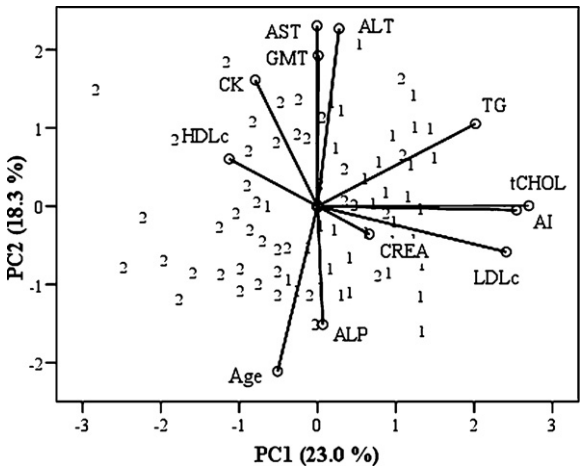

Fig. 1. PCA biplot in the PC2–PC1 plane for 84 men samples and 12 variables (biochemical tests). Software SPSS 15.0. (1) Values before the statin treatment, (2) values after 1-year treatment by statins.

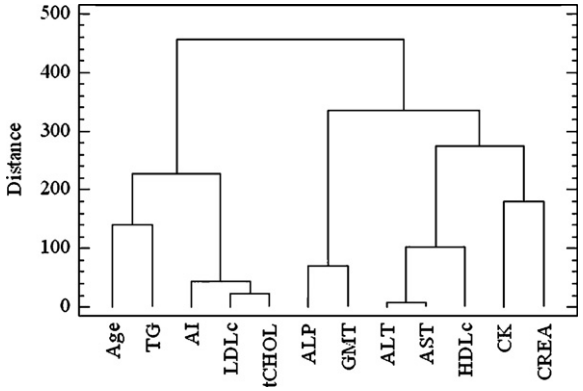

Fig. 2. Dendrogram of cluster analysis of RCBA values for 88 women samples and 12 variables (11 biochemical tests plus *Age*). Ward's clustering method, squared Euclidean distance. Software Statgraphics Plus 5.1.

Age (different to men!) were dependent on this risk partially. In general, the PCA biplot for women did not exhibit the effects of the statin treatment in the same way as in the case of men; the position of *AI* and *TG* with respect to *tCHOL* and *LDLc* was not close but perpendicular. However, it should be noted that in both cases (men as well as women) first two principal components express only about 41–42% of the data variability, which is not sufficient. Therefore information taken from PCA is not complete and has to be supplemented by the results of further chemometrical techniques.

3.2. Cluster analysis

In the studied case, cluster analysis was performed distinctively for the men and the women samples. Ward hierarchical cluster analysis was applied using squared Euclidean distance between the variables. For the men as well as the women data three main clusters appeared at the dendrogram, which is a common output of these techniques [21]. In both cases one cluster (showing the maximum similarity) was formed by the "bad" cholesterol variables *tCHOL*, *LDLc* and *AI*, which all represent high cardiovascular risk. As the closest to it a small cluster of *TG* and *Age* appeared but only for the women samples (Fig. 2), which is different to the situation observed for men. The definition of the relative change of the individual biochemical test (RCBA) is given in part 3.6. With regard to the most important variables indicating lipidaemia, the situation for men and women is similar so that, in general, the results of cluster analysis are supporting the PCA outputs.

3.3. Classification by discriminant analysis and KNN

Table 1 shows a summary of the results of linear discriminant analysis, LDA, quadratic discriminant analysis, QDA, and logistic regression, LR, for two data sets belonging to *men* and *women*. All these methods distribute the patient samples into two classes – before treatment by statins (Class 1) and after it (Class 2). The results for the training data set (calculating the classification model) and the leave-one-out validation, LOO, are represented in per cents. In the LOO technique, one object is left out from the training set and used as the only object for the purpose of validation until all objects are successively changed off. The result in the validation step, related to the samples independent of those used in the training process, is more important for overall valorization of the classification method.

The KNN results were also evaluated on the basis of successful classification (in %) for (a) the training data set, and for (b) the samples excluded from the training set by the leave-one-out method, used for the cross-validation purposes. The best classification performance in Table 1 was achieved when using seven nearest neighbours ($K = 7$) for the men samples and eleven ($K = 11$) for the

Table 1
Results of classification of men and women samples into two categories – before and 1 year after statin treatment – by discriminant analysis (LDA, QDA), logistic regression (LR) and KNN method calculated by two software packages SPSS and SAS.

Test	Results	Training set		Leave-one-out	
		Men	Women	Men	Women
LDA	True/all	73/84	81/88	67/84	74/88
	% True	86.9	92.0	79.8	84.1
QDA	True/all	70/84	76/88	55/84	61/88
	% True	83.3	86.4	65.5	69.3
LR	True/all	75/84	79/88	72/84	77/88
	% True	89.3	89.8	85.7	87.5
KNN	True/all	69/84	76/88	64/84	70/88
$K = 5$	% True	82.1	86.4	76.2	79.5
KNN	True/all	69/84	75/88	66/84	73/88
$K = 7$	% True	82.1	85.2	78.6	82.9
KNN	True/all	70/84	76/88	64/84	73/88
$K = 9$	% True	83.3	86.4	76.2	82.9
KNN	True/all	69/84	76/88	63/84	75/88
$K = 11$	% True	82.1	86.4	75.0	85.2

Table 2
Effect of statin administration on the level of all investigated variables (laboratory tests as well as multicomponent variables) expressed by the ROC curve area and Gini coefficients.

Men			Women		
Variable	*A*	*G*	Variable	*A*	*G*
DF1	0.949	0.898	Logit	0.973	0.946
Logit	0.935	0.870	DF1	0.968	0.936
tCHOL	0.931	0.862	tCHOL	0.940	0.880
LDLc	0.921	0.842	LDLc	0.880	0.760
PC1	0.892	0.784	AI	0.748	0.496
AI	0.798	0.596	PC1	0.710	0.420
TG	0.713	0.426	TG	0.676	0.352
HDLc	0.629	0.258	ALT	0.598	0.196
CK	0.554	0.108	HDLc	0.562	0.124
CREA	0.529	0.058	CK	0.557	0.114
ALT	0.512	0.024	ALP	0.533	0.066
AST	0.509	0.018	GMT	0.517	0.034
ALP	0.501	0.002	AST	0.512	0.024
GMT	0.490	−0.020	CREA	0.509	0.018

Note: *A* – area under the corresponding ROC curve; *G* – Gini coefficient.

women samples. Nevertheless, these results are not better compared to logistic regression, which exhibits the best results for the leave-one-out validation amongst all.

The importance of classification methods in the investigated statin problem is diminished by the fact that in this case they are not expected to provide prediction of the patient category, e.g. in the form of positive or negative diagnosis. Instead, their more modest aim is to demonstrate that the difference of the patients' status before and after statin treatment is highly significant. This quantitative output complements the qualitative outputs of principal component analysis and cluster analysis.

3.4. ROC analysis

The predictive value of any test can be displayed by constructing the plot of sensitivity against (1 – specificity). Sensitivity and specificity are here defined in the way common in clinical chemistry; otherwise they may be called sensitivity measure and selectivity measure, respectively [22]. The area under the corresponding ROC curve, *A*, is used as a summary measure of the test effectivity [23]. The ideal ROC curve has an *A* of 1, while a totally ineffective test exhibits a ROC curve along the diagonal line and has an *A* of 0.5.

In the performed ROC analysis, all original laboratory variables plus three linearly composed multicomponent variables PC1, DF1 and logit were used. The values of the first principal component, PC1, are the same as computed by principal component analysis and used in Section 3.1. Applied software allows saving these values in a table form into the PC memory and used them for another purpose. The values of the first discriminant function, DF1, were similarly obtained in a way described under linear discriminant analysis in Section 3.3; logit is the calculated dependent variable in logistic regression (part Section 3.3). The patient values of DF1 and logit were saved into the PC memory and subsequently used together with the saved PC1 values in the data file prepared in MS Excel for the ROC analysis.

The best variables with the largest ROC curve area are shown in Fig. 3, which demonstrates that the *A* value belonging to DF1 is clearly larger compared to the best original variables; the logit A value is also insignificantly larger than the ROC area for tCHOL – the best individual marker. Among them the most significant statin effect (with $A > 0.6$) exhibit *tCHOL*, *LDLc*, *AI*, *TG* and *HDLc* for the men samples and *tCHOL*, *LDLc*, *AI* and *TG* for the women samples. The observed independence of six standard biochemical parameters (characterizing the liver and/or renal human body functions) upon the statin treatment can be expected. The Gini coefficients, *G*, surveyed in Table 2, represent an alternative comparison of the statin effect on 11 investigated variables in the more convenient interval (0, 1):

$$G = 2A - 1 \quad (1)$$

It is worth noting that the observed negative values of Gini coefficients are caused by random errors affecting the part of the ROC curve below the diagonal line, which represents the ROC area value of $A = 0.5$.

Table 3
One-way analysis of variance showing effect of statin administration upon biochemical tests using categorical variable *Class 2*.

Test		*F*	*p*	Test		*F*	*p*
tCHOL	Men	90.92	**6.2**E−**15**	*AST*	Men	1.214	0.274
	Women	91.68	**3.3**E−**15**		Women	0.093	0.762
LDLc	Men	73.29	**5.4**E−**13**	*CK*	Men	0.885	0.350
	Women	60.18	**1.6**E−**11**		Women	0.503	0.480
AI	Men	24.78	**3.5**E−**06**	*CREA*	Men	0.811	0.371
	Women	18.72	**4.1**E−**05**		Women	0.118	0.732
TG	Men	9.147	**0.0033**	*ALT*	Men	0.616	0.435
	Women	7.217	**0.0087**		Women	3.503	0.0647
HDLc	Men	2.176	0.144	*GMT*	Men	0.268	0.606
	Women	1.347	0.249		Women	0.339	0.562
				ALP	Men	0.014	0.906
					Women	0.471	0.495

Categories: (1) Before the drug administration, (2) after 1-year drug administration (separately for men and women). Critical *F*-values are: $F(0.05, 1, 82) = 3.958$ for men samples, $F(0.05, 1, 86) = 3.952$ for women samples. Significant results obtained by the given laboratory test are indicated in bold (separately for men and women samples).

Table 4
Output of the ANOVA least significance difference post hoc test indicating all significant differences between the pairs (*I* vs. *J*) of four investigated categories of *Class 4*.

Multiple comparison							
Dependent variables	(*I*)	(*J*)	Mean difference (*I*–*J*)	Standard error	*p*[a]	95% Confidence interval	
						Lower bound	Upper bound
tCHOL	1	2	1.567	0.172	2.1E−16	1.228	1.905
	1	4	1.746	0.170	1.5E−19	1.411	2.081
	3	2	1.489	0.170	1.8E−15	1.154	1.824
	3	4	1.668	0.168	1.3E−18	1.337	1.999
LDLc	1	2	1.243	0.156	2.1E−13	0.936	1.550
	1	4	1.431	0.154	7.1E−17	1.128	1.735
	3	2	1.061	0.154	1.0E−10	0.758	1.365
	3	4	1.250	0.152	5.2E−14	0.950	1.550
TG	1	2	0.555	0.169	0.0012	0.226	0.888
	1	3	0.478	0.167	0.0046	0.150	0.807
	1	4	0.879	0.167	4.0E−07	0.551	1.208
	3	1	−0.478	0.167	0.0046	−0.817	−0.150
	3	4	0.401	0.165	0.0159	0.076	0.726
AI							
	1	2	0.908	0.179	1.0E−06	0.555	1.260
	1	3	0.916	0.177	6.2E−07	0.567	1.265
	1	4	1.657	0.177	4.5E−17	1.308	2.006
	3	1	−0.916	0.177	6.2E−07	−1.265	−0.567
	3	4	0.741	0.175	3.6E−05	0.396	1.086

Categories: (1) Men before the statin administration, (2) men after 1-year statin administration, (3) women before the statin administration, (4) women after 1-year statin administration.

[a] Significance level is expressed by *p*-values rounded to 2 or 3 valid figures; the mean difference is considered significant when $p < 0.05$. Other explored laboratory tests did not provide significant differences.

3.5. Analysis of variance (ANOVA) and descriptive statistics

In this work, ANOVA was performed for each of 11 quantitative variables dependent on the selected single factor. This factor may be represented by two categorical variable, named here as *Class 2* and *Class 4*, which indicates whether (a) the treatment by statins was carried out (after drug administration) or not (before it) and (b) the patient is a man or a woman. Two categories of *Class 2* are (1) the patient sample before the drug administration, (2) the patient sample after the drug treatment; ANOVA was evaluated separately

Table 5
Descriptive statistics for biochemical tests before statin administration (denoted *b*) and after one year statin administration (denoted *a*) using common and robust estimates and relative change of biochemical test defined as RCBA = $(a - b)/b$ for all tests.

Test	Mean (x)			Standard deviation (s)			Median ($\tilde{x}$)			*IQR* (adjusted)			$t = xn^{1/2}/s$
	b	*a*	RCBA	*b*	*a*	RCBA	*b*	*a*	RCBA	*b*	*a*	RCBA	RCBA
Men													
tCHOL	6.60	5.04	−0.235	0.722	0.783	0.107	6.40	4.90	−0.230	0.815	0.723	0.112	**14.149**
LDLc	4.19	2.94	−0.288	0.690	0.640	0.161	4.08	3.03	−0.297	0.625	0.576	0.149	**11.553**
AI	3.95	3.05	−0.218	0.828	0.843	0.193	3.85	3.00	−0.234	0.908	0.649	0.168	**7.316**
TG	2.34	1.78	−0.200	0.839	0.844	0.382	2.34	1.68	−0.291	0.788	0.523	0.143	**3.393**
HDLc	1.36	1.28	−0.056	0.216	0.265	0.127	1.31	1.27	−0.065	0.161	0.193	0.104	**2.878**
CK	1.87	2.03	0.131	0.749	0.824	0.327	1.78	1.86	0.107	0.804	0.812	0.306	**2.594**
CREA	82.7	80.2	−0.024	13.9	12.0	0.098	79.4	79.7	−0.018	14.233	13.028	0.087	1.586
GMT	0.705	0.778	0.120	0.548	0.734	0.549	0.485	0.515	−0.012	0.456	0.415	0.293	1.416
ALT	0.541	0.609	0.250	0.230	0.509	1.330	0.475	0.505	−0.071	0.250	0.235	0.366	1.216
AST	0.434	0.443	0.045	0.104	0.146	0.297	0.415	0.395	−0.028	0.089	0.111	0.204	0.973
ALP	1.26	1.25	0.009	0.338	0.340	0.220	1.21	1.22	−0.035	0.302	0.263	0.101	0.267
Women													
tCHOL	6.53	4.86	−0.253	0.802	0.832	0.112	6.40	4.80	−0.271	0.723	0.834	0.092	**14.973**
LDLc	4.00	2.75	−0.302	0.709	0.799	0.199	3.94	2.61	−0.324	0.643	0.658	0.196	**10.064**
AI	3.04	2.30	−0.229	0.836	0.769	0.202	3.00	2.25	−0.247	0.778	0.815	0.213	**7.500**
TG	1.86	1.46	−0.197	0.734	0.664	0.209	1.86	1.34	−0.199	0.752	0.517	0.194	**6.251**
HDLc	1.70	1.59	−0.054	0.445	0.401	0.119	1.70	1.58	−0.058	0.324	0.369	0.134	**3.031**
CK	1.65	1.79	0.163	0.896	0.992	0.469	1.45	1.55	0.055	0.573	0.686	0.339	**2.312**
ALT	0.405	0.353	−0.079	0.146	0.111	0.266	0.37	0.34	−0.120	0.154	0.120	0.289	1.961
ALP	1.29	1.24	−0.022	0.359	0.321	0.166	1.24	1.18	−0.045	0.413	0.406	0.140	0.898
GMT	0.435	0.375	0.017	0.620	0.306	0.453	0.31	0.30	−0.058	0.143	0.148	0.249	0.248
CREA	70.2	69.3	−0.004	12.2	11.5	0.125	68.75	67.65	0.014	11.490	10.823	0.131	0.212
AST	0.399	0.393	0.003	0.098	0.091	0.196	0.40	0.38	−0.024	0.069	0.095	0.156	0.113

Note: For men $t\_crit = 2.020$ ($\alpha = 0.05$, $\nu = 41$); for women $t\_crit = 2.017$ ($\alpha = 0.05$, $\nu = 43$). Significant *t*-test values for RCBA are denoted by bold typefaces and indicate the significant difference of the test results after 1-year treatment and before statin administration. *IQR* – interquartile range (the difference between the upper and the lower quartile adjusted by the factor of 0.7413 to fit better to the standard deviation).

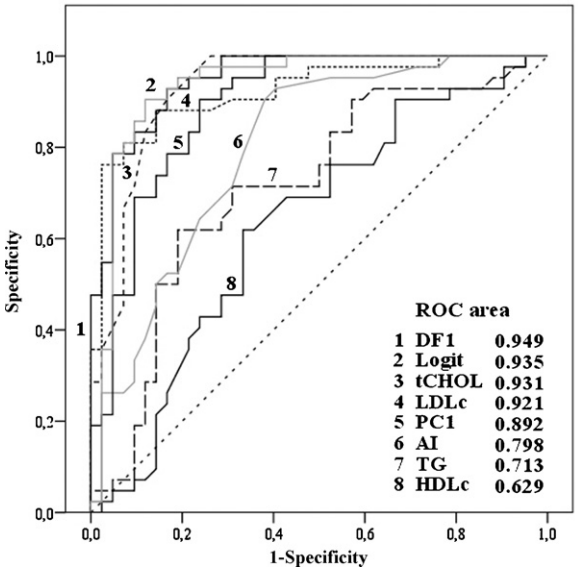

Fig. 3. ROC curves indicating the effectivity of the statin administration for the best biochemical tests *tCHOL*, *LDLc*, *AI* and *TG* (selected by the largest area under the ROC curve) and the calculated multicomponent variables DF1, logit and PC1 for 84 men samples. Software SPSS 15.0.

for men and women. The results collected in Table 3 show that the statin treatment affected significantly the level of four following tests: *tCHOL*, *LDLc*, *AI* and *TG* with the *p*-values less than 0.05 (95% probability).

Four categories of *Class 4* are (1) men before the drug administration, (2) men after the drug treatment, (3) women before the treatment, (4) women after the treatment. In this case two post hoc ANOVA tests were applied: (1) the least significant difference test [16,24], (2) Bonferroni test [16]. Both tests provided the same results: the variables *tCHOL*, *LDLc*, *TG* and *AI* are capable to separate the category 1 from category 2 and the category 3 from category 4 (which is the main goal) and also 1 from 4 and 3 from 2. The selection of the most important ANOVA outputs is summarized in Table 4 where only the significant combinations of the categories are included.

Even though descriptive statistics is often considered as something deficient compared to more sophisticated multivariate techniques, it can bring valuable information adding a new insight to the studied problem. A simple tool, well demonstrating the effects of the statin treatment, is based on the relative change of biochemical test, RCBA, which we have defined as

$$\mathrm{RCBA} = \frac{a-b}{b} \qquad (2)$$

where *b* denotes the test result before the statin treatment and *a* denotes it after 1-year treatment. This approach allows for hypotheses testing effective using not only the mean and the standard deviation but also their robust counterparts – the median and the interquartile range (*IQR*), and performs finally the tests of significance for every laboratory test in a very simple way. The relevant data including also the final *t*-test results are summarized in Table 5. This table demonstrates six laboratory tests with the *t* value larger than the critical one so that *tCHOL*, *LDLc*, *AI*, *TG*, *HDLc*, and also *CK* (which is close to the critical test value) may be considered significantly affected by the statin treatment for the samples regarding both genders of the patients. The order of the tests in this table shows how much they are affected. At the same time the close values of the pairs x vs. $\tilde{x}$ and s vs. *IQR* signify no outliers existing in the investigated data sets.

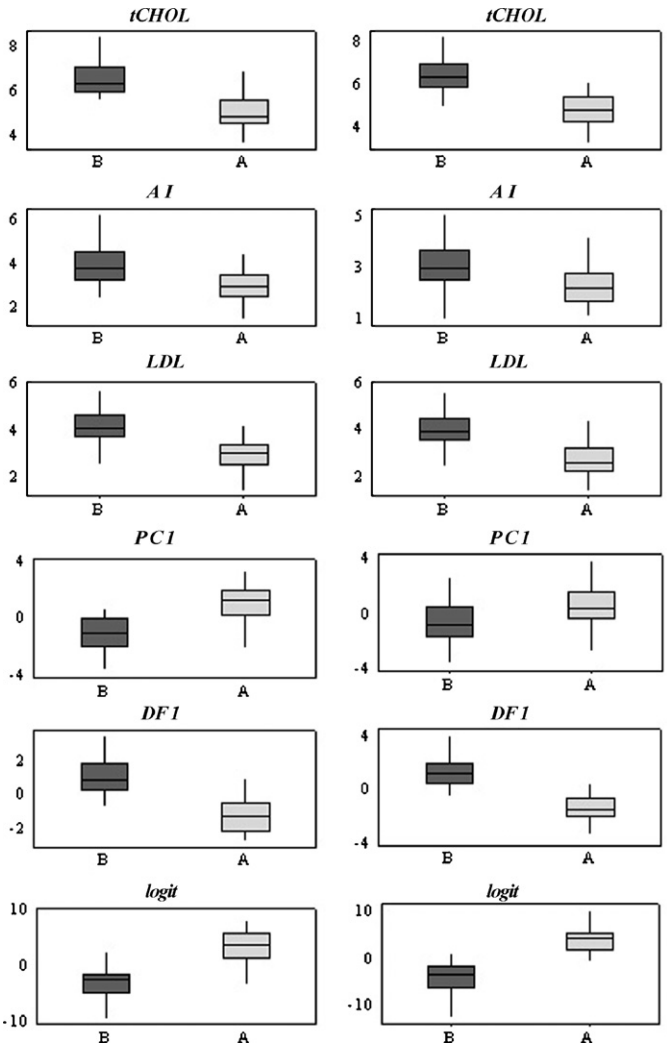

Fig. 4. Comparison of the box- and whisker plots for the best laboratory tests and the multicomponent variables logit, DF1 and PC1. Box- and whisker plots are constructed using 10%, 25%, 50%, 75% and 90% percentiles of the ranked variable or the investigated test. (B) blood serum samples before taking statins, (A) after the statin treatment. Men samples are located in the left column, women samples are in the right column. Most successful results are in the bottom.

Box- and whisker plots are useful means of descriptive statistics for comparing different sets of one-dimensional data and their construction allows a visual representation of the data [16]. The box itself covers inner 50% of all data values, starting from the lower quartile (25% of the ordered data) and ending by the upper quartile (75% of the ordered data). The whiskers, represented by the abscisses, cover the lowest and highest part of the variable data; the line across the box represents the median.

Based on the known categorization of the proband samples into two categories – before and after administration of statins, a comparison of the statin effect on the blood serum levels of all studied tests is visualized using box- and whisker plots. In addition, three multicomponent variables, namely PC1 (the first principal component), DF1 (the first discriminant function) and logit (the dependent variable in logistic regression) are also shown and compared to individual biochemical tests. The multicomponent variables were calculated by linear combinations of all original variables (tests) by principal component analysis, discriminant analysis and logistic regression. Box-plots in Fig. 4 demonstrate that the selected two categories of probands are relatively well separated using the variables *tCHOL*, *LDLc*, and *AI* for the men as well as the women sam-

ples. Nevertheless, it is evident that even better separation of two categories is achieved when using *PC1*, *logit* and *DF1* for the men samples and *DF1* and *logit* for the women samples. The effective utilization of the multicomponent variables in prediction and confirmation of clinical diagnosis was discovered in our previous works [25,26] and is supported by the present results. A relatively small difference of the *HDLc* results is in accordance with several abovementioned results; the same is valid for practically insignificant changes of further six biochemical parameters.

4. Conclusions

Positive changes in lipid metabolism after statin treatment of the patients with cardiovascular risk can be unambiguously determined and monitored by means of statistical and chemometrical techniques, which provide qualitative as well as quantitative judgment related to the laboratory tests, which are mostly affected by the administration of statin drugs.

In this work, biplots of principal component analysis, dendrograms of cluster analysis, ROC curves and box- and whisker plots provide visualization of the statin effects. Discriminant analyses, logistic regression and KNN classification methods allow a clear discrimination of the patients' samples into two categories – before and after statin treatment. Analysis of variance revealed that four variables are capable to differentiate the statin treatment with regard to the patient gender: total cholesterol, low-density lipoprotein cholesterol, triacylglycerols and aterogenity index. High-density lipoprotein cholesterol as well as all further investigated biochemical parameters were not efficient in differentiating neither *men* nor *women* groups before and after statin treatment.

A very high diagnostic effectiveness of three calculated multicomponent variables, composed by linear combination of individual laboratory tests, represents a special feature of the achieved results. It predestinates their further utilization in cardiovascular risk confirmation and prediction.

There may be considered two potentially serious side effects of statins, of which patients need to be aware. Occasionally, statin use cause an increase in liver function tests (*ALT*, *AST*, *ALP*, *GMT*). If the increase is severe, the patient may need to stop taking the drug, which usually reverses the problem. If there is no increase or it is only mild, one can continue to take the drug. In general, our study *has not proved* a significant change in the level of liver function tests with the statin uptake. In present work, among the non-lipid tests only creatine kinase, *CK*, (indicating a possible myopathia) were found possibly affected by the statin treatment on the basis of the *t*-tests, both for men and women. However, neither ROC curves nor ANOVA results were not decisive to confirm the previous suspicion. It is important to underline that in such a case (often occurring in real life) only the use of several statistical tools can provide an objective general statement. Such an effect should be assessed individually for the corresponding patients and their further monitoring should be made aimed to its evaluation.

Acknowledgements

The support of this work by the grants VEGA 1/1005/09 and 1/0066/09 as well as VVCE-0004-07 is gratefully acknowledged.

References

[1] A. Endo, The discovery and development of HMG-CoA reductase inhibitors, J. Lipid Res. 33 (1992) 1569–1582.
[2] E. Istvan, J. Deisenhofer, Statin inhibition of HMG-CoA reductase: a 3-dimensional view, Atheroscler. Suppl. 4 (2003) 3–8.
[3] A.S. Wierbzbicki, R. Poston, A. Ferro, The lipid and non-lipid effects of statins, Pharmacol. Ther. 99 (2003) 95–112.
[4] R.S. Rosenson, Statins in atherosclerosis: lipid-lowering agents with antioxidant capabilities, Atherosclerosis 173 (2004) 1–12.
[5] A. Blum, R. Shamburek, The pleiotropic effects of statins on endothelial function, vascular inflammation, immunomodulation and thrombogenesis, Atherosclerosis 203 (2009) 325–330.
[6] C. Cortese, L. Liberatoscioli, Effects of statins on lipoprotein fractions, Int. Congr. Ser. 1253 (2003) 247–252.
[7] A. Branchi, A.M. Fiorenza, A. Torri, F. Muzio, A. Rovellini, C. Berra, D. Sommariva, Effects of atorvastatin 10 mg and simvastatin 20 mg on serum triglyceride levels in patients with hypercholesterolemia, Curr. Ther. Res. 62 (2001) 408–415.
[8] M.J. Chapman, F. McTaggart, Optimizing the pharmacology of statins: characteristics of rosuvastatin, Atheroscler. Suppl. 2 (2002) 33–37.
[9] M. Schachter, Chemical, pharmacokinetic and pharmacodynamic properties of statins: an update, Fundam. Clin. Pharmacol. 19 (2004) 117–125.
[10] G. Anfossi, P. Massucco, K. Bonomo, M. Trovati, Prescription of statins to dyslipidemic patients affected by liver diseases: a subtle balance between risks and benefits, Nutr. Metab. Cardiovasc. Dis. 14 (2004) 215–224.
[11] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 2002.
[12] M. Otto, Chemometrics: Statistics and Computer Application in Analytical Chemistry, Wiley, Weinheim, 1999.
[13] N. Bratchell, Cluster analysis, Chemom. Intell. Lab. Syst. 6 (1987) 105–125.
[14] J. Pacheco, S. Casado, L. Núnez, O. Gómez, Analysis of new variable selection methods for discriminant analysis, Comput. Stat. Data Anal. 51 (2006) 1463–1478.
[15] S. Sharma, Applied Multivariate Techniques, Wiley, New York, 1996.
[16] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997.
[17] R. Khattree, D.N. Naik, Multivariate Data Reduction and Discrimination, SAS Institute, Cary, NC, 2000.
[18] M.P. Derde, D.L. Massart, Comparison of the performance of the class modelling techniques UNEQ, SIMCA, and PRIMA, Chemom. Intell. Lab. Syst. 4 (1988) 65–93.
[19] B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part B., Elsevier, Amsterdam, 1998.
[20] D.G. Kleinbaum, M. Klein, E.R. Pryor, Logistic Regression, Springer, Heidelberg, 2005.
[21] M. Daszykowski, B. Walczak, D.L. Massart, Density-based clustering for exploration of analytical data, Anal. Bioanal. Chem. 380 (2004) 370–372.
[22] J. Mocak, B. Balla, A. Bobrowski, P. Blazicek, Proper ways of comparison of two laboratory methods, Chem. Pap. 57 (2003) 143–146.
[23] C.D. Brown, H.T. Davis, Receiver operating characteristics curves and related decision measures: a tutorial, Chem. Intell. Lab. Syst. 80 (2006) 24–38.
[24] A.M. Brown, A new software for carrying out one-way ANOVA post hoc tests, Comput. Methods Program. Biomed. 79 (2005) 89–95.
[25] B. Balla, J. Mocak, H. Pivovarnikova, J. Balla, Comparative study of cardiovascular markers data by various techniques of multivariate analysis, Chemom. Intell. Lab. Syst. 72 (2004) 259–267.
[26] V. Mrazova, J. Mocak, E. Varmusova, D. Kavkova, A. Bednarova, Use of multidimensional data analysis for prediction of lung malignity, J. Pharm. Biomed. Anal. 50 (2009) 210–215.