CMST 10(1) 73-82 (2004) DOI:10.12921/cmst.2004.10.01.73-82

COMPUTATIONAL METHODS IN SCIENCE AND TECHNOLOGY **10**,73-82 (2004)

ADVANCED METHODS OF HEART RATE SIGNALS PROCESSING AND THEIR USEFULNESS IN DIAGNOSIS SUPPORT II. UNIVARIATE STATISTICAL TECHNIQUES

JERZY A. MOCZKO

Chair and Department of Computer Science and Statistics
University of Medical Sciences in Poznań
Dąbrowskiego 79, 60-529 Poznań, Poland

e-mail address: jmoczko@am.poznan.pl

(Rec. 29 February 2004)

Abstract: A tentative cardiological database was established using virtual instrumentation described in the first part of presented paper. Some additional not heart rate variability parameters were added. Three selected univariate statistical techniques were used for illustration diagnosis support techniques in discrimination between healthy and coronary heart disease people. Comparison of nonparametric Mann-Whitney test, receiver operating characteristic ROC analysis and univariate logistic regression results was performed. In all used methods long term heart rate variability indices were most useful in prediction of patient's status. The correctness of classification was between 55 to 79 percent with ROC technique and 68 to 78 percent with logistic regression. However high number of false negative FN cases excludes univariate techniques as reliable screening test.

1. INTRODUCTION

In first part of the paper a wide set of heart rate variability (HRV) indices in time, frequency and time-frequency domains was described [1]. The specially designed virtual instrumentation allows for interactive choice of ECG segment of interest (SOI) in anyone of 12-leads, detects *R*-waves, estimates $T_{R\text{-}R}$ intervals and heart rate (HR), calculates HRV indices and finally saves them in ASCII format database. This database was supplemented with additional information about patients. Whereas all data obtained from signal analysis are measured in statistical interval scale, these last are expressed in interval, ordinal or nominal scale. The first scale is represented by age, body mass index BMI, systolic RR1 and diastolic RR2 blood pressure, duration of the heart disease, echocardiographic measurements (left ventricle LK, left atrium LP, right ventricle PK, ventricular septum PMK, posterior wall left ventricle TSLK), lipids levels (total cholesterol, high- and low-density cholesterol, triglycerides levels), electrolytes levels (sodium and potassium), glucose and glycolysed hemoglobin HBA1C. In ordinal scale two parameters are measured contractility and coronary angiography results. Gender, pharmacological treatment, family history, other diseases and tobacco smoking are the measurements expressed in nominal scale. In presented preliminary methodological studies 61 data records were analyzed: 19 patients with diabetes, 27 with

coronary heart disease (CHD) and 15 healthy patients. The last group consisted of volunteers both age- and sex-matched. Clinical symptoms of CHD were confirmed with coronary angiography. ECG was recorded in five following situations: in supine position, in standing position, in supine position after administration of single-dose one milligram of atropine subcutaneously, in supine position after administration of single oral dose 40 milligrams of propranolol and in supine position after administration of single oral dose 25 milligrams of captopril. The main goal of presented part of study is to compare properties of measured parameters in discrimination between patients with stable coronary heart disease and control group in univariate approach.

2. DESCRIPTIVE AND CLASSICAL HYPOTHESIS TESTING STATISTICAL RESULTS

In the first step of statistical analysis normal distribution of interval scale parameters in both analyzed data groups was checked. Lillierforce's and Shapiro-Wilk's tests revealed not Gaussian distribution of data in prevalent number of measured parameters. On the other hand Levene test demonstrated lack of homogeneity of variances. These results and relatively small sample sizes forced us to use nonparametric Mann-Whitney test as a nonparametric alternative to the t-test for independent samples [2]. High power of Mann-Whitney test (95.5% in comparison to t-Student test) does not change substantially the ability of distinction between central tendency measures in analyzed groups. All mentioned calculations were done with StatSoft STATISTICA Data Miner version 6.1 (2004). The example of obtained results for standing position is presented in Fig. 1.

Mann-Whitney U Test (cardiobase1-interval)
By variable grupa_num
Marked tests are significant at p < ,05000

variable	Rank Sum Group 1	Rank Sum Group 2	U	Z	p-level	Z adjusted	p-level	Valid N Group 1	Valid N Group 2	2*1sided exact p
PI_Yeh II	397.0000	464.0000	113.0000	2.21946	0.026456	2.21956	0.026450	15	26	0.026142
PI_Organ BAND	413.0000	448.0000	97.0000	2.65253	0.007990	2.65253	0.007990	15	26	0.007211
PI_van Geijn ID	381.0000	480.0000	129.0000	1.78640	0.074036	1.78640	0.074036	15	26	0.076071
PI_Huey STV	393.0000	468.0000	117.0000	2.11119	0.034756	2.11119	0.034756	15	26	0.034813
PI_Huey LTV	429.0000	432.0000	81.0000	3.08659	0.002032	3.08659	0.002032	15	26	0.001540
PI_Dalton MABD	339.0000	522.0000	171.0000	0.64980	0.515952	0.64980	0.515952	15	26	0.529161
PI_Dalton SD	372.0000	489.0000	138.0000	1.54280	0.122881	1.54280	0.122881	15	26	0.127262
PI_Zugaib STV	368.0000	493.0000	142.0000	1.43453	0.151422	1.43464	0.151333	15	26	0.157029

Fig. 1. Fragment of Mann-Whitney test results for standing position - comparison between coronary heart disease and control groups

The prominent distinction between CHD and healthy people was observed in all five clinical situations in Huey LTV index and in w(2) wavelet-transform standard deviation. Four indices (Organ BAND, oscillation index OSC and w(1) and w(3) wavelet-transform standard deviations detected the difference after captopril and propranolol administration and in

Table 1. Usefulness of analyzed indices in discrimination between CHD and healthy people using nonparametric Mann-Whitney test in five different clinical situations. Plus sign denotes statistical significance at least at 0.05

Index	Supine Position	Standing Position	Captopril	Atropine	Propranolol
De Haan STI					
De Haan LTI					+
Yeh DI					
Yeh II		+			+
Organ BAND		+	+		+
van Geijn ID				+	+
Huey STV		+			+
Huey LTV	+	+	+	+	+
Dalton MABB					
Dalton DSD					+
Zugaib STV					
Zugaib LTV		+			+
OSC		+	+		+
SDNN					+
RMSSD	+		+		
Allan		+			
FFT total power		+			+
FFT VLF					
FFT LF					
FFT HF					
FFT LF/HF					
w(1)		+		+	+
w(2)	+	+	+	+	+
w(3)	+		+		+
w(4)					
w(5)		+			
AR(1)					+
AR(2)		+			
AR(3)					
AR(4)					
AR(5)					
AR(6)					
AR(7)					
AR(8)					
AR(9)					
AR(10)					
AR(11)					
AR(12)					
AR exc.noise					
AR total power					+
AR ULF					
AR VLF					
AR LF			+		+
AR HF					+
AR LF/HF			+		+

standing position. We did not found statistically significant difference in all short term variability indices (de Haan STI, Yeh DI, Dalton MABB, Zugaib STV and in majority of nonparametric frequency domain indices. In Table 1 the details of univariate approach are presented. On similar principle different clinical experiments may be performed. For example one may treat particular index value in supine position before drug administration as an initial status and compare the statistical significance of differences between this status and status after β-blockade in different heart diseases.

3. RECEIVER OPERATING CHARACTERISTIC CURVES

In the very majority of clinical situations we want to require a yes/no decision on the presence of a disease. This decision is done by comparing the output value of clinical test to some particular threshold. If obtained result of the test is above the threshold, the test is said to be positive (i.e. the disease is present, true positive TP). If the output is below the threshold, the test is said to be negative (i.e. the disease is absent, true negative TN). Of course, the opposite direction of decision making may be established, too. If the threshold unambiguously divides region of these two possible situations (healthy versus ill), we have no problem in diagnosis support. However, in real life high variability of variables describing actual status of living beings makes the discrimination problem more complex. The corresponding probability distribution functions (pdf) of measured parameters frequently overlap and unambiguous threshold does not exist. Figure 2 presents exemplary probability distribution functions for Zugaib's LTV index after propranolol administration for healthy and CHD people.

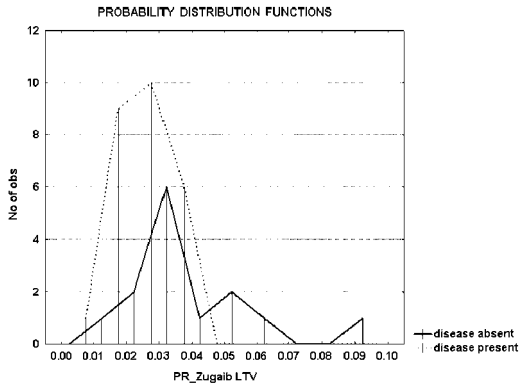

Fig. 2. Probability distribution functions for Zugaib's long term variability index after propranolol administration in control and coronary heart disease group

It is easy to find, that there does not exist any threshold level which uniquely discriminates mentioned groups of patients. We observe false positive results FP (when diagnostic test result is positive i.e. abnormal and true subject's condition is negative i.e. normal) or false negative results FN (in inverse situation). Statisticians introduce some indices to quantify quality of lab test: sensitivity (fraction of all cases with the disease who get a positive test result) and

specificity (fraction of all cases without the disease who get a negative test result). In other words sensitivity measures the ability of the test to detect the disease, while specificity describes the ability to exclude those objects who do not have the disease. Both indices belong to the interval <0,1>. Optimal situation is when sensitivity and specificity are equal to one. It corresponds to the situation when number of FP and FN cases is equal to zero (pdf curves do not overlap). The plot of sensitivity as a function of (1-specificity) is known as Receiver Operating Characteristic (ROC) curve [3, 4]. If the curve does not significantly differ from diagonal line (area under ROC curve does not significantly differ from 0.5) examined lab test does not effectively discriminate compared groups. Full discrimination is obtained when area under the curve is equal to one (number of FP and FN is equal to zero). Figure 3 presents results of ROC analysis obtained for Zugaib's long term variability index after propranolol administration in control and coronary heart disease group estimated with Analyse-It Software, ver. 1.62. *p*-value, which is less than assumed significance level $\alpha = 0.05$ confirms, that Zugaib's LTV index may be successfully used for CHD detection after propranolol administration. This result is consistent to Mann-Whitney test results in Table 1.

Curve	Area	SE	p	95% CI of Area	grupa_num = 2
PR_Zugaib LTV	0.821	0.0703	<0.0001	0.684 to 0.959	have lower values

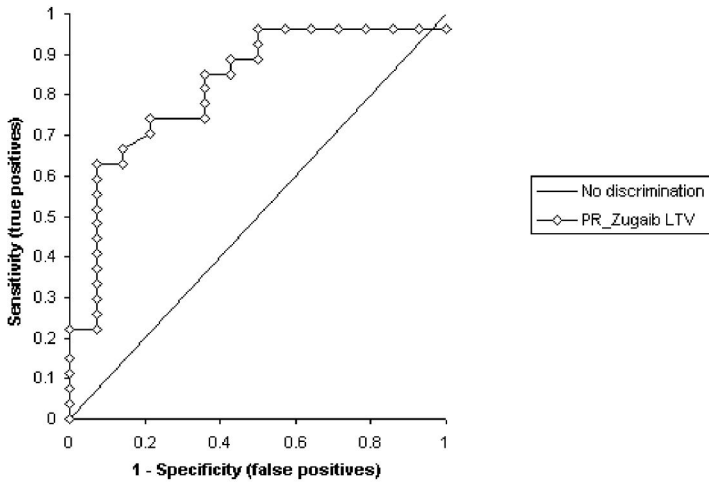

Fig. 3. ROC curve for Zugaib's long term variability index after propranolol administration in control and coronary heart disease group

In hypothesis testing theory the decrement of first type error increases the second type error (and vice versa) when sample size is constant. We have analogous situation in diagnosis

support. Shifting the threshold level in the range of examined parameter we change number of FP and FN cases. We may follow two different ways:

1. find a threshold level which simultaneously maximizes both sensitivity and specificity indices or
2. find a threshold level which corresponds to parameter value assuring assumed misclassification ratio.

PR_Zugaib LTV (abnormals below cut-off)	Sensitivity	Specificity	TP	TN	FP	FN
0.0083	0.0%	100.0%	0	14	0	27
0.0127	3.7%	100.0%	1	14	0	26
0.0130	7.4%	100.0%	2	14	0	25
0.0135	11.1%	100.0%	3	14	0	24
0.0153	14.8%	100.0%	4	14	0	23
0.0157	22.2%	100.0%	6	14	0	21
0.0163	22.2%	92.9%	6	13	1	21
0.0166	25.9%	92.9%	7	13	1	20
0.0170	29.6%	92.9%	8	13	1	19
0.0198	33.3%	92.9%	9	13	1	18
0.0209	37.0%	92.9%	10	13	1	17
0.0216	40.7%	92.9%	11	13	1	16
0.0221	44.4%	92.9%	12	13	1	15
0.0225	48.1%	92.9%	13	13	1	14
0.0227	51.9%	92.9%	14	13	1	13
0.0238	55.6%	92.9%	15	13	1	12
0.0262	59.3%	92.9%	16	13	1	11
0.0264	63.0%	92.9%	17	13	1	10
0.0265	63.0%	85.7%	17	12	2	10
0.0283	66.7%	85.7%	18	12	2	9
0.0298	70.4%	78.6%	19	11	3	8
0.0311	74.1%	78.6%	20	11	3	7
0.0314	74.1%	64.3%	20	9	5	7
0.0323	77.8%	64.3%	21	9	5	6
0.0339	81.5%	64.3%	22	9	5	5
0.0362	85.2%	64.3%	23	9	5	4
0.0366	85.2%	57.1%	23	8	6	4

Fig. 4. Sensitivity and specificity values as a function of threshold level for Zugaib's long term variability index after propranolol administration in control and coronary heart disease group (fragment of results)

The second situation is used when first and second type errors have not equivalent weight (i.e. it is better to detect the disease in healthy patient in a screening test than treat sick person as healthy one). In Fig. 4. we present sensitivity and specificity values obtained for different threshold level. Highlighted threshold 0.0311 maximizes simultaneously sensitivity and specificity and it may be used as "gold standard" if we assume that both types of mis-

classification are equivalently dangerous. We must remember that "gold standard" estimation should be confirmed by results obtained from sufficient sample size.

1	Parameter	Cutoff	Sensitivity	Specificity	TP	TN	FP	FN	STATUS
2	Organ BAND	6.0225	77.8%	60.0%	21	9	6	6	SUPINE
3	Huey LTV	84.7644	66.7%	80.0%	18	12	3	9	SUPINE
4	Zugaib LTV	0.0299	66.7%	60.0%	18	9	6	9	SUPINE
5	OSC	12.8852	70.4%	66.7%	19	10	5	8	SUPINE
6	RMSSD	47.7956	63.0%	100.0%	17	15	0	10	SUPINE
7	ALLAN	0.1213	66.7%	73.3%	18	11	4	9	SUPINE
8	FFT TOTAL POW	7.3692	70.4%	60.0%	19	9	6	8	SUPINE
9	W2	2.7260	74.1%	66.7%	20	10	5	7	SUPINE
10	W3	3.0031	74.1%	60.0%	20	9	6	7	SUPINE
11	YEH II	0.0459	84.6%	66.7%	22	10	5	4	STANDING
12	Organ BAND	7.1106	92.3%	60.0%	24	9	6	2	STANDING
13	van Geijn ID	4.3467	73.1%	60.0%	19	9	6	7	STANDING
14	Huey STV	104.5312	76.9%	66.7%	20	10	5	6	STANDING
15	Huey LTV	109.1154	76.9%	80.0%	20	12	3	6	STANDING
16	Zugaib LTV	0.0357	80.8%	66.7%	21	10	5	5	STANDING
17	OSC	15.5660	84.6%	66.7%	22	10	5	4	STANDING
18	SDNN	17.7806	73.1%	66.7%	19	10	5	7	STANDING
19	ALLAN	0.1094	57.7%	100.0%	15	15	0	11	STANDING
20	FFT TOTAL POW	6.3805	73.1%	66.7%	19	10	5	7	STANDING
21	W1	1.0437	69.2%	86.7%	18	13	2	8	STANDING
22	W2	3.1168	88.5%	66.7%	23	10	5	3	STANDING
23	W5	4.5053	80.8%	60.0%	21	9	6	5	STANDING
24	AR EXC. Noise	0.1327	73.1%	53.3%	19	8	7	7	STANDING
25	Organ BAND	3.5488	63.0%	73.3%	17	11	4	10	CAPTOPRIL
26	van Geijn ID	2.0780	51.9%	80.0%	14	9	6	13	CAPTOPRIL
27	Huey LTV	54.9069	66.7%	66.7%	18	10	5	9	CAPTOPRIL
28	OSC	8.4879	66.7%	80.0%	18	12	3	9	CAPTOPRIL
29	RMSSD	39.7928	81.5%	60.0%	22	9	6	5	CAPTOPRIL
30	FFT TOTAL POW	3.7616	74.1%	66.7%	20	10	5	7	CAPTOPRIL
31	LF	0.4232	63.0%	60.0%	17	9	6	10	CAPTOPRIL
32	W2	1.6347	77.8%	73.3%	21	11	4	6	CAPTOPRIL
33	W3	2.4349	65.4%	73.3%	17	11	4	9	CAPTOPRIL
34	W4	2.0850	57.7%	80.0%	15	12	3	11	CAPTOPRIL

Fig. 5. Threshold levels, corresponding sensitivity, specificity, number of true positives TP, true negatives TN, false positives FP and false negative FN cases in different clinical status-discrimination between CHD and control groups

In Fig. 5 and Fig. 6. we present the gathered results for ROC analysis of examined patients. One can find that the proper classification belongs to the interval <55, 79> percent, with mean value 76% after propranolol administration, 73% - in standing position, 70% - in supine position, 69% - after atropine administration and 63% - after captopril administration. In each clinical situation we observe prevalent number of FN cases in misclassification, what should

1	Parameter	Cutoff	Sensitivity	Specificity	TP	TN	FP	FN	STATUS
35	AR TOTAL POW	16.0632	66.7%	53.3%	18	8	7	9	CAPTOPRIL
36	AR LF	0.2757	59.3%	86.7%	16	13	2	11	CAPTOPRIL
37	AR HF	0.6292	59.3%	60.0%	16	9	6	11	CAPTOPRIL
38	AR LF/HF	0.4397	59.3%	80.0%	16	12	3	11	CAPTOPRIL
39	Organ BAND	6.3784	77.8%	60.0%	21	9	6	6	ATROPINE
40	van Geijn ID	3.8816	63.0%	66.7%	17	10	5	10	ATROPINE
41	Huey STV	100.6193	63.0%	60.0%	17	9	6	10	ATROPINE
42	Huey LTV	121.5879	81.5%	66.7%	22	10	5	5	ATROPINE
43	OSC	14.6160	70.4%	66.7%	19	10	5	8	ATROPINE
44	RMSSD	45.9777	63.0%	66.7%	17	10	5	10	ATROPINE
45	FFT TOTAL POW	7.1042	70.4%	60.0%	19	9	6	8	ATROPINE
46	W1	1.3987	81.5%	66.7%	22	10	5	5	ATROPINE
47	W2	2.7912	70.4%	66.7%	19	10	5	8	ATROPINE
48	de Haan LTI	57.0547	77.8%	85.7%	21	12	2	6	PROPRAN
49	YEH DI	0.0150	59.3%	71.4%	16	10	4	11	PROPRAN
50	YEH II	0.0390	74.1%	85.7%	20	12	2	7	PROPRAN
51	Organ BAND	4.4329	70.4%	92.9%	19	13	1	8	PROPRAN
52	van Geijn ID	4.6170	81.5%	71.4%	22	10	4	5	PROPRAN
53	Huey STV	101.2665	81.5%	71.4%	22	10	4	5	PROPRAN
54	Huey LTV	83.8887	63.0%	78.6%	17	11	3	10	PROPRAN
55	Dalton SD	42.6800	77.8%	64.3%	21	9	5	6	PROPRAN
56	Zugaib STV	0.0072	63.0%	71.4%	17	10	4	10	PROPRAN
57	Zugaib LTV	0.0311	74.1%	78.6%	20	11	3	7	PROPRAN
58	OSC	12.8453	81.5%	71.4%	22	10	4	5	PROPRAN
59	SDNN	21.3372	77.8%	64.3%	21	9	5	6	PROPRAN
60	FFT TOTAL POW	5.7684	77.8%	71.4%	21	10	4	6	PROPRAN
61	W1	5.7684	77.8%	71.4%	21	10	4	6	PROPRAN
62	W2	5.7684	77.8%	71.4%	21	10	4	6	PROPRAN
63	W3	2.4825	77.3%	69.2%	17	9	4	5	PROPRAN
64	Noise	0.0643	59.3%	78.6%	16	11	3	11	PROPRAN
65	AR HF	18.7057	66.7%	64.3%	18	9	5	9	PROPRAN
66	AR LF	0.2617	74.1%	78.6%	20	11	3	7	PROPRAN
67	AR LF/HF	0.4004	74.1%	78.6%	20	11	3	7	PROPRAN

Fig. 6. Threshold levels, corresponding sensitivity, specificity, number of true positives TP, true negatives TN, false positives FP and false negative FN cases in different clinical status-discrimination between CHD and control groups (continued)

be interpreted as insufficient power of used ROC method to detect abnormal cases (patients with CHD). The best discrimination between analyzed groups of patients we obtain after propranolol administration in de Haan LTI, Yeh II, Organ BAND, van Geijn ID, OSC and Huey STV indices. First five of them describe pure long term heart rate variability.

4. LOGISTIC REGRESSION

Logistic regression finds an equation that best predicts a binary outcome variable from one (or more) predictors [5]. In presented paper we analyzed only prediction based on single

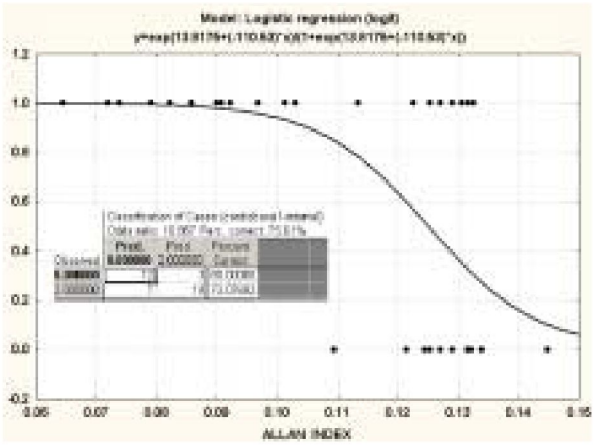

Fig. 7. Exemplary results of logistic regression for Allan index measured in standing position (code 0 - healthy, code 2 - CHD)

Table 2. Usefulness of analyzed indices in discrimination between CHD and healthy people in five different clinical situations - logistic regression results

Parameter	*p*-value	Correct classification [%]	TP	TN	FP	FN	STATUS
RMSSD	0.00192	69.05	21	8	7	6	SUPINE
HUEY LTV	0.00162	75.61	24	7	8	2	STANDING
OSC	0.00107	78.05	24	8	7	2	STANDING
ALLAN	0.00002	75.61	19	12	3	7	STANDING
YEH II	0.01070	73.17	26	4	10	1	PROPRANOLOL
ZUGAIB LTV	0.00476	73.17	26	4	10	1	PROPRANOLOL
W1	0.00223	75.61	25	6	8	2	PROPRANOLOL
W2	0.02720	68.29	24	4	10	3	PROPRANOLOL

variable. In our case binary dependent variable denotes affiliation either to control group or to CHD group. As a loss function maximum likelihood was used. In all cases Quasi-Newton estimation method with convergence criterion 0.00001 was applied (Fig. 7). We obtained fair results only in eight cases. All other logistic models were statistically not significant. However the general classification quality was between 68 and 78%, we observed high number of false negatives FN similarly as in ROC results (Table 2). The superiority of logistic method on previous cited is that it allows for parametrization and construction of simple prediction model.

5. CONCLUSIONS

Three different statistical techniques were used to check usability of single heart rate variability HRV index as a marker in screening test differentiating healthy people with coronary heart disease group. All three methods confirmed that reaction for propranolol administration is prevalent in comparison to other four analyzed clinical situations. It mani-

fested in changes of rather long-term than short-term heart rate variability. Most promising parameters, which usefulness in discrimination was confirmed simultaneously with all mathematical techniques are: RMSSD in supine position, Huey LTV, OSC and Allan indices in standing position, Yeh II, Zugaib LTV, w1 and w2 after administration of propranolol. However prediction of patients status on the basis of univariate analysis is rather poor. Predominant number of false negative misclassifications suggests that probably use of multivariate approach incorporating combinations of HRV indices and some additional parameters will improve quality of classification. Intentionally no one additional (not HRV) descriptor was used in presented analysis. We want to emphasize, that high degree of biological variability effectively makes impossible diagnosis support based exclusively on single descriptor. Because of small sample size obtained preliminary results should be rather treated as methodology discussion than real medical "gold standards".

References

[1] J. Moczko, *Advanced methods of heart rate signals processing and their usefulness in diagnosis support I. Mathematical heart rate descriptors and virtual instrumentation,* Comp. Meth. in Sci. and Tech., **8(2),** 65-76, (2002).
[2] M. Hollander, D. A. Wofe, *Nonparametric Statistical Methods.* John Wiley & Sons, Inc. (1973).
[3] S. W. Smith, *Digital Signal Processing. A practical guide for engineers and scientists.* Newnes-Elsevier Science (2003).
[4] Y. Lu and J.-Q. Fang, *Advanced Medical Statistics,* World Scientific (2003)
[5] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression,* John Wiley & Sons, Inc. (1989)