THE METHOD OF PRINCIPAL COMPONENTS AS AN INSTRUMENT FOR ANALYSIS OF ORAL GLUCOSE TOLERANCE TEST DATES

Avshalumov A.Sh., Filaretov G.F., Markovsky V.B., Sinitsyna E.N.

Moscow Research Institute for Cybernetic Medicine (MRICM)

ABSTRACT

It is proposed to use the method of principal components (MPC) as an instrument for date's analysis of oral glucose tolerance test (OGTT), which plays very important role in diabetic diagnoses. As it was shown on the basis of sufficiently representative clinical material, using of the MPC for OGTT data processing is very perspective. This is established, that basic information about the presence and level of hyperinsulinemia is contained in the first principal component FACTOR1 and FACTOR1 values for healthy patients and patients with signs of pathology are clearly separated. There is a significant correlation values FACTOR1 and standard indices CARO, HOMA-IR, HAFFNER. The dynamics of change in F1 values can clearly judge about the efficiency of therapeutic procedures.

Index Terms - Oral glucose tolerance test, method of principal components, diabetic diagnoses

1. INTRODUCTION

One of the most important problems in the medicine nowadays is the problem of the so called insulin resistance syndrome. Term "insulin resistance syndrome" means sensitivity reduction insulin-dependent tissues to the hormone insulin. This hormone produced by pancreas and such a sensitivity reduction lead to the chronic hyperinsulinemia. Insulin resistance syndrome and hyperinsulinemia are the two main factors for the progress of diabetes mellitus and its complications, as well as the whole number of pathological changes of metabolism and metabolic control united by the term "metabolic syndrome".

Investigations of the insulin resistance syndrome and of the associated pathological states may lead to development of the effective methods of diabetes mellitus prevention as well as the whole range of symptoms of the metabolic syndrome.

Many methods for estimation of the insulin resistance exist. The most popular amongst them is oral glucose tolerance test (OGTT) [1]. This test is realized by the taking fasting of dry 75 g glucose dissolved in 250 ml of water, and measuring concentrations of immunoreactive insulin (IRI), and in enlarged version also glucose and C-peptide in blood plasma at the initial (zero) time and 30, 60, 120 minutes after taking. Standard processing of test results is usually limited to fixing the initial and end levels of IRI and there confrontation [2], estimation of various indexes (CARO, HOMA-IR [3], "area under insulin curve", HAFFNER, etc.). The last indices are considered as the most informative ones and calculate as the sum of IRI concentrations in blood plasma at the initial (zero) time and 30, 60, 120 minutes after, or (for HAFFNER index [1]) as the same sum with weights 0,25, 0,5, 0,75 and 0,5 respectively.

In practice we can see that findings from OGTT in the every individual case reflect good enough type and level of malfunction in insulin secretion and hyperinsulinemia. However, it is clear that such data processing does not allow to extract all useful information especially when IRI, glucose and C-peptide are measured simultaneously. In this case it is desirable to get integral representation about level of variability for each factor, that possibly reveals liaisons with clinical signs of insulin resistance. In one's turn it gives opportunity for a new look to the estimation and objectification of the findings from the insulin resistance diagnostics and to an effectiveness of its correction methods. Further it is proposed to use the method of principal components (MPC) [4] as an instrument for data analysis of oral glucose tolerance test (OGTT).

2. INVESTIGATION OF MPC AVAILABILITY FOR DATA ANALYSIS OF OGTT

As well-known MPC serves for optimal dimensionality reduction of multidimensional initial data by means of the transformation to a new variables (principal components or factors) that are uncorrelated normalized linear combinations of the initial variables. MPC is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. It is assumed that only a few number of principal components (usually first two or three ones) can explain the largest part of common variance. The values in the remaining dimensions, therefore, tend to be highly correlated and may be dropped with minimal loss of information.

Investigation of MPC applicability for data analysis of OGTT included two steps.

Step 1. The aim of this step was to reveal a principal possibility and usefulness of MPC for data processing of OGTT. Thereto it was important to answer a fundamentally question: is it possible with help of MPC to reliably classify data on healthy or sick patients, using all OGTT factors (IRI, glucose and C-peptide).

Clinical material was obtained at the Clinic of the Moscow Institute of Cybernetic Medicine. Sample of 32 patients with the signs of insulin resistance syndrome was considered (group 1). Control group (group 2) included 10 persons.

Results of data processing of OGTT are shown on the plane of two first principal components in Figure 1, where circles represent patients from group 1 and triangles represent control group.

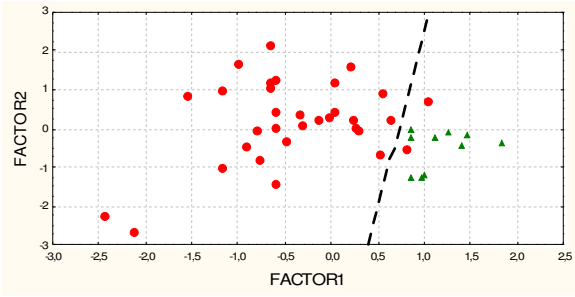

Figure 1: results of data processing (variant 1)

It is clear, although points from distinct groups are generally separated, nevertheless it is quite difficult to draw a boundary line amongst them. For example if one draws a boundary line shown in the Figure 1 only two points get wrong classification (instead of group 1 they get group 2).

In spite of quite favorable result let's improve it by modification of the initial variables. It is proposed to use as initial variables concentrations of IRI, glucose and C-peptide in blood plasma at the initial (zero) time and their differences between relevant values in the two time moments: D1 – is the difference between values in point 30 minutes and zero time; D2- is the difference between values in points 60 and 30 minutes, D3 - is the difference between values in points 120 and 60 minutes and D4 – is the difference between values in points 120 and 0 minutes. One can see that amongst variables D1,…, D4 only 3 ones are independent (each of considering variables may be found from 3 other ones), so one of them may be excluded.

The analysis shows, that the best separation of points is obtained if initial values and differences D2, D3, D4 are used. Results of data processing for this case are shown in the Figure 2.

Here we can see that all the points from the different groups are strictly and simply separated by means of the only one factor (FACTOR 1). Further we will consider exactly this case.

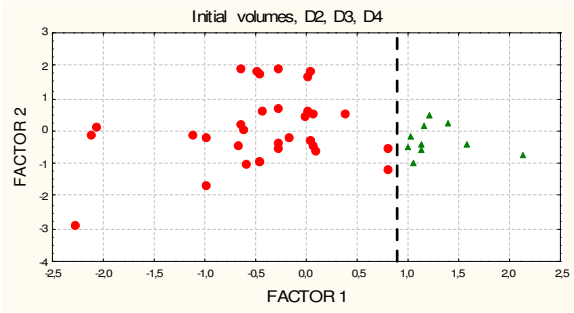

Figure 2: results of data processing (variant 2)

Statistical properties of the derived two-factor model are represented in the Table 1.

Table 1

Components	Paired correlation coefficients	
	FACTOR 1	FACTOR 2
Initial Insulin	***-0,810192***	-0,054413
Initial Glucose	-0,303697	0,470927
Initial C-peptide	***-0,828973***	0,043168
D2 Insulin	-0,514240	0,591197
D3 Insulin	-0,5121118	***-0,741789***
D4 Insulin	***-0,772964***	-0,271312
D2 Glucose	***-0,712550***	0,407881
D3 Glucose	0,116337	***-0,849556***
D4 Glucose	***-0,722621***	-0,391643
D2 C-peptide	-0,5691115	0,621869
D3 C-peptide	-0,637267	-0,558332
D4 C-peptide	***-0,768277***	0,060255
Part of total scattering data	41%	25%
	Total 66%	

Pair correlation coefficients represented in the Table 1. allow to estimate which of the initial variables have the biggest correlation with FACTOR 1 and FACTOR 2. Particularly it is clear that more informative in the term of the correlation with the component FACTOR 1 are initial values of IRI, C-peptide, all variables P4 and variable P2 for glucose (respective significant coefficients highlighted in bold italics in the Table 1). Table 1 also represent such part of the common variance of the all experimental data, that may be explained by means of the two first components. Remaining 34% most probably relate to individual patients organism's features or to IRI, C-peptide and glucose concentration measurement errors.

To find out correlation between derived factors and the standard indices CARO, HOMA, HAFFNER, one should draw a diagram representing experimental points in the respective coordinate system and then the linear models describing relationship between factors and indexes can be built (see Figure 3 and 4).

Correlation between the component FACTOR 1 and the standard indices clearly expressed, in the meantime there is no such an obvious correlation between standard indices and FACTOR 2.

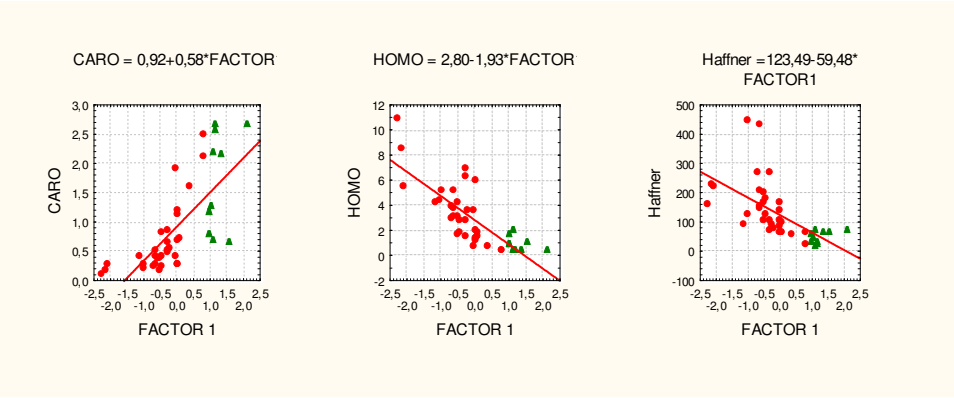

Figure 3: relationship between Factor 1 and standard indexes

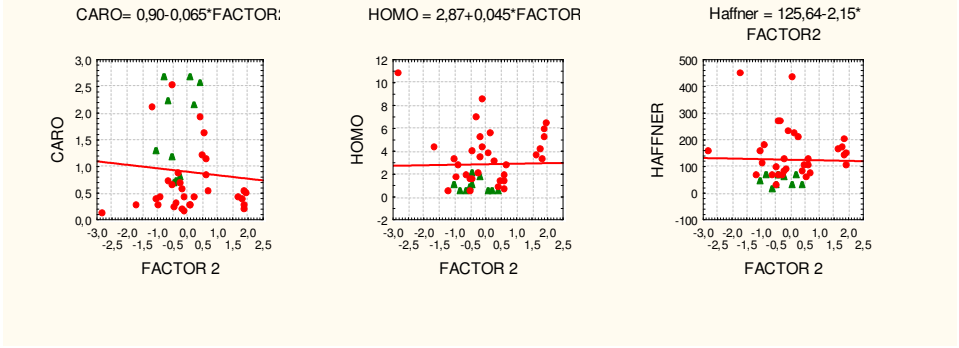

Figure 4: relationship between Factor 2 and standard indexes

Therefore the FACTOR 1 can be used as one more quantitative characteristic of the insulin resistance. Let's call it *FIR index* further. As to the component FACTOR 2 its interpretation is not obviously clear yet and may be an object for the further research.

Self-descriptiveness from the position of the distinction healthy and sick patients for the FIR index obviously higher than for the standard ones. This fact is reflected on the diagram (see Figure 5), where one can see that points from the two different groups (patients and the control group) does not overlap only for FIR index, meanwhile for the other indices there is a considerable intersection. It comes to be clear because of FIR index includes much more initial information than the other ones.

In whole results derived on the *Step 1* show that MPC can be successfully used for data analysis of oral glucose tolerance test.

Step 2. The aim of this step is the further analysis of the effectiveness of the introduced method on the extended patients contingent as well as investigation of its potential to estimate therapeutic procedures effectiveness by observation over a points dynamics on the principal component's plane. This analysis based on the results obtained from the previous step, namely previously found linear relationships have been used. These relationships are associated the initial observed values with principal components (let's call them F1_calc, F2_calc). They provide an opportunity to find the position of experimental points in the plane (F1_calc, F2_calc). Moreover the component F1_calc was shifted on –0.9 in order to bring the point which separates the two groups (patients and the control group) on the zero level for visual clearness: FIR = (F1_calc – 0,9).

Total number of the patients on this step was 52, including 32 patients from the Step 1.

Results of data processing of OGTT for the total set of the patients on the plane (FIR; F2_calc) are shown on the Figure 6. Here rhombs represent new patients, circles represent old patients and triangles represent the control group.

Although as was shown above level of insulin resistance defined only by the component FIR (or F1_calc), for the better visual clarity we use the plane (F1_calc; F2_calc); moreover such approach may be useful for understanding of the FACTOR 2 interpretation.

Obviously clear that points from the different groups are divided as before

For therapeutic procedure effectiveness estimation on the same plane (F1_calc; F2_calc) pares of points representing the initial conditions of the patients and their conditions after the therapy were drawn. Examples of such representation are shown on the Figure7, where arrows mark movement of patients representing points during the therapy. In whole only one patient from the 48 have passed course of the treatment ones, did not show any positive results.

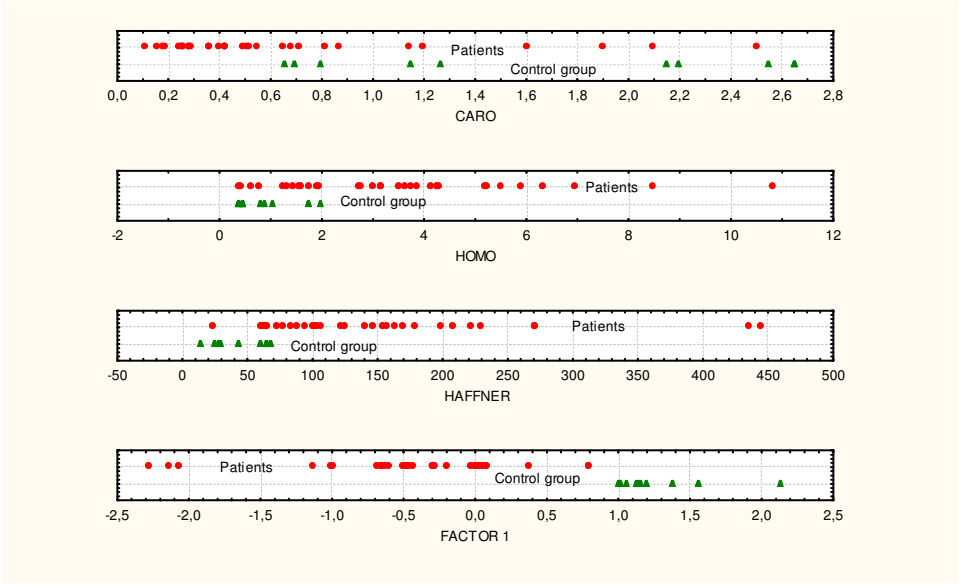

Figure 5: comparison of standard indexes and FIR index

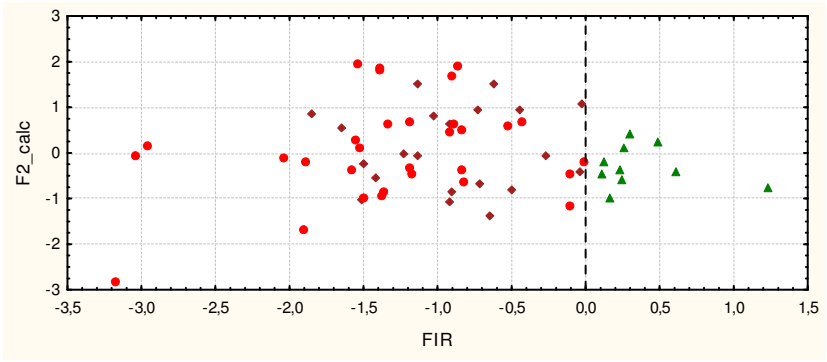

Figure 6: results of data processing for the total set of the patients

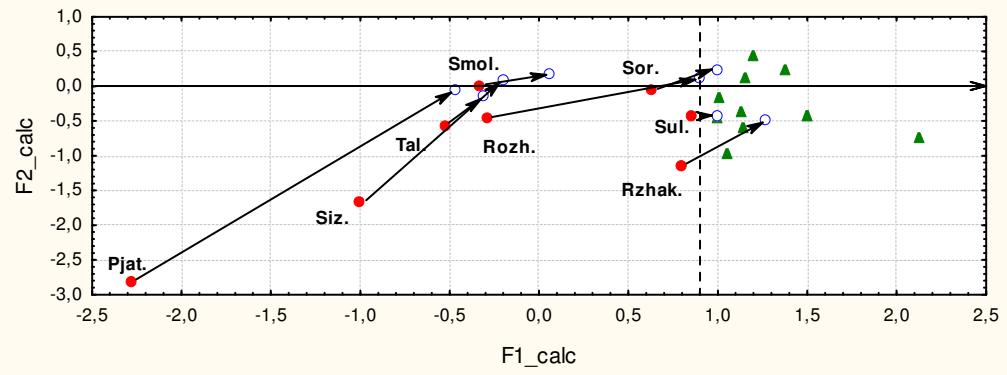

Figure 7: estimation of therapeutic procedures effectiveness

Obtained results indicate that the used method allows to get visual estimation of the effectiveness of the applied treatment and visual estimation of the quality of the resulting effect.

3. ADDITIONAL CONSIDERATIONS

As well-known, veracity and adequacy of the results derived by means of any statistical methods entirely depend on sampling representativeness and on its volume. In this sense MPC is not an exception. Therefore expressions derived from the data processing (data obtained from 42 persons) described above should consider only as some intermediate result. So these expressions need to be corrected while number of the patient increases. At the same time the problem of the method's robustness (in a sense of whether the conclusions derived from smaller sampling will stay correct while the sampling grow) is a very important one.

For the analysis of the result's stability the sample, containing 110 measurement (where 96 obtained

from 48 patients have passed course of the treatment, 4 measurement obtained from 4 patients have refused the treatment and 10 – from the control group) was used. With this sample the new and more accurate linear relationships were obtained to calculate location of experimental points in the plane (FIR; FACTOR 2). As an example of their using the scattering graph for the points from the first sample (analog of Figure 2) is presented on Figure 8).

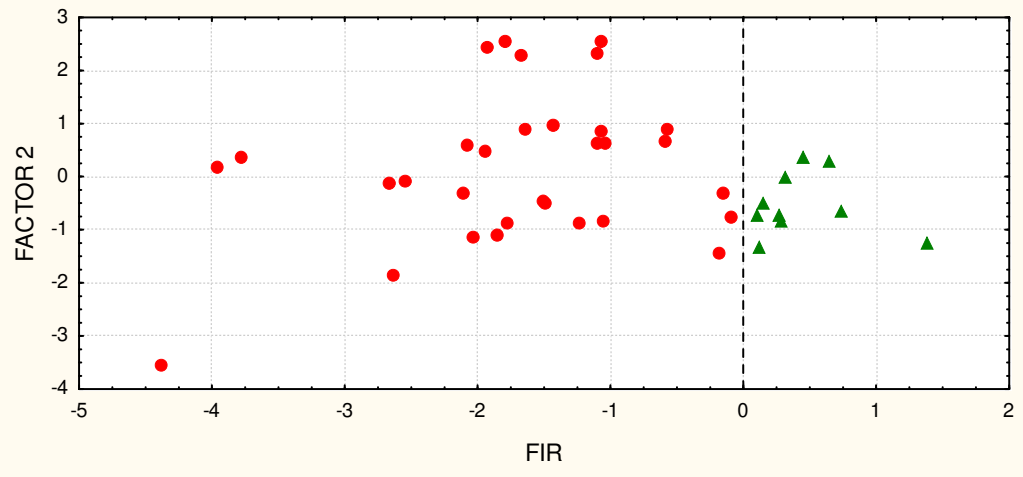

Figure 8: results of data processing (new variant)

Comparison of the above-mentioned diagrams shows that though positions of the points have been slightly changed their general type of the conclusions about distinction of the points from the different groups (patients and the control groups) practically without any changes. It indicates that in the mentioned sense the method is stable (robust) and may be practically used for analysis of insulin resistance. Moreover for the further analysis and practical application more accurate formulas for calculating FIR and FACTOR 2, derived for 110 measurements, should be used. These formulas have the form:

$$FIR = 0{,}922 - 0{,}008*Gl(0) + 0{,}081*Gl(30) - 0{,}083*Gl(60) - 0{,}075*Gl(120) - 0{,}082*Cp(0) + 0{,}043*Cp(30) - 0{,}003*Cp(60) - 0{,}095*Cp(120) - 0{,}022*IRI(0) + 0{,}002*IRI(30) - 0{,}006*IRI(120)$$

$$FACTOR\ 2 = -0{,}989 + 0{,}151*Gl(0) - 0{,}084*Gl(30) + 0{,}257*Gl(60) - 0{,}233*Gl(120) + 0{,}010*Cp(0) + 0{,}091*Cp(30) - 0{,}142*Cp(60) + 0{,}042*Cp(120) + 0{,}002*IRI(0) + 0{,}008*IRI(30) - 0{,}015*IRI(60) + 0{,}008*IRI(120).$$

Here Gl(t), Cp(t), IRI(t) – measured values of the glucose (mmol/l), C-peptide (nmol/l) and IRI (mkME/ml) respectively, where t equals to 0, 30, 60, 120.

These expressions should be considered as the second iteration for determination of the required formulas. They may be corrected again while the sampling grows. As one can expect level of variability of the coefficients in the considering expressions will gradually reduce and their values will tend to some theoretical limit.

4. RESUME

1) Obtained results indicate availability of MPC for data analysis of oral glucose tolerance test (OGTT).

2) Research of the potential of the described method should be proceeded at least in two directions:

- analysis of the boundary line between two sets of points from the two groups: patients group and the control group;
- analysis of the connection between points clustering on the principal components plane and the standard characteristics of the insulin resistance and metabolic syndrome.

5. REFERENCES

[1] Tvorogova M.G., Yaskova K.N., Mychka V.B., Chasova I.E. Insulin resistance syndrome and its diagnostic // Laboratory Medicine, 2003, №6

[2] National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. // Diabetes. - 1979: 28: 1039-1057.

[3] Caro F. Insulin resistance in obese and nonobese man. // J. Clin. Endocrinol. Metab. 1991; vol. 73: 691-695.

[4] Ayvazyan S.A., Mkhitaryan V.S. Applied statistics and fundamentals of economics. Manual for colleges – M.: UNITI, 1998. 1022 p.