Dietary habits of men from Podlasie region of Poland in the years 1987-1998 analysed with self-organizing neural networks

Jamiołkowski J, Szpak A, Pawłowska D

Department of Public Health, Medical University of Białystok, Poland

Abstract

The study was conducted on a group of 556 men. Their dietary habits were evaluated 3 times in the years 1987-1998 by the frequency of consumption of 41 food items during last three months. Obtained frequencies were processed by self-organizing Kohonen neural network, allowing to group persons of similar dietary habits into 3 clusters. After analysing frequencies of consumption of each food item in each cluster, in view of health value, one model was described as appropriate, while remaining were described as two different inappropriate models. In three studies during 11 years, statistically significant increase in frequency of appropriate model was observed. That increase was linked with decrease of occurrence of inappropriate models. Additional verification of described models revealed significant differences between them in nutritive ingredients intake, and also in concentrations of HDL cholesterol in the blood serum of men assigned to those dietary patterns.

Key words: prospective studies, neural networks (computer), diet, atherogenic.

Introduction

Analysis of dietery habits, based on frequencies of various food consumption, includes large number of variables, which have to be evaluated simultaneously. These variables, are tied with complicated relationships [1-5]. These essentials are reasons for complexity of the problem and main source of difficulties in drawing conclusions upon nutrition as a whole and its impact on health condition.

The goal of study is the applying of self-organizing Kohonen neural networks to separate prevalent dietary patterns in studied population. The use of neural network would allow the concentration of diffused information about dietary habits into one variable representing dietary pattern. Such approach enables further investigation of relationships between dietary habits and health condition, especially concentrations of plasma lipids.

Material and methods

Studies were conducted from 1987 to 1998 in group of men settled down in north-eastern region of Poland. Each of 556 men participating in the studies was surveyed three times. The first screening was performed in the years 1987-1989, the second one in 1991-1993 and the third in 1996-1998.

The nutrition of studied persons was examined by means of two methods. To determine dietary habits during past 3 months, food frequency questionnaire was used. Its questions concerned 41 food items. Usual frequency of consumption was recalculated to mean monthly frequency of consumption.

Quantitative estimation of dietary habits involved 24-hour consumption questionnaire. In every case it was polled during definite time of year, namely from 25th of January to 31st of March, to exclude seasonal fluctuations of diet composition. From the information on the consumed products and dishes, including their amounts, the energy value and nutritive value, as well as content of the food groups, were calculated.

All men present in the study had their levels of serum lipids measured. Due to objective reasons, it was difficult to ensure employing the same methods during 9 years of the study. Some of the methods have passed out of use and moreover, manufacturers have desisted the production of obsolescent reagents. In case of necessity of method switching, estimated parameter was measured in a number of samples using both new and previ-

ADDRESS FOR CORRESPONDENCE:
Jacek Jamiołkowski, M.D.
Departament of Public Health, Medical University of Białystok
1 Kilińskiego Str, 15-089 Białystok, Poland
Tel: +48 085 748 55 82

Received 17.01.2005 Accepted 09.02.2005

ous methods. The new method was then adjusted so its output would match the results, obtained using the previous method. The concentrations of triglycerides in blood serum during first stage of the study were estimated using reagents produced by Technicon Instruments Corporation Tarrytown. During second and third stage kits from Ciba-Corning Diagnostics Corp. were used. The concentrations of total cholesterol were determined by Libermann-Burchard method during first and second periods of the study, while Ciba-Corning Diagnostics reagents in enzymatic reaction were used during the third one. The levels of HDL cholesterol in the first and second study were estimated using precipitation method by Lopes-Virelle. In the third study enzymatic method with utilization of Ciba-Corning Diagnostics Corp. Reagents was used. To estimate the concentrations of LDL cholesterol in blood serum Friedewald formula was used:

$$LDL\,cholesterol = total\,cholesterol - HDL\,cholesterol - 0.2 \times Triglycerides$$

The frequencies of consumption of various food items in 3-month period preceding the polling were analysed using Kohonen neural network. Obtained models were validated by comparson in view of quantitative consumption of nutritive ingredients and food groups received from 24-hour consumption questionnaire.

To determine the impact of dietary habits on blood serum lipids, the variable describing dietary pattern, which was generated by Kohonen network was used.

Because of lack of normality in distributions of most variables was observed, statistical reasoning involved nonparametric tests. The dependence of quantitative variables from nutrition model was tested with Kruskal-Wallis procedure and *post hoc* comparisons were executed according to Dunn's procedure [6]. The connection of categorical variables were tested with Pearson's chi-square and chi-square for trend [7] tests estimated. All hypotheses in statistical test were verified under desired type I error (alfa) set on 0.05. Statistical analyses of the data were performed with "Statistica" version 6.0 software from StatSoft.

Results

The use of Kohonen neural network allowed to isolate three dietary patterns, grouping persons of similar dietary habits (*Tab. 1*). The main factors differentiating these models were frequencies of consumption of food items containing fats. The 1st dietary pattern is characterized by high frequency of consumption of fresh margarine, fried vegetable oil and low frequency of consumption of fresh butter, fried lard and bacon. In addition, the 1st dietary pattern presented the most frequent consumption of poultry, fresh fish and fruits. The 2nd dietary pattern demonstrated adverse attributes than the 1st one. Very high frequency of consumption of fresh butter and fried lard was distinctive for that pattern. Furthermore, it indicated higher frequencies of consumption of giblets, smoked second-rate meats, bacon, eggs, sweet baking, sugar, jam, sweets, cookies and sweet beverages. The 3rd dietary pattern was characterised by low frequency of consumption of both butter and margarine, as well as the lowest frequency of consumption of milk and cheese.

Table 1. **Mean monthly frequencies of consumption of 41 food items in three separated dietary patterns**

Food items	Dietary pattern		
	1st	2nd	3rd
White bread	26.7	28.3	27.5
Dark bread	5.9	5.0	4.4
Sweet baking	6.7	8.5	4.8
Flour dishes	7.7	8.0	6.3
Cereal and rice	6.5	6.3	5.5
Milk	18.9	18.5	14.3
Cottage cheese	6.8	8.4	4.9
Cheese	7.6	7.3	4.7
Meat	20.3	22.3	21.8
Poultry	5.0	4.1	4.2
Giblets	2.8	3.2	3.2
Sausages	19.1	19.7	18.5
Smoked luxury meats	6.1	6.0	4.8
Smoked second-rate meats	4.8	5.2	5.1
Bacon	3.8	6.0	5.0
Canned meat	1.6	2.0	1.8
Canned fish	3.1	3.1	2.8
Fresh fish	4.2	2.6	2.9
Eggs	5.3	6.8	5.3
Fresh butter	2.2	26.4	6.3
Fried butter	1.0	2.0	1.3
Fresh margarine	26.9	1.3	3.0
Fried margarine	6.5	7.2	5.5
Fresh lard	1.2	1.9	2.5
Fried lard	5.9	14.3	10.6
Fresh oil	8.7	9.7	8.4
Fried oil	9.8	6.1	6.7
Potatoes	22.5	24.7	23.4
Boiled vegetables	8.6	9.3	8.5
Raw vegetables	13.0	15.5	13.2
Leguminous plants	3.6	4.2	4.1
Fruit	18.2	14.0	12.3
Sugar	24.7	27.9	22.8
Jam	5.1	6.7	3.8
Honey	3.0	2.8	2.1
Sweets	6.8	8.3	4.9
Cookies	6.7	7.7	4.6
Sweet beverages	18.8	20.1	17.2
Beer	4.7	3.9	4.6
Wine	0.6	0.8	0.6
Vodka	2.5	3.1	3.2

Described above patterns were created based on frequencies of consumption of various food items during past 3 months. It was purposeful to extend the characteristics of these models by information gained from 24-hour questionnaire, that is quantitative composition of the diet. *Tab. 2* and *Tab. 3* present energy value, nutritive ingredients and groups of products of three discussed patterns. Performed analyses reveal that, energy value, nutritive ingredients values and consumption of most groups of products differed significantly between patterns. There were no significant distinction only for consumption of alcohol, fruits and vegetables containing vitamin C and leguminous plants. The

Table 2. **Influence of dietary pattern on medians of energy value and nutritive values. Only significant p values are stated**

Variable	Dietary pattern			p value*	Pairs of patterns differing significantly**
	1st	2nd	3rd		
Energy [kcal]	2595.6	2946.1	2434.6	0.0001	1-2.2-3
Animal protein [g]	48.2	53.4	47.2	0.0001	1-2.2-3
Vegetable protein [g]	30.3	31.7	28.7	0.0001	1-3.2-3
Animal fat [g]	70.9	99.8	79.2	0.0001	1-2.1-3.2-3
Vegetable fat [g]	25.2	14.7	12.9	0.0001	1-2.1-3
Saccharosis [g]	52.6	71.1	55.2	0.0001	1-2.2-3
Carbohydrates without saccharosis [g]	263.3	274.7	243.7	0.0001	1-2.1-3.2-3
Calcium [mg]	516.7	511.7	385.9	0.0001	1-3.2-3
Ferrum [mg]	13.8	14.9	13.6	0.001	1-2.2-3
Vitamin A [μg]	627.9	690.6	557.4	0.0001	1-2.1-3.2-3
Vitamin B_1 [mg]	1271.0	1449.1	1281.0	0.0001	1-2.2-3
Vitamin B_2 [mg]	1326.4	1476.1	1254.8	0.0001	1-2.2-3
Vitamin C [mg]	45.2	49.6	46.2	0.025	1-2.
Alcohol [g]	0.043	0.029	0.071		

* Kruskal-Wallis test
** Dunn's test

Table 3. **Content of groups of foods in three separated dietary patterns. Only significant p values are stated**

Groups of foods	Dietary pattern			p value*	Pairs of patterns differing significantly**
	1st	2nd	3rd		
Corny products [g]	258.1	266.0	227.4	0.0001	1-3.2-3
Milk and dairy products [g]	250.0	285.5	22.2	0.001	1-3.2-3
Eggs [g]	5.2	7.3	5.6	0.003	1-2
Meat and fish [g]	279.9	307.7	290.2	0.002	1-2
Butter [g]	5.0	22.5	6.0	0.0001	1-2.1-3.2-3
Other fats [g]	44.2	32.6	31.5	0.0001	1-2.1-3
Potatoes [g]	462.0	462.0	462.0		
Vegetables and fruits containing vitamin C [g]	50.0	44.8	47.4		
Vegetables and fruits containing carotene [g]	57.2	60.5	50.0	0.006	2-3
Other vegetable and fruits [g]	237.2	279.4	229.8	0.001	1-2.2-3
Leguminous vegetables [g]	0	0	0		
Sweets [g]	56.6	77.6	60.0	0.0001	1-2.2-3

* Kruskal-Wallis test
** Dunn's test

Table 4. **Influence of dietary pattern on medians of blood lipids profile in each period of study. Only significant p values are stated**

Period of study	Variable	Dietary pattern			p value*	Pairs of patterns differing significantly**
		1st	2nd	3rd		
1st	Total cholesterol [mg/dl]	206	214	217		
	LDL cholesterol [mg/dl]	136.4	135.85	139.4		
	HDL cholesterol [mg/dl]	48	47.15	47.8		
	Triglycerides [mg/dl]	124	124	126		
2nd	Total cholesterol [mg/dl]	209.5	215	218		
	LDL cholesterol [mg/dl]	135.15	142.8	142		
	HDL cholesterol [mg/dl]	47	44	45.4	0.001	1-2.2-3
	Triglycerides [mg/dl]	122	122	123		
3rd	Total cholesterol [mg/dl]	212	202.5	214		
	LDL cholesterol [mg/dl]	132.6	128.8	132.2		
	HDL cholesterol [mg/dl]	49	51	49		
	Triglycerides [mg/dl]	121	104.5	129		

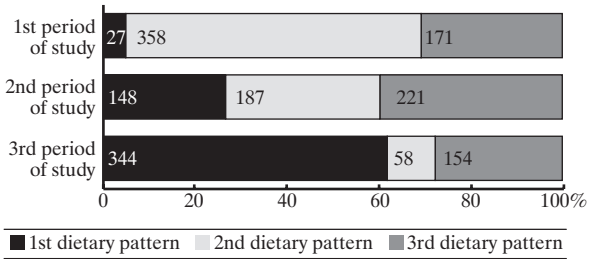
* Kruskal-Wallis test
** Dunn's test

1st dietary pattern was characterised, compared to 2nd dietary pattern, by lower energy intake, lower content of animal fat and saccharosis, whereas higher content of vegetable fat. In addition 1st pattern's diet contained lower quantities of eggs, meat, butter and sweets, and higher quantity of "other fats" (this group includes all fats except for butter, amongst other vegetable oil and margarine) than 2nd dietary pattern. Diet in 3rd pattern was characterised by lowest intake of calcium and content of milk compared to remaining patterns.

After all men have been classified to one of 3 dietary patterns, concentrations of triglycerides, cholesterol and its fractions in blood serum were compared between patterns. Because of changes of methods used to measure these parameters in the course of studies, the comparisons were conducted separately for each of three periods of the study – *Tab. 4*. In the second study, persons qualified by Kohonen network to 1st dietary pattern had significantly higher HDL cholesterol level than persons in 2nd pattern. There were no significant differences in concentrations of triglycerides and total and LDL cholesterol between patterns observed. The most likely reason, that levels of none of measured blood serum lipids differed significantly in the first and the third study, are considerable disproportions in occurrences of 1st and 2nd dietary pattern.

Fig. 1 presents counts of men classified to one of three dietary patterns in each of three periods of study. The frequencies of occurrence of the 1st dietary pattern in subsequent periods of study showed significant rising trend ($p<0.001$), while the 2nd pattern presented significant falling trend ($p<0.001$). The percentage of men qualified to the 1st dietary pattern rose from 4.9% to 61.9% and the percentage of men assigned to the 2nd dietary pattern fell from 64.4% to 10.4%.

Figure 1. **Occurrence of three separated dietary patterns in subsequent studies during 9-year observation**



Discussion

The objective of most neural networks is to find best fitted function transforming input vector data into output vector. Kohonen network in contrary to other types of neural networks, require to train only input data. In the process of learning, network by itself isolates repetitive patterns and learns to recognize them. For this reason, described method is well suited for partitioning large set of data into subsets consisted of similar entities [8,9]. In last 10 years time period, it was unable to find in Medline database resources any publication covering the topic of application of evaluating dietary habits using neural networks, especially Kohonen networks. Just few works applied this method in studies on food itself [10-16].

The application of Kohonen neural network allowed to partition studied group of men into 3 subsets, according to frequencies of consumption of various food items. The significance of that division was verified by estimating its influence on other variables, that were not used in process of learning the network, thus potentially independent of that division. Significant differences among patterns for these variables indicates, that assignment of persons to patterns is not random. The attributes of dietary patterns let describe 1st pattern as beneficial in view of atherosclerosis prevention, 2nd pattern as unfavourable in view of atherosclerosis and obesity prevention and the 3rd one as unfavourable in view of osteoporosis prevention.

The grouping obtained using self-organizing neural network allowed to reveal favourable changes in the dietary habits of studied men during 9 years. In subsequent periods of study, the 1st dietary pattern, described as advantageous in view of atherosclerosis prevention, were displacing the 2nd pattern, likely to stimulate atherosclerosis development.

The work demonstrated relations between dietary habits and quantitative composition of diet, as well as blood serum lipids profile. The blood level of HDL cholesterol of men from the 1st dietary pattern – beneficial in view of atherosclerosis was significantly higher, than in group, that indicated atherogenic features of nutrition – the 2nd dietary pattern.

The advantage of presented method is, that it simultaneously analyses all variables describing dietary habits by frequencies of consumption of various food items and defines homogenous subsets – dietary patterns. Separate analyses of frequencies of consumption of each food item are biased by probability of making an error while testing hypotheses (desired type I error – alfa, usually set for 0.05). While drawing final conclusions from many partial, biased conclusions, these errors accumulate and value of such general conclusions is decreased.

It seems, that application of multivariate analysis methods, such as Kohonen neural networks, in the field of nutritional science has relevant advantages and improves available possibilities of reasoning and generalization.

Conclusions

1. Prospective study of men from Podlasie region in Poland, conducted from 1987 to 1998 indicated significant, beneficial changes of their dietary habits, described by frequencies of consumption of various food items. The percentage of men qualified to 1st dietary pattern, that is characterized by high frequency of consumption of food rich in vegetable plants and low frequency of consumption of foods containing large quantities of animal fat, rose significantly from 4.9% to 61.9%. The percentage of men classified to 2nd dietary pattern, that is indicating high frequency of consumption of food rich in animal plants and low frequency of consumption of foods containing vegetable fat, fell significantly from 64.4% to 10.4%. These changes are advantageous in view of prevention of atherosclerosis, cardiovascular disease and obesity.

2. Dietary pattern determined by self-organizing neural network is variable, that encloses differences in dietary habits, estimated by frequencies of consumption of food items, in a synthetic way. Dietary habits are treated as a whole and not as a set of separate components. The use of such created variable – dietary pattern – allows to define relations between dietary habits and other attributes, characterizing health state.

References

1. Tokudome S, Imaeda N, Tokudome Y, Fujiwara N, Nagaya T, Sato J, Kuriki K, Ikeda M, Maki S. Relative validity of semi-quantitative food frequency questionnaire versus 28 day weighed diet records in Japanese female dietitians. Eur J Clin Nutr, 2001; 55: 735-42.

2. Sygnowska E, Waśkiewicz A, Pardo B. Zmiany zwyczajowego sposobu żywienia populacji Warszawy objętej programem Pol-MONICA w latach 1984-1993. Żywienie Człowieka i Metabolizm, 1997; XXIV: 234-48.

3. Prevost T, Whichelow MJ, Cox BD. Longitudinal dietary changes between 1984-5 and 1991-2 in British adults: associations with socio-demographic, lifestyle and health factors. Br J Nutr, 1997; 78: 873-88.

4. Shea S, Melnik TA, Stein AD, Zansky SM, Maylahn C, Basch CE. Age, sex, educational attainment, and race/ethnicity in relation to consumption of specific foods contributing to the athergenic potential of diet. Prev Med, 1993; 22: 203-18.

5. Pardo B, Piotrowski W. Analysis of the relationship between age, body mass, smoking, nutritional habits and plasma lipids in Pol-MONICA Warsaw population. Żywienie Człowieka i Metabolizm, 1992; XIX: 3-10.

6. Dunn OJ. Multiple comparisons using rank sums. Technometrics, 1964; 6: 241-52.

7. Armitage P. Metody statystyczne w badaniach medycznych. Warszawa, PZWL; 1987.

8. Kohonen T. Self-organizing maps. New York, Springer-Verlag New York, Inc, Secaucus; 1997.

9. Osowski S. Sieci neuronowe w ujęciu algorytmicznym. Warszawa, Wydawnictwa Naukowo-Techniczne; 1997.

10. Bloch HA, Petersen M, Sperotto MM, Kesmir C, Radzikowski L, Jacobsen S, Søndergaard I. Identification of barley and rye varieties using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry with neural networks. Rapid Commun Mass Spectrom, 2001; 15: 440-5.

11. Sørensen HA, Sperotto MM, Petersen M, Kesmir C, Radzikowski L, Jacobsen S, Søndergaard I. Variety identification of wheat using mass spectrometry with neural networks and the influence of mass spectra processing prior to neural network analysis. Rapid Commun Mass Spectrom, 2002; 16: 1232-7.

12. Søndergaard I, Jensen K, Krath BN. Classification of wheat varieties by isoelectric focusing patterns of gliadins and neural network. Electrophoresis, 1994; 15: 584-8.

13. Jacobsen S, Nesić L, Petersen M, Søndergaard I. Classification of wheat varieties: use of two-dimensional gel electrophoresis for varieties that can not be classified by matrix assisted laser desorpiton/ionization-time of flight-mass spectrometry and an artificial neural network. Electrophoresis, 2001; 22: 1242-5.

14. González G, Pena Méndez EM, Sánchez Sánchez MJ, Havel J. Data evaluation for soft drink quality control using principal component analysis and back-propagation neural networks. J Food Prot, 2000; 63: 1719-24.

15. Díaz C, Conde JE, Estévez D, Pérez Olivero SJ, Pérez Trujillo JP. Application of multivariate analysis and artificial neural networks for the differentiation of red wines from the Canary Islands according to the island of origin. J Agric Food Chem, 2003 Jul 16; 51: 4303-7.

16. Frías S, Conde JE, Rodríguez MA, Dohnal V, Pérez-Trujillo JP. Metallic content of wines from the Canary Islands (Spain). Application of artificial neural networks to the data analysis. Nahrung, 2002; 46: 370-5.