

On an algorithm for decision-making for the optimization of disease prediction at the primary health care level using neural network clustering

PETRO SELSKYY^{1, A, C-F}, DMYTRO VAKULENKO^{2, D, F}, ANATOLII TELEVIK^{3, D-F}, TARAS VERESIUK^{4, D-F}

¹ Department of Pathological Anatomy with Forensic Medicine, I. Horbachevsky Ternopil State Medical University, Ternopil, Ukraine

² Medical Informatics Department, I. Horbachevsky Ternopil State Medical University, Ternopil, Ukraine

³ Department of Human Anatomy, I. Horbachevsky Ternopil State Medical University, Ternopil, Ukraine

⁴ Ternopil District Hospital, Ukraine

A – Study Design, **B** – Data Collection, **C** – Statistical Analysis, **D** – Data Interpretation, **E** – Manuscript Preparation, **F** – Literature Search, **G** – Funds Collection

Summary Background. A number of studies have been aimed at solving the problems of introducing information technology and systems, but the questions of informatization in rural medicine are not completely resolved. It is important to optimize the prognosis of diseases using available and inexpensive information methods for the improvement of primary healthcare.

Objectives. The aim of our study was to develop an algorithm to optimize the decision-making prognosis of disease at the primary health care level based on information methods.

Material and methods. The data used for analysis originated from the survey results of 63 patients with hypertension in educational and practical centers of primary health care (EPCPHC) of Ternopil region (Ukraine). For a deeper analysis and clustering, the neural network approach was used with the NeuroXL Classifier add-in application for Microsoft Excel.

Results. Thirteen (19.40%) patients experienced health deterioration and the development of complications. It has been established that neural network clustering could effectively and objectively allocate patients to the appropriate category in terms of the average survey results. Cluster analysis results have shown that the combination of high blood pressure (systolic, diastolic and pulse) gave reason to anticipate the deterioration of patients' conditions.

Conclusions. A decision algorithm was created in order to optimize the prediction of diseases at the primary health care level, and also to correct examination and treatment based on an analysis of average values of patients' examination and the use of neural network clustering.

Key words: primary health care, hypertension, algorithms, neural networks (computer), cluster analysis.

Selskyy P, Vakulenko D, Televiak A, Veresiuk T. On an algorithm for decision-making for the optimization of disease prediction at the primary health care level using neural network clustering. *Fam Med Prim Care Rev* 2018; 20(2): 171–175, doi: <https://doi.org/10.5114/fmPCR.2018.76463>.

Background

Information technology and systems in medicine have become increasingly important [1–3]. The medical sector is closely linked with the accumulation and processing of large volumes of information. In order to ensure the proper quality of medical care, there is a need for the accumulation and analysis of data over a long period of time, and the use of electronic expert systems [2]. The volumes of medical information are constantly increasing due to the description of new pathological states, syndromes, the introduction of registers of new nosologies, diagnostic methods and medicines. Therefore, among the most important health problems is its informatization. According to WHO A58/21, 'eHealth' "Opens a unique opportunity for the development of public health. Strengthening healthcare through the eHealth system can contribute to the implementation of basic human rights through increased equity, solidarity, quality of life and quality of health care".

Scientific research on the economic effectiveness of the introduction of medical information systems (MISs) suggests savings of up to 60% of time for medical personnel to keep records, and for over-data collection – up to 50% [1, 4]. According to a report from the American Institute of Medicine, about

44 000 deaths annually in the United States are the cause of diagnostic errors. According to the data of the indicated institute, electronic document management allows not only an increase in efficiency and reduction of the cost of treatment, but also the reduction of the number of medical errors. National European health informatization programs aim at uniting medical facilities into one network, and providing for the electronic storage of medical data of the patient throughout life. A number of studies have been aimed at solving the problems of introducing innovative medical information technology and procedures [1, 5–9], but the questions of informatization in rural medicine are not completely resolved.

At the primary health care level the most significant criteria are information support for the prevention and early diagnosis of disease, timely medical care for patients in different risk groups and for persons against socially significant diseases, effective use of resources in health care, support in medical decision-making, conducting medical examinations for persons who are unable to attend health care facilities, quality training, and carrying out effective research [2]. Thus, it is important to optimize the prognosis of diseases using available and inexpensive information methods for the improvement of primary health care.



Objectives

The aim of our study was to develop an algorithm to optimize the decision-making prognosis of disease at the primary health care level based on the analysis of average cardiac values of patients with hypertension, correlation coefficients, and results of the multiparameter neural clustering.

Material and methods

Study design

The exploration of predictors of patients with hypertension residing in rural areas was conducted with the standard statistical methods, correlation analysis, as well as with neural network clustering. Resulting from the outcome of the investigation, information modeling and an algorithm for decision making for the optimization of disease prediction at the primary health care level were developed.

Setting

The data used for analysis originated from the survey results of patients admitted to educational and practical centers of primary health care (EPCPHC) in the villages Hnylycia and Zarubyntsi, Ternopil region (Ukraine). These studies were carried out during the first two years of operation of these centers (2011–2012).

The research received a positive opinion from the Commission on Bioethics of I. Horbachevsky Ternopil State Medical University, Ministry of Health of Ukraine, dated 23 October, 2017 (protocol No. 43).

Participants

The exploration of predictors of 63 patients with hypertension who comprised the general study group was conducted. The group comprised 50 patients with stable disease course, and 13 patients with deterioration in their condition. The control group was composed of 19 residents of these rural areas without any disease.

Variables

All patients with hypertension and the ones in the control group were included in the analysis. A separate study of the indicators of patient groups with stable disease course and with deterioration in their condition was also performed. The group of patients with deterioration in their condition was determined by the second survey of the signs of a more severe stage of the disease or its complications.

Bias

Patients were included in the general study group by random sampling. Random sampling was also performed in the control group of the individuals who came to EPCPHC and who did not have any pathology. The indicators obtained under the same conditions were used in the study.

Data sources

This research utilized the diagnostic findings that were input by the medical staff of the EPCPHC into the "Registry" [10] database program during the patients' first visit and during the period of six to nine months of treatment (hereinafter referred to as the second survey). The "Registry" program makes provision for the input of personal details for all residents of these villages, the recording of diagnostic and medical procedures, and storage of patients' survey results.

Study size

The study size was calculated by representative sampling of the general group. From this group the research groups were formed. Each next stage of the study was carried out taking into account the results of the previous one. By generalization of the results of the use of statistical methods and neural network clustering, an algorithm for decision-making was developed.

Quantitative variables

The following quantitative variables were handled in the analyses: relative proportions (in percentages) of study groups, men and women separately, with the average values of age, sex, position of the electric axis of the heart, heart rate, systolic blood pressure, diastolic blood pressure, pulse blood pressure value. The study of correlation indicators was conducted, as well as the parameters study for the algorithm for neural clustering offered by the program NeuroXL Classifier.

Statistical methods

Statistical analysis of the data was conducted using the software package "Microsoft Excel" (Microsoft Office 2007). The statistical significance of differences between the arithmetic average and relative values was estimated by Student's *t*-test (*t*) for the normal distributed data set. During the comparison of similar groups a liaison analysis of the correlation coefficient (*r*) was conducted by Pearson's method.

For a deeper analysis and clustering of patients in a group in order to optimize the prognosis of pathology at the primary level, the neural network approach was used with the NeuroXL Classifier add-in application for Microsoft Excel. The neural network is a proven and very common technology used to solve complex classification problems. They are modeled on the basis of the human brain and are composed of interconnected networks of independent processors that are arranged by changing the connections (a process known as training) through learning and problem-solving processes. The program NeuroXL Classifier (development of the company AnalyzerXL) creates self-organizing neural networks, which perform categorization by studying trends and correlations within the data. Despite the high performance, neural networks are not often used because of their complexity and the extensive training which is necessary for their proper implementation. The NeuroXL Classifier eliminates such barriers, using the complexity of methods based on already known neural networks, and taking advantage of the Microsoft Excel computational engine. The key advantages of using the NeuroXL Classifier are simplicity in operation, optional in-depth knowledge database of the neural networks, integration with Microsoft Excel, provision of reasonable neural network technology for high-precision classification, and the determination of trends and correlations that cannot be shown by traditional methods [11, 12].

Results

Participants

A total of 63 patients (general group) were examined at first examination and the second survey, comprising – 15 (23.81%) males and 48 females (76.19%). Among them, 50 patients (79.37%) had disease progression classified as stable (first group), while the remaining 13 patients (20.63%) suffered worsening hypertension and the development of complications (second group).

Descriptive data

All of the people under test, including the control group, were rural residents, who lived permanently in the respective

settlements. All patients were farm workers. No occupational hazards were found.

Outcome data

The average age of patients in the general group was (64.30 ± 1.81) years. The average position of the electrical axis of the heart was $(37.83 \pm 1.92)^\circ$. At first examination, the average heart rate was (78.24 ± 1.15) beats per minute. Blood pressure values were as follows: systolic – (154.76 ± 2.29) mm Hg. Art., diastolic – (92.94 ± 1.04) mm Hg. Art., pulse pressure – (61.83 ± 1.95) mm Hg. Art.

The average age of the first group was (63.76 ± 1.80) years. The group of patients with stable disease course comprised $(72.0 \pm 6.35)\%$ women. The average electrical heart axis provisions value was $(37.72 \pm 2.00)^\circ$. Average heart rate during the first examination was (77.28 ± 1.17) beats per minute. Blood pressure values were as follows: systolic – (155.80 ± 2.29) mm Hg. Art., diastolic – (93.10 ± 1.01) mm Hg. Art., pulse pressure – (62.70 ± 1.95) mm Hg. Art.

The average age of the second group was (66.38 ± 3.74) years. The group of patients with deterioration in their condition comprised $(92.31 \pm 7.69)\%$ women. The average electrical heart axis provisions value was $(38.23 \pm 2.67)^\circ$. The average heart rate during the first examination was (81.92 ± 1.64) beats per minute. Blood pressure values during the first observation of patients with deterioration of hypertension were as follows: systolic – (150.77 ± 4.28) mm Hg. Art., diastolic – (92.31 ± 2.40) mm Hg. Art., pulse pressure – (58.46 ± 3.74) mm Hg. Art.

The control group was composed of 11 males and 8 females. The average age in the control group was (38.21 ± 3.91) years. The average heart rate was (77.58 ± 1.02) beats per minute. Blood pressure values were as follows: systolic – (123.42 ± 1.49) mm Hg. Art., diastolic – (78.42 ± 1.32) mm Hg. Art., pulse pressure – (45.00 ± 1.58) mm Hg. Art.

Main results

During the first examination in all groups with hypertension blood pressure values were higher ($p < 0.05$) in comparison with those of the control group. The average heart rate was significantly higher ($p < 0.05$) for the patients with deterioration of the condition, but not so different ($p > 0.05$) for the patients in the group with stable disease course, and in the general group compared to the same period in the control group.

Comparative analysis was also conducted in the first and the second groups of patients. It was shown that the group of patients with deterioration in their condition (second group) was also dominated by female patients, and this percentage was significantly higher compared to the group with stable disease course ($p < 0.05$). The average age of patients did not significantly differ ($p > 0.05$). The average value for the position of the electrical axis of the heart also did not significantly differ ($p > 0.05$). The average heart rate was significantly higher ($p < 0.05$) compared with the group with stable disease course.

Other hemodynamic parameters did not differ compared with the group of non-progressive hypertension ($p > 0.05$).

At the second survey in the general group of patients, the average indicator heart rate ((78.59 ± 1.07) beats per minute) and pulse pressure ((57.06 ± 1.57) mm Hg. Art.) did not differ statistically ($p > 0.05$) from the analogical indicators at the first examination. The indicators of systolic and diastolic pressure were much lower in comparison to the prior analogical indicators at the first examination: systolic – (145.86 ± 2.01) mm Hg. Art. ($p < 0.01$), diastolic – (88.49 ± 1.08) mm Hg. Art. ($p < 0.001$).

During the second survey in the group of patients with stable disease course (first group), the average heart rate ((78.62 ± 1.07) beats per minute) was not statistically different from the similar value obtained during the first examination ($p > 0.05$). Blood pressure values were significantly lower compared with those before medical treatment: systolic – (145.28 ± 1.84) mm Hg. Art. ($p < 0.01$), diastolic – (88.30 ± 1.04) mm Hg. Art. ($p < 0.01$) and pulse pressure – (56.60 ± 1.36) mm Hg. Art. ($p < 0.05$).

During the second survey in the second group, the average heart rate was at (78.46 ± 2.20) beats per minute, and it was statistically different ($p > 0.05$), while blood pressure did not vary: systolic – (148.08 ± 5.51) mm Hg. Art., diastolic – (89.23 ± 2.49) mm Hg. Art. and pulse pressure – (58.85 ± 4.86) mm Hg. Art. ($p > 0.05$), compared to the measurements before treatment (during the first survey).

Correlation analysis showed a direct correlation between the number of hemodynamic values during the first and second surveys of the patients in the group with stable hypertension course (pulse ± 0.5 , blood pressure, diastolic ± 0.3 , pulse ± 0.1) and deterioration (pulse ± 0.6 , blood pressure systolic ± 0.5 , diastolic ± 0.7 pulse ± 0.3).

In order to establish the value of a combination of changes in certain parameters for prognosis of the disease, the neural clustering of patients with hypertension was performed based on the following parameters: Age – age, s – sex, ECG – position of the electric axis of the heart, 1 Ps – heart rate during the first survey, 1 max – systolic blood pressure value during the first survey, 1 min – diastolic blood pressure value during the first survey, 1 d – index of pulse blood pressure during the first survey, 2 Ps – heart rate during the second survey, 2 max – systolic blood pressure value during the second survey, 2 min – diastolic blood pressure value during the second survey, 2 d – index of pulse blood pressure during the second survey. Index of patients' changes ("change") was recorded as "1" in front of each case of patient deterioration, and "2" – with stable hypertension course. The parameters for the algorithm of neural clustering were offered by the program. The permitted number of clusters was chosen equal to five.

Figures 1a and 1b show some results of the data analysis for the patients with hypertension. The largest proportion of patients (30.16%) was found in the 5th cluster (Figure 1a). This cluster is formed from the largest proportion of patients with deterioration (26.32%) compared to 1–4 clusters (1st – 21.43%, 2nd – 25.0%, 3rd – 11.11% and 4th – 11.11%).

Clusters weights

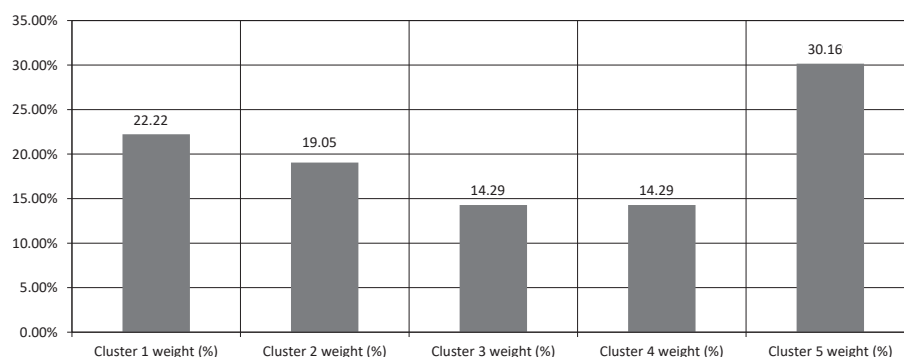


Figure 1. Results of clustering of patients with hypertension:

a) cluster size – the percentage of patients in a cluster;

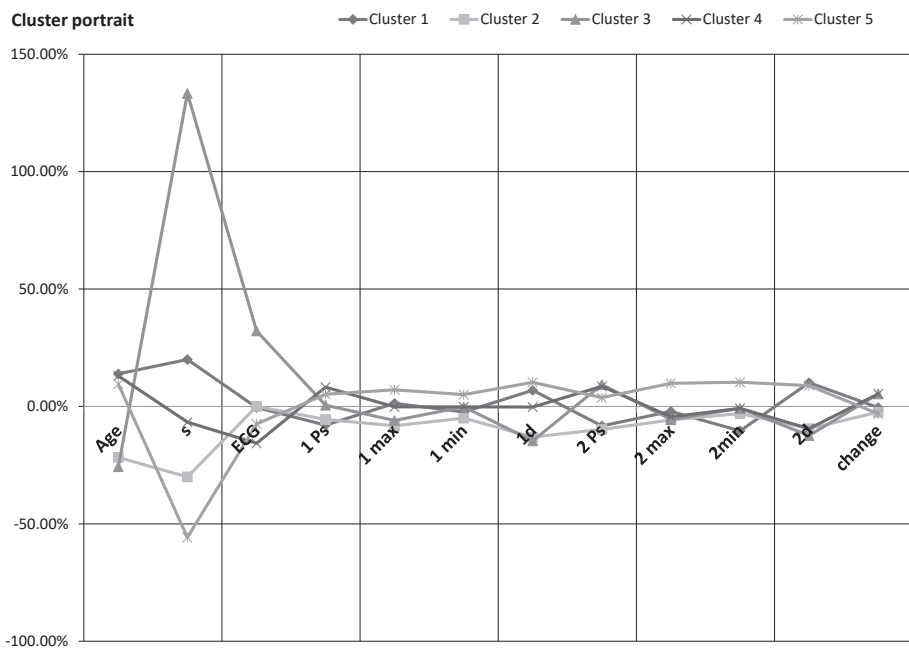


Figure 1. Results of clustering of patients with hypertension: b) cluster schema – values within the parameters of chosen clusters

Using the cluster analysis (Figure 1b), it can be determined that the fifth cluster, along with the lowest “change”, showed the highest systolic (1 max, 2 max), diastolic (1 min, 2 min) and pulse (1 d, 2 d) blood pressure in the first and second surveys, compared to all other clusters. However, while the values for the parameters of age (Age) and pulse (Ps) were high in this cluster, they were not significantly higher than in certain other clusters, including: “Age” – the 1st and 4th, “1 Ps” – 4th and “2 Ps” – 3rd and 4th clusters.

Discussion

Key results

As the results of the study show, the simple analysis based on average values and the correlation coefficients between age, the position of the electrical axis of the heart, and series of hemodynamic parameters is the primary tool among researchers. But such simple data processing makes it impossible to determine a combination of changes for certain parameters in order to predict disease course and outcome, such as further deterioration or improvement. At the same time, neural network clustering can effectively and objectively allocate patients to the appropriate category – either deterioration or with stable outcome. Cluster analysis results have shown that the combination of high blood pressure (systolic, diastolic and pulse) gave reason to anticipate the deterioration of patients’ conditions, while the combination of high age and heart rate (tachycardia) were essential, but not a priority for the prediction.

Tactics phased analysis of indicators of examination of patients with hypertension has been proposed in the form of an algorithm for decision-making in order to optimize the prediction of disease and to improve patient examination and treatment (Figure 2).

During the primary level examination and after the results of the verbal consultation, including teleconsultation, the required indicators are recorded and analyzed, in order to predict disease progression, including using the results of cluster portraits analysis. For identifying groups of indicators and their combined

changes which are important for prognosis, cluster analysis of data from the observation in the region is recommended.

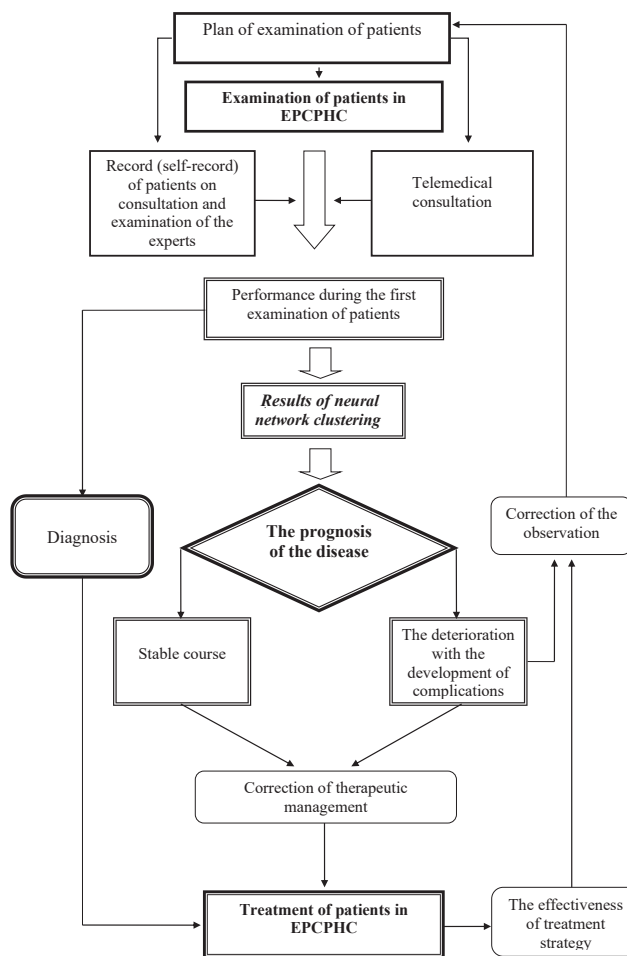


Figure 2. Decision algorithm for the optimization of disease prediction at primary health care facilities in order to correct examination and treatment

Limitations of the study

The proposed approach (as exemplified by the case of patients with hypertension) can be widely used to optimize the prediction of the course of diseases at the primary level of provision of medical care. At the first stage of its introduction to the family doctor, the assistance of a specialist in informatics will be necessary. However, after conducting neural network clustering using the original data array, the results can be applied in the future, according to the proposed algorithm, without the necessity for expert opinion in each clinical case.

Conclusions

Therefore, this article describes a method of analysis consisting of the survey results of patients with hypertension in educational and practical centers of primary health care units based on primary cardiac function indicators, their average values, correlation coefficients, and algorithms of network clustering.

Neural network clustering has been used for the purpose of effective and objective allocation of patients to relevant categories in terms of certain indicators obtained during observation, which enables the determination of a combination of changes in certain parameters for prognosis of the disease towards deterioration or improvement.

The algorithm for decision-making has been created in order to optimize the prediction of diseases at the primary level, and for correct examination and treatment through the use of neural network clustering.

Generalizability

The suggested algorithm can be used to determine the risk of developing complications due to various diseases that are most common in a particular region. The suggested method is easy to use and does not require large financial investments, which is important in the organization of primary health care in rural areas.

Source of funding: This work was funded from the authors' own resources.

Conflicts of interest: The authors declare no conflicts of interest.

References

1. Epstein RM, Franks P, Fiscella K, et al. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. *Soc Sci Med* 2005; 61: 1516–1528.
2. Mintzer OP, Woronenko YV, Babintseva LY, et al. Health informatization concept in Ukraine. *Medical Informatics and Engineering* 2012; 3: 5–29.
3. Schiff GD. Minimizing diagnostic error: the importance of follow-up and feedback. *Am J Med* 2008; 21(5, Suppl.): S38–S42.
4. Wong D, Galleqos Y, Weinger M, et al. Changes in intensive care unit nurse task activity after installation of a third-generation intensive care unit information system. *Crit Care Med* 2003; 31(10): 2488–2494.
5. Perera C. The evolution of e-Health – mobile technology and mHealth. *JMTM* 2012; 1(1): 1–3, doi: <http://dx.doi.org/10.7309/jmtm.1>.
6. Vostrov GN, Mintser OP, Pavlov OO, et al. Information model of remote medical services. The first message. *Medical Informatics and Engineering* 2010; 3: 37–47.
7. Hartzband P, Groopman J. Off the record – avoiding the pitfalls of going electronic. *N Engl J Med* 2008; 358: 1656–1658.
8. Martsenyuk VP, Kravets NO. On the software development environment of intelligent medical databases. *Clinical Informatics and Telemedicine* 2004; 1: 47–53.
9. Florczak B, Scheurich A, Croghan J, et al. An observational study to assess an electronic point-of-care wound documentation and reporting system regarding user satisfaction and potential for improved care. *Ostomy Wound Manage* 2012; 58(3): 46–51.
10. Kovalchuk LA. Results of the latest techniques of the educational process in the I. Horbachevsky Ternopil State Medical University and future plans. *Med Educ* 2012; 2: 11–17.
11. Martsenyuk V. On an indirect method of exponential estimation for a neural network model with discretely distributed delays. *Electron J Qual Theo* 2017; 23: 1–16, doi: 10.14232/ejqtde.2017.1.23.
12. Martsenyuk V. Indirect method of exponential convergence estimation for neural network with discrete and distributed delays. *EJDE* 2017; 246: 1–12. Available from URL: <https://ejde.math.txstate.edu/Volumes/2017/246/martsenyuk.pdf>.

Tables: 0

Figures: 2

References: 12

Received: 02.12.2017

Reviewed: 05.12.2017

Accepted: 03.02.2018

Address for correspondence:

Prof. Petro Selskyy, MD, PhD, DSc

I. Horbachevsky Ternopil State Medical University

Maydan Voli 1

46000 Ternopil

Ukraine

Tel.: +38 0673985279, +38 0352 253998

E-mail: selskyy@tdmu.edu.ua