Development of the information system for HIV progression modeling in inmates

D S Ponomarev¹, S B Ponomarev¹, M M Gorohov¹, E L Averyanova² and V E Polishchuk¹

¹ Research Institute of the Federal Penitentiary Service of Russia, 125130, Moscow, Russia
² Pskov State University, 180000, Pskov, Russia

E-mail ponomarev.dmitry1990@mail.ru

Abstract. The study presents the stages of developing a HIV prognosis model in inmates. Correlation data analysis and regression modeling were used to develop the model. It was estimated that the developed model gives the best results for patients under the age of 40 years. On the obtained results there were developed an information system and a computer program. The study results application will help to predict correctly the HIV severity and to take timely treatment measures.

1. Introduction
The adequate therapy for HIV-infected inmates is of particular interest. Without any doubts, it can help to predict the HIV progression in this group of patients [1]. The study of HIV progression was based of the methods of system analysis and mathematical modeling [2]. The use of these methods will make it possible not only to assess the condition of patients more accurately (in numerical terms), but also to predict the course of the disease in the future. It should be said that the development of the information system makes it possible to single out the most important properties of the process and detach them from its insignificant characteristics [3]. Often, information systems and mathematical modeling allow to formulate new hypotheses and to gain new knowledge about the researched object those that was not available during the research. The development of models based on the formalization of links between its elements is of particular interest. This will allow not only to eliminate gaps in knowledge about the object, but also to identify new qualitative problems that initially could not be determined in a practical way. It should be noted that research in this area has already been carried out, and this fact confirms the relevance of the research [4,5]. In this case, the use of artificial neural networks is of particular interest [6,7]. However, conducting developments in this direction it is necessary to identify the relationship between the parameters and justify their choice. Many methods are known for modeling and diagnosing HIV stages. Nevertheless, there are problems remaining unresolved and leaving the urgency of the problem at a sufficiently high level. Based on the foregoing, it should be developed a method based on the most informative indicators that allows a simple, non-invasive and accurate risk prognosis of developing AIDS in a HIV-infected person.

To create an information system for assessing for HIV progression modeling in inmates, it is necessary, first of all, to assess the relationship between indicators of the patient's condition and the development of the disease itself.
The goal of the research is to study the possibility of using the methods of mathematical statistics and systems analysis to identify the relationship between the development of HIV and the basic medical indicators data.

2. Research materials and methods

Previous studies have identified a connection between the transition of HIV to AIDS with the stages of concomitant tuberculosis. But the proposed prognosis methods do not always take into account the whole range of indicators characterizing the condition of the patient. Today a lot of methods for identifying statistical relationships exists, but the most popular of named statistical method is correlation data analysis.

All analyzed parameters of patients were considered in point and numerical equivalents on the base of the previously conducted metrological study [8]. To carry out a correlation analysis and establish statistical links between the main medical parameters and the AIDS development parameter it was used a sample formed on the data of 2 years observation of HIV-carriers inmates in Russian penitentiary system.

There were selected main indicators and parameters that are monitored regularly and systematically among inmates with HIV (in total - 29 parameters). Each of these parameters was assigned the value X. For the parameter that reflects the rate of transition of HIV to the terminal stage - AIDS (AIDS development indicator), was assigned the value Y. The condition was set that the parameters X are considered as input signals, and Y as an output signal. 29 parameters X were analyzed.

The correlation coefficient was calculated using the formula (1) [9]:

\[
 r = \frac{\sum_{i=1}^{n} (x_n - \bar{x})(y_n - \bar{y})}{n\sigma_x\sigma_y},
\]

where: \(n\) – the number of observations; \(x\) – input variable; \(y\) – output variable; \(\sigma\) - the variance of the above parameters.

Based on the methods of mathematical statistics in clinical research, optimal partitioning methods for data analysis, in the course of the correlation analysis from analyzed 29 parameters (X) the following were noted as significant:

- determination of HIV progression stages in points (Stage 1 - 0 points, Stage 2 - 1 point, Stage 3 - 2 points, Stage 4 A-B - 3 points, Stage 4 B-B - 4 points) (X2);
- determination of the stage of tuberculosis in points (absence of tuberculosis - 0 points, tuberculosis in the anamnesis, metatuberculous changes in the lungs - 1 point, tuberculosis of thoracic lymph nodes - 2 points, focal tuberculosis - 3 points, infiltrative tuberculosis - 4 points, disseminated tuberculosis - 5 points, tuberculosis with pleurisy - 6 points, generalized tuberculosis - 7 points, tuberculosis recurrence - 8 points) (X3);
- body temperature (0 points – temperature from 36.6 °C to 37.0 °C, 1 point - temperature from 37.0 °C to 38.0 °C; 2 points - temperature above 38.0 °C (high temperature)) (X5);
- the presence or absence of chronic obstructive pulmonary disease (the presence of obstructive pulmonary disease - 1 point, the absence - 0 points) (X7);
- the measurement of liver size according to the method developed by M.G. Kurlov in centimeters (X9);
- the number of CD4, thousand cells / ml (X12).

To identify the risk of developing AIDS (Y), the following coefficients with moderate and strong correlation were obtained for the above parameters, which have: \(R_{X2}=0.49\); \(R_{X3}=0.50\); \(R_{X5}=0.52\); \(R_{X7}=0.58\); \(R_{X7}=0.47\); \(R_{X12}=0.49\).
3. Discussions

Regression analysis of the data was carried out on the basis of the revealed correlations of the parameters. As a result, the coefficients of the regression equation were established.

It should be said that the value of the outgoing parameter Y was divided into point equivalents. Limitations were set that Y less than 2 corresponds to a low risk development of HIV-infection to AIDS, with Y from 2 to 3 - the average risk development of HIV-infection to AIDS, Y more than 3 corresponds to a high risk development of HIV-infection to AIDS. It should be noted that the obtained model is especially effective for predicting the risk of developing AIDS only for young people aged from 18 to 40 years old. After the age of 40, the results become less accurate and sometimes erroneous.

Coefficients were obtained with the use of regression analysis, on its basis it was built a model, which helps to predict the risk of developing AIDS [9]. The developed method has been tested in practice.

The developed model can be implemented in practice as follows: determination of HIV progression stages in points (Stage 1 - 0 points, Stage 2 - 1 point, Stage 3 - 2 points, Stage 4 A-B - 3 points, Stage 4 B-B - 4 points); determination of the stage of tuberculosis in points (absence of tuberculosis - 0 points, tuberculosis in the anamnesis, metatuberculous changes in the lungs - 1 point, tuberculosis of thoracic lymph nodes - 2 points, focal tuberculosis - 3 points, infiltrative tuberculosis - 4 points, disseminated tuberculosis - 5 points, tuberculosis with pleurisy - 6 points, generalized tuberculosis - 7 points, tuberculosis recurrence - 8 points); determination of body temperature (0 points – temperature from 36.60°C to 37.00°C, 1 point - temperature from 37.00°C to 37.50°C; 2 points - temperature above 38.00°C (high temperature)); determination of liver size according to the method developed by M.G. Kurlov in centimeters; determination of the presence or absence of chronic obstructive pulmonary disease (the presence/the absence - 1 / 0 points); determination of the number of CD4, thousand cells / ml; then the formula predicts the outcome of the disease in points (2):

\[ Y = 1.34 + 0.25X_2 + 0.17X_3 + 0.11X_4 + 0.34X_5 + 0.09X_9 - 0.03 \cdot 10^{-2}X_{12} \]  

(2)

Y – prognosis of the risk of developing AIDS, points; \(X_2\) – HIV progression stage, points; \(X_3\) – the stage of tuberculosis, points; \(X_5\) – body temperature, points; \(X_9\) – chronic obstructive pulmonary disease, points; \(X_{12}\) – liver size according to the method developed by M.G. Kurlov, centimeters; \(X_{12}\) – CD4, thousand cells/ml.

The risk of developing AIDS is determined in points equivalent. At Y <2 points, a low risk of HIV-infection to AIDS is predicted, at Y from 2 to 3 - the average risk of HIV-infection to AIDS, at Y> 3 a high risk of HIV-infection to AIDS is predicted. The advantages of the method are high accuracy, ease of implementation, low cost and the ability to improve the quality of the assessment of the severity of the disease in inmates [9].

Further, on the basis of the obtained model, a program was developed [10], with which it can be predicted the possible course of the disease. It is assumed that the user will enter data in points equivalent for the parameters discussed above. Then, according to the above formula, which was obtained as a result of regression analysis of the data, the program calculates the point equivalent (in this case, parameter Y) and, accordingly, the development of the disease in general.

The results of the model and the work of the program were tested on real clinical cases. Despite the fact that the studies were conducted using data for inmate patients, the results obtained can be used for the health care system as a whole.

4. Results

The article proves the effectiveness and relevance of the use of systems analysis methods in the development of the information system for HIV progression modeling. The possibilities of applying the methods of statistical analysis and modeling, the development of an information system based on them are considered.

Correlation analysis was carried out and a regression model was developed, which makes it possible to predict the development of AIDS more accurately. It should be noted that the developed model is mainly applicable for persons under 40 years old.
Based on the above-mentioned regression model, a computer program was developed that allows a more convenient way to apply the research in practice. On the basis of the study was obtained invention patent [9] and a certificate for registration of a computer program was received [10].

It should be repeated that the obtained model is especially effective for predicting the risk of developing AIDS only for young people aged from 18 to 40 years old. After the age of 40, the results become less accurate and sometimes erroneous.

Reference

[1] Sizyakina L and Andreeva I 2005 Pat. of the Russian Federation 2251701 appl. 07.06.2004, publ. 10.05.2005.
[2] Denisenko V and Simovanyan A 2010 Pat. of the Russian Federation 2444299 appl. 09.03.2011 publ. 10.03.2012.
[3] Tariq R 2016 Make Your Own Neural Network (Kindle Edition) p 11-101.
[4] Ponomarev S and Averyanova E 2016 Pat. of the Russian Federation 2597805 appl. 16.06.2015 publ. 20.09.2016.
[5] Vostroknutov M, Dyuzheva E and Ponomarev S 2018 Risk indicators of lethal outcome in patients with comorbid HIV infection and tuberculosis held in penitentiary institutions Ural Medical Journal 8(163) 29-32.
[6] Muller A and Guido S 2016 Introduction to Machine Learning with Python: A Guide for Data Scientists (O'Reilly) p 21-115.
[7] Kartun-Giles A, Krioukov D, Gleeson J, Moreno Y and Bianconi G 2018 Sparse power-law network model for reliable statistical predictions based on sampled data Entropy 20(4) 156-210.
[8] Dyuzheva E, Ponomarev S, Vostroknutov M, Polishchuk V and Ponomarev D 2019 The development of the forecasting information system of HIV and tuberculosis coinfection’s outcome in inmates Journal of Physics: Conference Series 33023.
[9] Ponomarev D, Ponomarev S and Averyanova E 2020 Pat. of the Russian Federation 2725242 appl. 14.08.2019 publ. 30.06.2020.
[10] Ponomarev S, Ponomarev D and Averyanova E 2019 Pat. of the Russian Federation 2019662308 appl. 09.06.2019 publ. 20.09.2019.