**Statistical analysis of the Charity hospital database**

*I. Maslova1, I. Tsepeleva2, A. Bydanov3, P. Pchelintseva4, E.Tomilov5*

*1  I.M. Sechenov First Moscow State Medical University, 119991, Moscow, Russia*

*2D. Mendeleev University of Chemical Technology of Russia, Miusskaya Square, 9, 125047, Moscow, Russia*

*3 Center for Genetics and Life Science, Sirius University of Science and Technology, 354340, Sochi, Russia*

*4 InSysBio, Moscow, Russia*

*5 Russian Gerontology Clinical Research Center of Pirogov Russian National Research Medical University of Ministry of Healthcare of the Russian Federation, Moscow, Russia, 129226*

Charity hospital is a non-profit organization that provides medical and social assistance to homeless people in Saint Petersburg. Volunteer doctors consult patients, vaccinate them, provide them with glasses, etc. Since 2021, the REDCap electronic data capture system has been collecting information about all the Charity hospital patients.

The aim of this study was to analyze the patient database of the Charity hospital which contains information about 1633 unique patients and 4427 visits. We perform data polishing, descriptive statistics in order to form a typical portrait of homeless patients. Next we investigated correlation of variables with the place of reception, analyzed HIV patients, factors for ambulance call and main factors affecting the patients' life quality.

The analysis of universal portrait of homeless showed the following results: most people without a fixed place of residence who applied to ANO "Charity Hospital" are male, with Russian citizenship, have a basic set of documents (Passport, SNILS, compulsory medical insurance policy), have a permanent registration or do not have it at all. For the majority of patients, consent to the processing of personal data was not filled out at the first appointment. Most patients do not have socially significant diseases, allergies, but have bad habits (nicotine and alcohol addiction). Of the diseases, the most common are diseases of the skin and subcutaneous tissue, diseases of the musculoskeletal system and connective tissue, diseases of the respiratory system, diseases of the digestive system and diseases of the circulatory system. Patients are mostly unmarried. Approximately equally there were patients who were and were not in places of detention.

Data on the visual acuity of the 111 homeless patients were also analyzed. 74% had farsightedness. The most demanded glasses are 2.5 diopters.

To search for associations with HIV-positive status, the most populated variables were identified andt an exploratory regression model was built.

Based on the results of the analysis, the following conclusions could be drawn:

* If a patient is infected with Hepatitis C and B, then the likelihood of having an HIV infection increases.
* Every year the chance of contracting HIV decreases by 0.94 times compared to the previous year.

A similar analysis was carried out for Hepatitis C. It was found that every year the chance of infection with hepatitis C decreases by 0.96 times compared to the previous year. Gender in this case does not play a significant role (p-value>0.05). At the same time, the presence of HIV infection increases the chance by 3 times, and hepatitis B - by 10 times.

To search for indicators on which the call to the emergency medical services to the patient depends, Cramér’s V was used for nominal variables. Regression analysis was unavailable due to lack of data. To calculate the strength of association, we built a table, the values of which were then divided into three categories: weak, moderate and strong connection. No strong association between the variables was found. A moderate association was observed between hospitalization and the following variables: complaints, edema (anasarca), vomiting (uncontrollable), SpO2 saturation.

For further analysis, we identified several places of reception with the highest attendance, between which some differences were found. The place of reception that stands out most in terms of a number of indicators was designated as MS. Significant differences in systolic pressure were between the points of reception of MS and NP, diastolic - between MS and all other points of reception.

To conclude, the results described above can be useful for optimizing the work of the charity hospital.

All our results can be found in the GitHub repository:

<https://github.com/User00342/BioStat_2022/tree/main/%D0%9D%D0%BE%D0%B2%D0%BE%D0%B5>