

Abstract

N. A. Halushko,
T. O. Tretska,
A.V. Halushko,
*Sumy State University, 2,
Rymkogo-Korsakova st., 40007
Sumy, Ukraine*

FEATURES OF THE TRANSMISSION MECHANISM OF VIRAL
HEPATITIS C IN UKRAINE

Introduction/objective. The significant part of young people in the structure of hepatitis C virus (HC/HCV infection) incidence, a great deal of latent cases of this infection and the lack of specific prevention may complicate the epidemic situation regarding this infection in Ukraine in the coming years. The authors developed a mathematical model of the HC epidemiological process to determine the most significant factors in this infection transmission in the country.

Materials and methods. The study is based on correlation-regression analysis of the relationship between dependent (or responding) and explanatory (factorial or predictors) variables. In total, the analysis involved 3 dependent variables y_1, y_2, y_3 , corresponding to the annual number of acute and chronic HC cases and the number of HC virus seropositive individuals, and 17 predictors $x_1 - x_{17}$, including patients who received etiotropic treatment; patients with mental and behavioral disorders due to narcotics use, including opioids; patients with sexually transmitted infections; the number of visits to dentists; the number of patients who had dentures placed; the number of surgical operations, blood transfusions, endoscopic examinations, laboratory blood tests, hemodialysis, etc. The number of observations (n) of dependent and explanatory variables was equal to 25, which corresponds to the number of administrative-territorial units in Ukraine (24 regions and Kyiv).

The quality of regression models was evaluated using multiple correlation coefficients (R), determination coefficients (R^2) and regression coefficients (b_0, b_1, b_2). Statistical significance of R^2 was determined by F-statistics, regression coefficients – by standard errors (m), t -test, p -value, and the range of 95% confidence intervals (CI).

To compare the degree of influence of factor variables over dependent variables in the two-factor regression model, standardized regression coefficients were calculated.

The reliability of regression models was evaluated by the statistics of Durbin–Watson (DW), Breusch–Godfrey (BG) and White (W) tests.

The relative risk (RR) of HC infection was retrospectively determined in the individuals from behavioral and medical risk groups.

Results. In mathematical model of the epidemic process of acute HC, statistical significance was demonstrated for only one variable effect – annual number of dentist visits. The obtained regression equation was as follows:

$$y_1 = 0.000021 x_5 - 11.353,$$

where y_1 = annual number of patients with acute HC; x_5 = annual number of dentist visits. Statistical characteristics of the model: $R = 0.892$, $R^2 = 0.796$; F-test: 89.9 for 1 and 23 degrees of freedom, statistical significance for F: 0.0000000021; regression coefficients: $b_1 = 0.000021$ ($m = \pm 0.0000023$; $t = 9.48$, $t_{crit} = 1.71$; $p = 0.0000000021$; 95% CI [0.000017; 0.000026]), $b_0 = -11.353$ ($m = \pm 3.982$; $t = 2.85$, $t_{crit} = 1.71$; $p = 0.009$; 95% CI [-19.59; -3.116]).

When developing a model of the epidemic process of acute HC taking into account the annual number of seropositive individuals, statistical significance was demonstrated only for two variables: annual number of the sexually transmitted infections and annual number of laboratory blood tests. The analytical relationship of variables in this model had the following mathematical expression:

$$y_3 = 4.563 x_4 + 0.0058 x_{15} - 36552.721,$$

where y_3 = number of HCV-seropositive individuals; x_4 = number of sexually transmitted diseases, x_{15} = number of laboratory blood tests.

Statistical characteristics of the model: $R = 0.92$, $R^2 = 0.842$; F-test: 58.62 for 2 and 22 degrees of freedom, statistical significance for F: 0.00000000153; regression coefficients: $b_0 = -36552.721$ ($m = \pm 10649.1$; $t = 3.43$, $t_{crit} = 1.71$; $p = 0.0024$; 95% CI [-58637.63; -14467.81]), $b_1 = 0.0058$; $m = \pm 0.00082$; $t = 7.1$, $t_{crit} = 1.71$; $p = 0.0000004$; 95% CI [0.0041; 0.0075]; $b_2 = 4.563$; $m = \pm 1.526$; $t = 2.99$, $t_{crit} = 1.71$; $p = 0.0067$; 95% CI [1.4; 7.73].

The Durbin–Watson and Breusch–Godfrey tests did not reveal autocorrelation of residues for both regression models: $DW_U < DW_p < 4 - DW_U$; $BG < \chi^2$. White's test shows no heteroscedasticity for both models: $W < \chi^2$. The test results indicate the reliability of both regression models.

Conclusions. According to our data, at least 84% of HC virus infection cases in Ukraine occur through sexual contact and during laboratory blood sampling, and the role of the latter route of transmission in the HC virus spread was even more significant (standardized regression coefficients are 0.3 and 0.7, respectively).

Almost 80% of acute HC cases are associated with dental interventions.

Etiotropic treatment of patients with HC at the current level of treatment coverage can reduce the incidence of complications and the risk of death, but it is ineffective as a measure of influence on the first stage of the epidemiological process (source of infection).

Drug users have little effect on the intensity of the HC epidemiological process in Ukraine as a whole, despite the fact that the relative risk of HC among this population is quite significant (RR = 6.5; 95% CI [6.39; 6.63]).

Keywords: Hepatitis C, epidemic process, dental procedures, laboratory blood sampling, sexual transmission.

Corresponding author: galnatmed@gmail.com

Резюме

Н. А. Галушко,

Т. О. Трецька,

А. В. Галушко,

Сумський державний університет, вул. Римського-Корсакова, 2, м. Суми, Україна, 40007

ОСОБЛИВОСТІ МЕХАНІЗМУ ПЕРЕДАЧІ ВІРУСНОГО ГЕПАТИТУ С В УКРАЇНІ

Вступ/мета. Значна кількість осіб молодого віку в структурі вірусного гепатиту С (ГС/ HCV-інфекція), суттєві масштаби прихованих випадків цієї інфекції та відсутність специфічної профілактики можуть ускладнити епідемічну ситуацію щодо цієї інфекції в Україні в найближчі роки. Ми побудували математичну модель епідемічного процесу ГС для встановлення найбільш значущих факторів передачі цієї інфекції в країні.

Матеріали та методи. Дослідження засновано на проведенні кореляційно-регресійного аналізу взаємозв'язків між залежними (або результативними) та пояснюючими змінними (або факторними, або предикторами). Всього в аналіз включено 3 залежних змінних y_1, y_2, y_3 , що відповідають щорічній кількості випадків захворювань на гострий, хронічний ГС і кількості серопозитивних до вірусу ГС осіб, та 17 предикторів $x_1 - x_{17}$, серед яких кількість осіб, які отримали етіотропне лікування, кількість осіб з розладами психіки та поведінки через вживання наркотичних речовин, включаючи опіоїди, кількість хворих на інфекції, що передаються статевим шляхом, кількість відвідувань стоматологів, число осіб, що отримали зубні протези; кількість хірургічних операцій, гемотрансфузій, ендоскопічних обстежень, лабораторних аналізів крові, гемодіалізів тощо. Кількість спостережень (n) залежних і пояснюючих змінних дорівнює 25, що відповідає кількості адміністративно-територіальних одиниць в Україні (24 області та м. Київ).

Якість регресійних моделей оцінювали за допомогою множинних коефіцієнтів кореляції (R), коефіцієнтів детермінації (R^2) та коефіцієнтів регресії (b_0, b_1, b_2). Статистичну значущість R^2 визначали за F-статистикою, коефіцієнтів регресії – за стандартними помилками (m), t -критерієм, p -значенням і діапазоном 95% довірчих інтервалів (ДІ).

Для порівняння ступеню впливу факторних змінних на залежну змінну у двохфакторній регресійній моделі розраховували стандартизовані коефіцієнти регресії ($b_{ст}$).

Надійність регресійних моделей оцінювали за статистикою тестів Дарбіна–Уотсона (DW), Бреуша–Годфрі (BG) і Уайта (W).

Ретроспективно визначали величину відносного ризику (RR) інфікування ГС осіб з груп поведінкового та медичного ризику.

Результати. При математичному моделюванні епідемічного процесу гострого ГС доведена статистична значущість ефекту лише однієї змінної – річної кількості відвідувань стоматологів. Отримане рівняння регресії має наступний вигляд:

$$y_1 = 0,000021 x_5 - 11,353,$$

де y_1 – щорічна кількість людей з гострим ГС; x_5 – річна кількість відвідувань стоматологів. Статистична характеристика моделі: $R = 0,892$; $R^2 = 0,796$; F-статистика: 89,9 для 1 і 23 ступенів свободи, статистична значущість F: 0,0000000021; коефіцієнти регресії: $b_1 = 0,000021$ ($m = \pm 0,0000023$; $t = 9,48$, $t_{крит} = 1,71$; $p = 0,0000000021$; 95% ДІ [0,000017; 0,000026]), $b_0 = -11,353$ ($m = \pm 3,982$; $t = 2,85$, $t_{крит} = 1,71$; $p = 0,009$; 95% ДІ [-19,59; -3,116]).

Моделювання епідемічного процесу хронічного ГС (тобто залежності змінної y_2 від усіх пояснюючих змінних) не виявило статис-

тично значущих коефіцієнтів регресії, що може бути пов'язане з неповною реєстрацією випадків хронічного ГС.

При моделюванні епідемічного процесу HCV-інфекції з врахуванням щорічної кількості серопозитивних осіб встановлена статистична значущість двох предикторів – річної кількості лабораторних досліджень крові та річної кількості інфекційних захворювань, що реалізуються статевим шляхом. Аналітичний зв'язок змінних у даній моделі має наступний математичний вираз:

$$y_3 = 4,563 x_4 + 0,0058 x_{15} - 36552,721,$$

де y_3 – кількість HCV-серопозитивних осіб; x_4 – кількість захворювань, що передаються статевим шляхом, x_{15} – кількість лабораторних аналізів крові.

Статистична характеристика моделі: $R = 0,92$; $R^2 = 0,842$; F-статистика: 58,62 для 2 і 22 ступенів свободи, статистична значущість F: 0,0000000153; коефіцієнти регресії: $b_0 = -36552,721$ ($m = \pm 10649,1$; $t = 3,43$, $t_{крит} = 1,71$; $p = 0,0024$; 95% ДІ [-58637,63; -14467,81]); $b_1 = 0,0058$ ($m = \pm 0,00082$; $t = 7,1$, $t_{крит} = 1,71$; $p = 0,0000004$; 95% ДІ [0,0041; 0,0075]); $b_2 = 4,563$ ($m = \pm 1,526$; $t = 2,99$, $t_{крит} = 1,71$; $p = 0,0067$; 95% ДІ [1,4; 7,73]); $b_{ст}$ для змінних x_4 та x_{15} становлять відповідно 0,3 та 0,7.

У тестах Дарбіна-Уотсона та Бреуша-Годфрі не виявлено автокореляції залишків обох регресійних моделей: $DW_U < DW_p < 4 - DW_U$; $BG < \chi^2$. Тест Уайта свідчить про відсутність гетероскедастичності обох моделей: $W < \chi^2$. Результати тестів свідчать про надійність обох регресійних моделей.

Висновки. За нашими даними, щонайменше 84% випадків зараження вірусом ГС в Україні відбувається через статеві контакти та під час лабораторного забору крові, причому вплив останнього шляху передачі на поширення вірусу ГС є більш вираженим (стандартизовані коефіцієнти регресії становлять відповідно 0,3 та 0,7).

Майже 80% випадків гострого ГС пов'язано з проведенням стоматологічних втручань.

Етіотропне лікування хворих на ГС при існуючому рівні охоплення лікуванням може знизити частоту ускладнень і ризик смерті, однак як захід впливу на першу ланку епідпроцесу (джерело інфекції) є малоефективним.

Особи, що вживають наркотики, мало впливають на інтенсивність епідпроцесу ГС в Україні у цілому, незважаючи на те, що відносний ризик захворювань на ГС серед цієї групи населення є досить високим ($RR = 6,5$; 95% ДІ [6,39; 6,63]).

Ключові слова: Гепатит С, епідемічний процес, стоматологічні процедури, лабораторний забір крові, статевий шлях передачі.

Автор, відповідальний за листування: galnatmed@gmail.com

Introduction

Viral hepatitis C (HC/HCV-infection) is one of the most burning problems in the world and in Ukraine due to the wide range of medical and social-economic consequences of its morbidity: severe course with chronic forms of the disease, cirrhosis and liver cancer, and significant cost of

diagnosis and treatment [1, 2, 3]. The complexity of combating this infection is determined by the lack of specific prevention methods, poor availability of etiotropic treatment, high chronogenic potential of the virus and huge latent epidemiological process. In these conditions, the proportional prevalence of young people in HC incidence among the population of Ukraine [4], as well as insufficient

etiotropic treatment coverage of patients with HC [5] may complicate the epidemiological situation in the coming years.

The epidemiological process of any infectious disease in the human population develops involving certain social factors. Determination of certain elements of social life that specifically influence the infectious disease spread is a prerequisite for successful management of the course of its epidemic process. These ideas determined the objective of our study.

The objective was to define the relative risk of HC in population groups with certain behavioural and medical risks; to develop a mathematical model of HC epidemic process in Ukraine in order to find out its most important determinants.

Materials and methods. The authors of the study used data from the analytical report of the Public Health Center on the incidence of acute and chronic HC in Ukraine, on seroprevalence of HC virus, on the number of patients with HC who received etiotropic treatment [4], and the data from open database of the Medical Statistics Center of the Ministry of Health of Ukraine on performance of health care institutions in different regions of Ukraine [6] for 2013–2018.

Retrospectively, according to serological studies conducted in Ukraine for prophylactic purposes, the relative risk of infection with HC pathogen was determined for different groups of population. Relative risk of infection (RR) is the ratio of the number of HC virus seropositive individuals in an exposed group to the number of seropositive individuals in an unexposed group. 95% confidence interval (CI) was calculated for RR. The relative risk of HC infection was determined in the groups of behavioural risk (narcotics users and patients with sexually transmitted diseases) and medical risk (health workers and long-term inpatients), and in comparison groups (pregnant women and donors).

In order to create a mathematical model of HC epidemic process the authors used correlation and regression analysis between dependent (or responding) and explanatory (or factorial feature, or predictor) variables that were logically related. The number of observations (n) of dependent and explanatory variables was equal to 25, which corresponds to the number of administrative-territorial units in Ukraine (24 regions and Kyiv). All parameters had absolute dimension. The groups of parameters with names, designations, as well as arithmetic mean values (M) and standard deviations (σ) are given in Table 1.

Statistical data were processed using MS Office Excel software package.

Study procedure:

1. The long-term average annual values for dependent and explanatory features (variables) were calculated for every administrative-territorial unit of Ukraine for 2013–2018.

2. The distribution of random variables (dependent and factorial features) was tested for normality. To do this, the coefficients of skewness (A) and kurtosis (E) were calculated using the functions $SKEW(\text{number1};[\text{number2}];\dots)$ and $KURT(\text{number1};[\text{number2}];\dots)$ of MS Office Excel software package, and also the variances for these coefficients $D(A)$, $D(E)$ were calculated. If the modulo coefficients of skewness and kurtosis did not exceed respectively 3 and 5 square roots of their variances, i.e. inequalities $|A| \leq 3\sqrt{D(A)}$ та $|E| \leq 5\sqrt{D(E)}$ were satisfied, then the distribution of random variables was supposed to be close to normal.

3. The matrix of Pearson correlation coefficients (r_{xy}) between dependent and explanatory features was obtained. The correlation strength and statistical significance of the correlation coefficients were determined. Possible grouping variants were evaluated for features of correlation-regression models.

The strength of the correlation between the responding and factorial features was evaluated according to the Chaddock scale: with the absolute value of r_{xy} up to 0.3, the strength of the correlation was considered weak, 0.3 to 0.5 – it was considered moderate, 0.5 to 0.7 – it was considered noticeable, 0.7 to 0.9 – it was considered high, > 0.9 – it was considered very high.

The hypothesis of the significance of the Pearson correlation coefficient (r_{xy}) for the n -sample was determined using the Student's t-test (t), which was calculated by the formula:

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

According to the table of critical values of the Student's distribution (bilateral critical region), when the level of significance $\alpha = 0.05$ and the number of degrees of freedom $f = n - 2$, the corresponding critical value of the Student's coefficient (t_{crit}) was found. When $t > t_{crit}$, the hypothesis $H_0: r_{xy} = 0$ was rejected and an alternative hypothesis was chosen, on the basis of which a conclusion was made about the significance of the correlation coefficient.

Table 1 – Characteristics of indicators that were included in the correlation-regression analysis (average annual values per an administrative-territorial unit)

| Groups and parameters | Corresp. variable | The arithmetic mean of the indicator and standard deviation ($M \pm \sigma$), $n = 25$ |
|--|-------------------|--|
| 1. Dependent (responding) features: | | |
| Number of acute HC cases | y_1 | 22 ± 19 |
| Number of chronic HC cases | y_2 | 232 ± 150 |
| Number of HC virus seropositive individuals | y_3 | 68088 ± 49919 |
| 2. Factorial (explanatory) features | | |
| <i>2.1. Factors influencing the first stage of the HC epidemiological process</i> | | |
| Patients having received etiotropic treatment | x_1 | 155 ± 98 |
| <i>2.2. Factors influencing the second stage of the HC epidemiological process</i> | | |
| <i>a) predictors of HC virus transmission due to narcotics use:</i> | | |
| Patients with mental and behavioral disorders due to narcotics use | x_2 | 2407 ± 2139 |
| Patients with mental and behavioral disorders due to opioid use | x_3 | 1681 ± 1453 |
| <i>b) predictor of HC sexual transmission:</i> | | |
| Patients with syphilis, gonococcal infection, sexually transmitted chlamydial infection, trichomoniasis and urogenital mycoplasmosis | x_4 | 4836 ± 3305 |
| <i>c) predictors of nosocomial route of HC transmission:</i> | | |
| Number of dentist visits | x_5 | 1585635 ± 798265 |
| Patients having had planned sanation or sanation on request | x_6 | 398635 ± 215572 |
| The number of teeth treated for caries in the reporting year | x_7 | 3548914 ± 1625789 |
| Number of patients who received dentures | x_8 | 55028 ± 26018 |
| Number of surgeries performed in outpatient polyclinics | x_9 | 84099 ± 36061 |
| Number of inpatient days | x_{10} | 4002184 ± 1656540 |
| Number of transfusions of blood or blood components | x_{11} | 629124 ± 400326 |
| Number of inpatient surgeries | x_{12} | 84208 ± 40158 |
| Number of endoscopic instruments | x_{13} | 143 ± 65 |
| Number of endoscopic examinations | x_{14} | 66245 ± 31182 |
| Number of hematological and biochemical lab tests | x_{15} | 14181122 ± 6151547 |
| Number of hemodialysis sessions performed | x_{16} | 29273 ± 14790 |
| Number of hemodialysis machines | x_{17} | 45 ± 29 |

95% confidence interval (CI) was calculated for correlation coefficient. The calculations were based on the Fisher z-transformation. The lower (z_L) and upper (z_U) limits of the transformed 95% confidence interval for the Pearson correlation coefficient are:

$$z_L = 0.5 \ln \left(\frac{1+r_{xy}}{1-r_{xy}} \right) - \frac{1.96}{\sqrt{n-3}} \text{ and } z_U = 0.5 \ln \left(\frac{1+r_{xy}}{1-r_{xy}} \right) + \frac{1.96}{\sqrt{n-3}},$$

where ln = natural logarithm, n = sample size.

The value of the correlation coefficient for the general population calculated from the sample in 95% of cases will be in the range:

$$\text{from } \frac{e^{2z_L-1}}{e^{2z_L+1}} \text{ to } \frac{e^{2z_U-1}}{e^{2z_U+1}},$$

where e = Euler number ($e \approx 2.7$).

The correlation coefficient was considered statistically significant, if its confidence interval did not include 0.

4. Using a graph of the empirical regression line, the form of the relationship was determined between responding and factorial features.

5. The authors solved the regression equations and evaluated the coefficients of multiple correlation R (that reflect the degree of dependence of a responding feature y on all factor variables x), the coefficients of determination R^2 (that show the proportion of the responding feature variance attributable to independent variables), and regression coefficients (that show what variation y is accounted for by a unit of variation x). It was supposed that the closer the coefficient of determination to 1, the greater variability of all variables is accounted for by the model, and, consequently, the better is the quality.

The significance of the coefficient of determination was found using Fisher's F-test. The critical point was found for significance level $\alpha = 0.05$ and the values of the degrees of freedom $f_1 = k$, $f_2 = n - k - 1$, where n = number of observations, k = number of factorial criteria. Provided that $F > F_{crit}$, the null hypothesis $H_0: R^2 = 0$ was rejected and the coefficient of determination was considered statistically significant ($R^2 \neq 0$). With $R^2 > 0.7$ the variation of responding feature y was considered to be influenced, generally, by factorial features included in the model.

Regression coefficients were not considered statistically significant if: the confidence intervals included both positive and negative values; predicted values of the t -test were less than the critical value, and standard errors exceeded half the values of the parameters.

6. The theoretically predicted values of the responding feature were analyzed. Mean error of approximation E (average deviation of the predicted values from the actual values) was evaluated:

$$E = \frac{\sum |y - \hat{y}|}{n} \times 100\% = \frac{\sum |\varepsilon|}{n} \times 100\%,$$

where \hat{y} = the value of the responding indicator (theoretical value, calculated according to the regression equation by substituting corresponding actual values of the factors); ε = error that is present in the model due to the fact that the responding indicator is also affected by other factors not taken into account in the regression equation.

With approximation error of up to 10%, the precision of the regression model selection was considered high, 10 to 20% – it was considered good, 20 to 50% – satisfactory, more than 50% – unsatisfactory.

7. Comparative analysis of variances (total variance, factor variance, residual variance) was performed and the strength of relationship between the features included in the model was evaluated.

8. Exclusion of insignificant and inclusion of additional factors were carried out followed by sections 1–6. At this stage, the algorithm of sequential selection was used (Stepwise) [7]: at each step after inclusion of a new variable in the model, significance of other variables entered earlier was tested. If the significance was not confirmed, then such variables were removed from the model. After adjusting the list of variables included in the model, another iteration of the procedure for finding a new variable was performed, which satisfied the conditions of its inclusion in the model.

9. The statistical significance of regression parameters was evaluated.

10. The reliability of the regression equation was analyzed.

Durbin–Watson test and Breusch–Godfrey test were used to detect possible distortion of standard errors and t -statistics of regression due to autocorrelation between the levels of the studied variable.

Darbin–Watson test evaluation was limited to testing of the hypothesis H_0 of absence of autocorrelation, when the autocorrelation coefficient ρ was equal to 0 ($\rho = 0$). In case of rejection, the alternative hypothesis $H_1: \rho > 0$ or $\rho < 0$ was accepted. The Darbin–Watson test was performed as follows:

$$DW_p = \frac{\sum (e_n - e_{n-1})^2}{\sum e_n^2},$$

where DW_p = prediction Durbin–Watson test, $(e_n - e_{n-1})$ = lag size (bias between runs of residuals of the regression model).

The critical lower (DW_L) and upper (DW_U) values of the Durbin–Watson test were found from the corresponding statistical tables at a significance level of $\alpha = 0.05$ provided the number of observations n and the number of factor variables k . If $DW_L < DW_p < DW_U$, H_0 was accepted. Otherwise, H_0 was rejected and H_1 was accepted.

The Breusch–Godfrey test was also used to test the null hypothesis of absence of serial correlation. If this hypothesis proved to be true, then the distribution for the criterion $BG = (n - f)R_{aux}^2$ (where R_{aux}^2 = coefficient of determination of the auxiliary model, f = autoregression order) for n -number of observations

was close to the chi-square distribution (χ^2) with f -degrees of freedom. The null hypothesis H_0 was rejected, if the predicted value $(n - f)R_{aux}^2$ exceeded the critical value χ^2 at the given level of significance α . In fact, a certain number of observations was omitted from the sample after developing Breusch–Godfrey auxiliary regression model with lag variables and was afterwards subtracted from the original sample size.

The heterogeneity of observations expressed in unequal (non-constant) variance of the random error of the regression model (heteroscedasticity) was supposed to lead to the inadequacy of the obtained statistical conclusions. To test the heteroscedasticity of random deviations, the White test was used to develop the dependence of the square of random deviations on all exogenous variables, their squares, and cross products. To test the null hypothesis H_0 (the presence of homoscedasticity of random deviations) the values of the White test statistics were used:

$$W = nR_{aux}^2,$$

where n = number of observations, R_{aux}^2 = coefficient of determination of the auxiliary model.

According to the table of critical values of the χ^2 -distribution and with regard to the number of degrees of freedom ($f = k - 1$, where k = number of regressors in the auxiliary regression model) the critical point was found. If the White test value (W) exceeded the critical value of χ^2 at a given level of significance α , then the null hypothesis was rejected

in favor of the alternative hypothesis of the presence of heteroscedasticity of random deviations, and vice versa.

To compare the degree of influence of factor variables x over dependent variables (in the two-factor linear regression equation), standardized regression coefficients were calculated. The standardized regression coefficient (b_{st}) shows which part of the standard deviation σ_y of the dependent variable y will change with the change of the corresponding factor x by the value of its standard deviation σ_x under the constant influence of another factor included in the equation. The standardized regression coefficient (b_{st}) was calculated by multiplying the regression coefficient b_x by the standard deviation of the variable x (σ_x) and dividing the result by the standard deviation of the variable y (σ_y). Standard deviations of factor and dependent variables are presented in Table 1.

Results. According to serological studies conducted in Ukraine, the relative risk of HC virus infection (Table 2) is highest for people who use drugs and for children born to HCV-positive mothers: respectively 6.5 and 5.4 times higher than the risk for the rest of population. The risk of HC infection in long-term inpatients and patients with sexually transmitted diseases is more than 2 times higher than that in the rest of population. At the same time, the risk of infection in healthcare workers, pregnant women and donors is 1.5, 2, and 6.8 times lower than that in the rest of population.

Table 2 – Characteristics of the relative risks (RR) of HC virus infection in different population groups (according to the determined serological markers of infection) including 95% confidence intervals

| Population groups | RR | 95% CI |
|---|-------|--------------|
| Narcotics users | 6.5 | 6.39; 6.63 |
| Long-term inpatients | 2.35 | 2.3; 2.38 |
| Patients with sexually transmitted diseases | 2.2 | 2.16; 2.3 |
| Orphan children | 1.4 | 1.33; 1.52 |
| Healthcare workers | 0.67 | 0.68; 0.7 |
| Pregnant women | 0.5 | 0.49; 0.51 |
| Donors | 0.147 | 0.145; 0.149 |
| Children born from HCV-positive mothers | 5.4 | 5.12; 5.71 |

The rationale for correlation-regression analysis to be used to study the relationship between responding and factorial features was determined through the coefficients of skewness (A), kurtosis (E) and variances of these coefficients D(A) and D(E). The obtained calculations showed that the distribution of the studied random variables in the

studied statistical series is close to normal distribution (Table 3).

Analysis of the correlation coefficients between responding and factorial variables (Table 3) helped to identify a number of important dependencies required for understanding the epidemiological process of HC:

1. The lack of correlation between the number of patients having received etiotropic treatment and the number of patients with acute HC, between the number of patients having received etiotropic treatment and the number of HC virus seropositive individuals, as well as a direct statistically significant correlation of noticeable relationship strength between the number of patients having received etiotropic treatment and the number of patients with chronic HC (r_{xy} is equal relatively to: 0.261, 95% CI [-0.149; 0.594]; 0.224, 95% CI [-0.187; 0.568] and

0.677, 95% CI [0.37; 0.84]). All three parameters indicate insufficient level of coverage of HC patients etiotropic treatment in order to effectively influence the epidemiological process of HC. The statistically significant correlation of noticeable relationship strength between the number of patients having received etiotropic treatment and the number of patients with chronic HC indicates only that the distribution of pharmaceuticals for free etiotropic treatment among regions of Ukraine is proportional to the number of patients with chronic HC.

Table 3 – Characteristics of random variables distribution in the statistical series of acute and chronic HC incidence and HC virus seroprevalence according to the coefficients of skewness (A), kurtosis (E), and corresponding variances (D)

| Parameters | A | E | $3\sqrt{D(A)}$ | $5\sqrt{D(E)}$ |
|---|------|------|----------------|----------------|
| Number of acute HC cases | 1.25 | 2.87 | 1.39 | 3.96 |
| Number of chronic HC cases | 1.05 | 0.63 | | |
| Number of HC virus seropositive individuals | 1.35 | 1.17 | | |

2. Statistically insignificant influence of narcotics use in general, including opioids, over the mechanism of HC transmission in Ukraine (r_{xy} is equal relatively to: 0.354, 95% CI [-0.047; 0.657] and 0.341, 95% CI [-0.062; 0.648])

3. Noticeable direct correlation between the number of transfusions of blood and blood components with the number of patients with acute HC and the number of HC virus seropositive individuals (r_{xy} is equal relatively to: 0.548, 95% CI [0.196; 0.755] and 0.574, 95% CI [0.232; 0.789]).

4. Noticeable direct correlation between the number of HC virus seropositive individuals and the number of patients with sexually transmitted diseases ($r_{xy} = 0.693$, 95% CI [0.411; 0.854]).

6. Noticeable and moderate direct correlations between HC incidence and indicators that reflect the therapeutic and diagnostic work in hospitals (Table 4).

In the course of successive model modification by exclusion of insignificant and inclusion of significant factors, only the indicators that describe HC epidemiological process in Ukraine best of all were kept and thus the following regression equations were obtained.

1. Regression equation for acute HC epidemiological process

$$y_1 = 0.000021 x_5 - 11.353$$

where y_1 = annual number of patients with acute HC; x_5 = annual number of dentist visits.

All regression coefficients were statistically significant, which was shown by the following data:

$$b_0 = -11.353; m = \pm 3.982; t = 2.85, t_{crit} = 1.71; p = 0.009; 95\% \text{ CI } [-19.59; -3.116];$$

$$b_1 = 0.000021; m = \pm 0.0000023; t = 9.48, t_{crit} = 1.71; p = 0.0000000021; 95\% \text{ CI: } 0.000017; 0.000026.$$

There was statistically strong relationship between dependent and factorial variables: R was equal to 0.892. The value of the coefficient of determination R^2 was 0.796. Thus, the regression equation was accountable for 79.6% of the responding feature variance (incidence of acute HC), and its residual variance was 20.4%, which indicated a satisfactory approximation and adequacy of the model. The results of F-test ($F = 89.9$, $F_{crit} = 0.0000000021$ at degrees of freedom $f_1 = 1$ and $f_2 = 23$) also confirmed the statistical significance of the influence of the factorial variable x (the number of dentist visits) over the number of acute HC cases.

The value of the approximation error E (44.9%) did not exceed 50%, and this characterized the model as satisfactory, which was clearly shown in Fig. 1. Interestingly, by excluding only two regions from the regression analysis – Volyn and Ivano-Frankivsk oblasts, where the predicted incidence of acute HC was, respectively, 1.67 and 3.38 times higher than the actual values of this indicator (apparently due to significant underdiagnosis of acute HC in these areas), the regression coefficient b_0 decreased to -9.8, and the approximation error decreased to 29.8%.

Table 4 – Correlation coefficients between dependent (responding) and explanatory (factorial) indicators and corresponding 95% confidence intervals

| Factorial features | Responding features | | | | | |
|--|---------------------|------------------|------------|------------------|-------------------------|------------------|
| | Acute HC | | Chronic HC | | HC virus seroprevalence | |
| | r_{xy} | CI | r_{xy} | CI | r_{xy} | CI |
| Patients having received etiotropic treatment | 0.261* | -0.149; 0.594 | 0.677 | 0.37; 0.84 | 0.224* | -0.187; 0.568 |
| Patients with mental and behavioral disorders due to narcotics use | 0.167* | -0.244; 0.527 | 0.312* | -0.094; 0.629 | 0.354* | -0.047; 0.657 |
| Patients with mental and behavioral disorders due to opioid use | 0.152* | -0.258; 0.516 | 0.264* | -0.146; 0.596 | 0.341* | -0.062; 0.648 |
| Patients with syphilis, gonococcal infection, sexually transmitted chlamydial infection, trichomoniasis and urogenital mycoplasmosis | 0.467 | 0.089; 0.727 | 0.329* | -0.076; 0.64 | 0.693 | 0.411; 0.854 |
| Number of dentist visits | 0.892 | 0.768; 0.951 | 0.659 | 0.357; 0.836 | 0.832 | 0.651; 0.923 |
| Patients having had planned sanation or sanation on request | 0.796 | 0.585; 0.906 | 0.525 | 0.164; 0.762 | 0.751 | 0.505; 0.883 |
| The number of teeth treated for caries in the reporting year | 0.755 | 0.513; 0.885 | 0.504 | 0.136; 0.749 | 0.747 | 0.5; 0.881 |
| Number of patients who received dentures | 0.749 | 0.503; 0.882 | 0.530 | 0.171; 0.764 | 0.732 | 0.474; 0.874 |
| Number of surgeries performed in outpatient polyclinics | 0.788 | 0.571; 0.902 | 0.468 | 0.09; 0.728 | 0.782 | 0.56; 0.899 |
| Number of inpatient days | 0.817 | 0.623; 0.916 | 0.636 | 0.322; 0.824 | 0.878 | 0.74; 0.945 |
| Number of transfusions of blood or blood components | 0.548 | 0.196; 0.755 | 0.261* | -0.158; 0.588 | 0.574 | 0.232; 0.789 |
| Number of inpatient surgeries | 0.856 | 0.697; 0.934 | 0.622 | 0.318; 0.822 | 0.848 | 0.682; 0.946 |
| Number of endoscopic instruments | 0.819 | 0.627; 0.917 | 0.604 | 0.275; 0.806 | 0.831 | 0.649; 0.923 |
| Number of endoscopic examinations | 0.661 | 0.36; 0.837 | 0.504 | 0.136; 0.749 | 0.660 | 0.359; 0.836 |
| Number of hematological and biochemical lab tests | 0.817 | 0.623; 0.916 | 0.602 | 0.272; 0.805 | 0.882 | 0.748; 0.947 |
| Number of hemodialysis sessions performed | 0.433 | 0.043; 0.707 | 0.358* | -0.043; 0.659 | 0.445 | 0.061; 0.714 |
| Number of hemodialysis machines | 0.218* | -0.193; 0.564 | 0.041* | -0.359; 0.429 | 0.523 | 0.162; 0.760 |

* Parameters that have no statistical significance

Durbin–Watson test: $DW_p=1.963$, $DW_{L(\alpha=0.05, n=25, k=1)}=1.206$, $DW_U(\alpha=0.05, n=25, k=1)=1.55$; $DW_U < DW_p < 4 - DW_U$. The calculation results indicated the absence of autocorrelation between the levels of the studied variable, which confirmed the high quality of the model.

Breusch–Godfrey test statistics also indicated the insignificance of the autocorrelation for the

model and therefore, its poor quality: $R_{aux} = 0.3279$; $BG = 2.58$; $\chi^2_{(\alpha=0.05; f=1)} = 3.84$; $BG < \chi^2$.

White test statistics show the homoscedasticity of the model: $R_{aux} = 0.148$; $W = 0.548$; $\chi^2_{(\alpha=0.05; f=2)} = 5.99$; $W < \chi^2$.

2. Regression equation for acute HC epidemiological process with regard to seroprevalence:

$$y_3 = 4.563 x_4 + 0.0058 x_{15} - 36552.721,$$

where y_3 – number of HC virus seropositive individuals; x_4 – number of patients with sexually transmitted diseases; x_{15} – number of lab blood tests.

The multiple correlation coefficient R for this model was equal to 0.92, which showed the linear nature of the relationship between

responding and factorial variables. The coefficient of determination R^2 was 0.842, therefore, the regression equation was accountable for 84.2% of the responding feature variance, and the part of the residual variance of the dependent variable y was 15.8%, which indicated a satisfactory approximation and adequacy of the regression model.

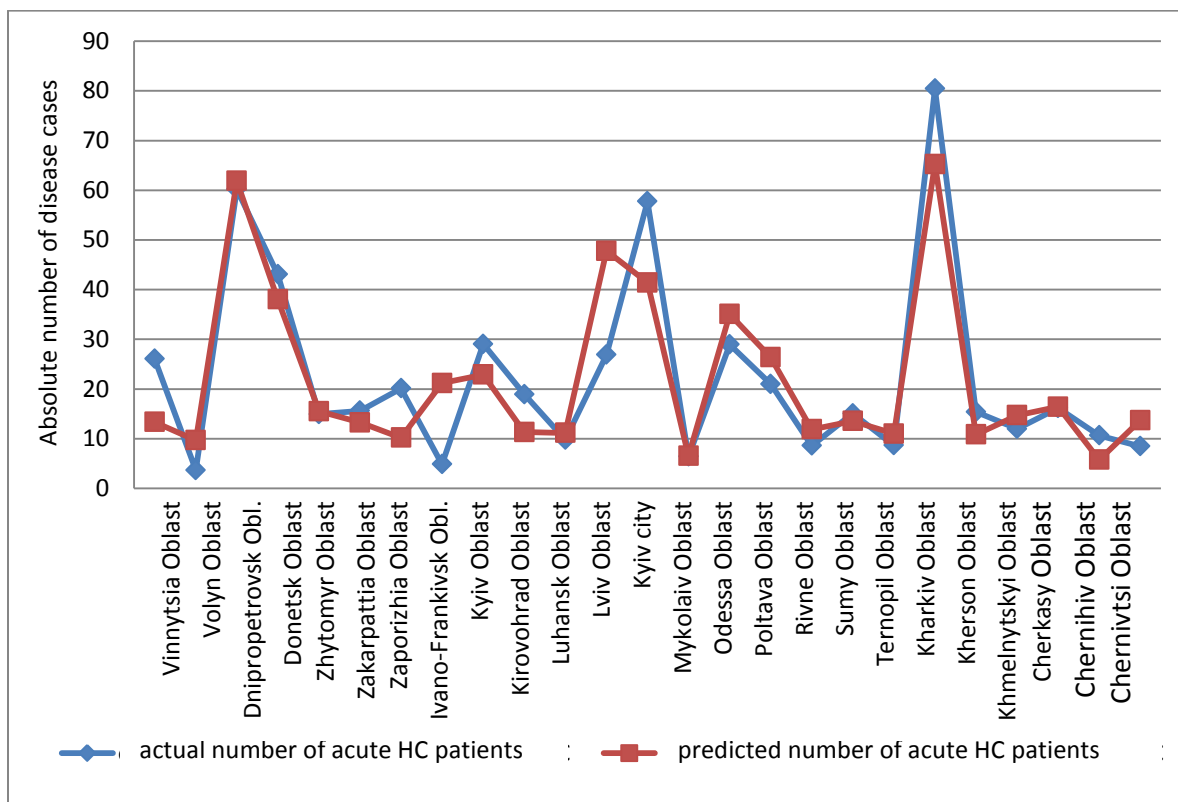


Figure 1 – Actual and simulated (acc. to the number of dentist visits) incidence of acute HC in different administrative-territorial units of Ukraine

All regression equation coefficients were statistically significant, which was shown by the following data:

$$b_0 = -36552.721; m = \pm 10649.1; t = 3.43, t_{crit} = 1.71; p = 0.0024; 95\% \text{ CI } [-58637.63; -14467.81];$$

$$b_1 = 0.0058; m = \pm 0.00082; t = 7.1, t_{crit} = 1.71; p = 0.0000004; 95\% \text{ CI } [0.0041; 0.0075];$$

$$b_2 = 4.563; m = \pm 1.526; t = 2.99, t_{crit} = 1.71; p = 0.0067; 95\% \text{ CI } [1.4; 7.73].$$

The statistical significance of the coefficient of determination was also confirmed by Fisher's test: $F(58.62)$ significantly exceeded $F_{crit}(0.0000000153)$ with $f_1 = 2$ and $f_2 = 22$, and this served to reject the null hypothesis of no influence of factors x_1 and x_2 on the responding variable and an alternative hypothesis that confirms the

statistical reliability of regression equation estimation was accepted.

The predicted value of the approximation error E for this model was equal to 26.9% and indicated its satisfactory adequacy, which was shown in Fig. 2.

Durbin–Watson test: $DW_p = 2.52, DW_{L(a=0.05, n=25, k=2)} = 1.288, DW_{U(a=0.05, n=25, k=2)} = 1.454; DW_L < DW_p < DW_U < 4 - DW_U$. There was no autocorrelation, which confirmed the high quality of the model.

Breusch–Godfrey test statistics also indicates the insignificance of the autocorrelation for the model and therefore, its poor quality: $R_{aux} = 0.3279; BG = 2.472; \chi^2_{(a=0.05; f=2)} = 5.99; BG < \chi^2$;

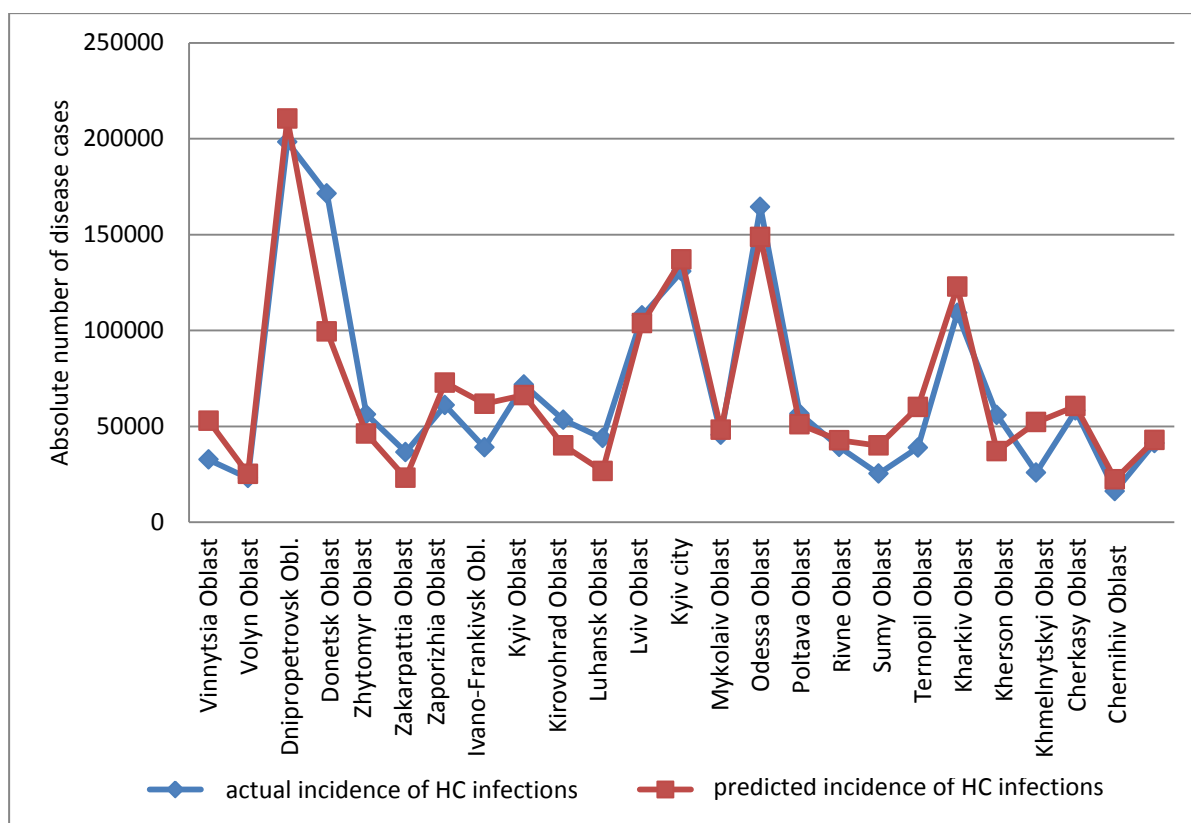


Figure 2 – Actual and simulated (acc. to the number of patients with sexually transmitted diseases and number of hematological and biochemical lab tests) incidence of infection in different administrative-territorial units of Ukraine

White test statistics: $R_{aux} = 0.148$; $W = 0.548$; $\chi^2_{(\alpha=0.05; f=5)} = 11.07$; $W < \chi^2$, that is the model was homoscedastic.

When analyzing the obtained equation of two-factor linear regression, the question was: which of the two factors – x_4 or x_{15} – had the greatest influence on the dependent variable y_3 ? In order to answer this question, the authors conducted a procedure of standardization for regression coefficients according to the above-mentioned method. The results showed that the standardized regression coefficient (b_{st}) for the variable x_{15} was twice as high as the standardized regression coefficient for the variable x_4 (0.7 vs 0.3), which indicated that the artificial route of infection associated with blood sampling in laboratories had a more pronounced effect on the epidemic process of HC compared to sexually transmitted infection.

When modeling chronic HC epidemic process, no statistically significant regression coefficients were found for any of the factors influencing the incidence: confidence intervals included both positive and negative values, the predicted t-test value was less than critical value, and standard errors exceeded half of the parameters. In our opinion, this

is the result of underreporting of chronic HC cases in Ukraine.

Discussion. In mathematical model of the epidemic process of acute HC, the authors demonstrated statistical significance of the effect of only one variable – annual number of dentist visits. The possibility of being infected with parenteral hepatitis in dental offices seems quite expectable, although there is very little direct evidence of this in the literature. In 2013, the US Center for Disease Control and Prevention (CDC) for the first time presented a documented report on patient-to-patient transmission of HC in the dental office [8]; infection was confirmed by molecular epidemiology methods. Somewhat earlier, in 2007, molecular-epidemiological evidence of hepatitis B virus infection of a patient during dental surgery was presented [9]. The source of infection in the latter case was another patient operated on in this office 161 minutes earlier. The details of the viral hepatitis transmission mechanism in both cases were not clarified. Meanwhile, inadequate disinfection of high-speed dental handpieces is considered a potential threat in modern dental practice in the context of HC spread [10]. Recommendations for

their safe handling include rinsing with water mechanical cleaning with detergents, and sterilization; however, in many cases only one stage of treatment is used – disinfection (at best with the use of alcohol-containing substances) without cleaning the instrument after each patient. In our opinion, the risk of cross-infection may also be related to the effect of aspiration at the time of cessation of air supply to the air-turbine handpiece, which has no option to prevent back suction: the pressure in the air duct becomes lower than in the turbine area and biological fluids are sucked into the air duct, contaminating not only the turbine handpiece, but also the coupling, tube and cleaning unit. This can lead to cross-infection, even if the turbine handpiece is replaced after each patient.

According to serological studies conducted in Ukraine [4], about 4% of the population have antibodies to HC virus. Indicators of HC virus seroprevalence show that transmission of the virus in the population occurs much more often as compared to the data on the incidence of acute and chronic HC, so the regression model of the epidemic process of HC, which is based on annual number of HC virus seropositive persons, reflects the basic patterns of its hidden component.

When developing a model of the epidemic process of acute HC taking into account the annual number of seropositive individuals, statistical significance was demonstrated only for two variables: annual number of the sexually transmitted infections and annual number of laboratory blood tests. Only a few reports of viral hepatitis C and B were found in the available literature, the transmission of which had been associated with clinical laboratories. For example, the risk of being infected with viral hepatitis C and B for laboratory staff was estimated to be about 10 times higher than that for the rest of the population [11] and almost 3 times higher than for other hospital staff [12]. In the context of our results, the unusual outbreak of

hepatitis B virus described in 1974 among the staff of a clinical diagnostic laboratory, who were involved in the processing of laboratory computer charts [13], is of great interest. According to this study, the determinant and statistically valid factor of the outbreak in the lab staff was represented by obvious and hidden skin lesions, which served as the gateway to the virus during the direct contact with potentially blood-contaminated objects. In our opinion, a similar mechanism of pathogen transmission from hands of a laboratory assistant to the scarification wound of the patient may occur during blood collection, if the laboratory assistant uses rubber gloves repeatedly in terms of glove shortage and significant flow of patients.

It seems quite logical that these factors (the annual number of sexually transmitted infections and the annual number of laboratory blood tests) did not fall into the first model describing the epidemiological process of acute HC. Obviously, when the virus gets into the scarification wound from the laboratory gloves during blood sampling or during sexual intercourse, a small number of HC pathogens enter the human body, so the infectious process will not be manifested by pronounced clinical symptoms.

The study showed statistically insignificant influence of narcotics use, including opioids, over the mechanism of HC transmission in Ukraine (r_{xy} is equal relatively to: 0.354, 95% CI [-0.047; 0.657] and 0.341, 95% CI [-0.062; 0.648]), whereas the relative risk of being infected for this group is quite significant (RR = 6.5; 95% CI [6.39; 6.63]). In our opinion, this may be a consequence of relative social isolation of this category of the population. Poor socialization of drug users makes unlikely active sexual intercourse with people outside the subculture of drug addicts, and is accompanied by restrictions in the use of medical services, including dental services.

Conclusions

According to our data, at least 84% of HC virus infection cases in Ukraine occur through sexual contact and during laboratory blood sampling, and the role of the latter route of transmission in the HC virus spread was even more significant (standardized regression coefficients are 0.3 and 0.7, respectively).

Almost 80% of acute HC cases are associated with dental interventions.

Etiotropic treatment of patients with HC at the current level of treatment coverage can reduce the incidence of complications and the risk of death, but it is ineffective as a measure of influence on the first stage of the epidemiological process (source of infection).

Drug users have little effect on the intensity of the HC epidemiological process in Ukraine as a whole, despite the fact that the relative risk of HC among this population is quite significant (RR = 6.5; 95% CI [6.39; 6.63]).

Prospects for future research

Prospects for further research are related to the study of patterns and current trends in the development of HC epidemiological process in Ukraine to improve epidemiological surveillance of this infection.

Conflict of interest

The authors declare no conflict of interest.

Відомості про авторів

Галушко Наталія Анатоліївна, кандидат медичних наук, старший викладач кафедри громадського здоров'я СумДУ, вул. Римського-Корсакова, 2, м. Суми, Україна, 40007 (galnatmed@gmail.com; +38(050) 061 42 98);

Трецька Тамара Олександрівна, кандидат медичних наук, старший викладач кафедри громадського здоров'я СумДУ, вул. Римського-Корсакова, 2, м. Суми, Україна, 40007 (ttretska@ukr.net; +38(050) 999 25 43);

Галушко Андрій Вікторович; лікар Сумської районної клінічної лікарні, вул. Марка Вовчка, 2, м. Суми, Україна, 40000 (Galushcoa@gmail.com; +38(050) 684 69 92).

References (список літератури)

1. World Health Organization [Internet] *Global health sector strategy on viral hepatitis 2016-2021; 2016* [cited 2020 Feb 12]. Available from: <http://apps.who.int/iris/bitstream/10665/246177/1/WHO-HIV-2016.06-eng.pdf>
2. Talalaiev KO, Vastianov RS, Hoidyk VS, Rusnak SV. Porivnialna kharakterystyka epidemichnoi situatsii z virusnykh hepatytiv B i C v Prychornomorskomu rehioni, Ukraini ta Yevropi v period z 2013 po 2017 roku [Comparative characteristics of the epidemic situation of viral hepatitis B and C in the Black Sea region, Ukraine and Europe in the period from 2013 to 2017] *Medychni ta ekolohichni problemy prymorskykh rehioniv*. 2018; 3(84):13-21[in Ukrainian]
3. Stepanova M, Younossi ZM. Economic Burden of Hepatitis C. *J. Infection Clin. Liver Dis*. 2017; 21(3): 579-594. doi: 10.1016/j.cld.2017.03.012
4. Centr Gromads'kogo zdorovya MOZ Ukrayiny [Internet] *Hepatyt C v Ukraini: epidemiolohichna kharakterystyka ta otsinka tiaharia (za rezultatamy analizu danykh z riznykh dzherel)* [Center of Public Health of the Ministry of Health of Ukraine. Hepatitis C in Ukraine: epidemiological characteristics and burden assessment (for the results of data analysis from different sources)]. Kyiv, 2018. 111 p. [in Ukrainian]. Available from: <https://phc.org.ua/kontrol-zakhvoryuvan/virusni-gepatiti/statistika-z-vg>
5. World Health Organization [Internet]. *Progress report on access to hepatitis C treatment: focus on overcoming barriers in low-and middle-income countries*. 2018. [cited 2020 Feb 22]. Available from: www.who.int/iris/handle/10665/260445
6. Open database: Materialy DZ «Tsentru medychnoyi statystyky MOZ Ukrayiny» [Materials state institution "Center for Medical Statistics of the Ministry of Health of Ukraine"] Available from: <http://medstat.gov.ua/ukr/statdan.html> [in Ukrainian]
7. Berry Michael JA. *Data Mining Techniques: for Marketing, Sales, and Customer Relationship Management*, Gordon Linoff, 3d ed. Wiley Publishing, Inc., 2011. 847 p.
8. Weaver JM. Confirmed transmission of hepatitis C in an oral surgery office. *Anesthesia progress*. 2014; 61(3): 93–94. doi: 10.2344/0003-3006-61.3.93
9. Redd JT, Baumbach J, Kohn W, Nainan O, Khristova M, Williams I. Patient-to-Patient Transmission of Hepatitis B Virus Associated with Oral Surgery. *The Journal of Infectious Diseases*. 2007; 195 (9): 1311–1314. doi:10.1086/513435
10. Flávia Morais Gomes Pinto, Camila Quartim de Moraes Bruna, Tamara Carolina Camargo, Maira Marques, Cely Barretos Silva, Suzete M. Sasagawa, Lycia Maria Jenné Mimica, Kazuko Uchikawa Graziano. The practice of disinfection of high-speed handpieces with 70% w/v alcohol: An

- evaluation. *American Journal of Infection Control*. 2017; 45 (1): e19 - e22
11. Hadler S, Doto I, Maynard J, Smith J, Clark B, Mosley J, Eickhoff Th, Himmelsbach C, Cole W. *Occupational Risk of Hepatitis B Infection in Hospital Workers. Infection Control*. 1985; 6:24-31. doi:10.1017/S0195941700062457
12. Dienstag JL, Ryan DM. Occupational exposures to hepatitis B virus in hospital personnel: infection or immunity? *American Journal of Epidemiology*. 1982; 115(1): 26–39
13. Pattison CP, Boyer KM, Maynard JE, Kelly PC. Epidemic Hepatitis in a Clinical Laboratory: Possible Association With Computer Card Handling. *JAMA*. 1974; 230(6): 854–857. doi: 10.1001/jama.1974.03240060024023

(received 12.05.2020, published online 29.06.2020)

(одержано 12.05.2020, опубліковано 29.06.2020)