PHYSICS AND MATHEMATICS

THE ORIGINAL FORMULA FOR PREDICTING THE SURVIVAL OF GASTRIC CANCER PATIENTS UNDERGOING SURGICAL TREATMENT

[1]Sira O. V.,
[2]Kirkilevsky S. I.,
[3,5]Tkachenko O. I.,
[3]Dubinina V. G.,
[2,3,4]Mashukov A. A.,
[3,4]Bilenko A. A.,
[3,4]Merlich S. V.,
[5]Maksimovsky V. E.,
[4]Zgura A. N.,
[4]Ratsiborsky D. V.,
[4]Shilin I. V.,
[4]Sirbu V. N.,
[4]Boychenko A. I.

[1]Ukraine, Kharkiv, NTU Kharkiv National Technical University "Kharkov Polytechnic Institute", Department of Distributed Information Systems and Cloud Technologies;
[2]Ukraine, Kiev, National Institute of Cancer, Research Department of tumors of the chest cavity;
[3]Ukraine, Odessa, Odessa National Medical University;
[4]Ukraine, Odessa, KU "Odessa Regional Oncology Center", Department of abdominal oncology;
[5]Ukraine, Odessa, Odessa National Medical University, Center of Reconstructive and Plastic Surgery

Abstract. *The purpose of this work was to find ways to predict the survival of gastric cancer patients. The study included 221 patients who were radically operated in the abdominal department of the Odessa Regional Oncology Center from 2007 to 2013. The life expectancy of this group of patients was measured in months. From the factors given in the article, only the age of the patient, the presence and invasion in neighboring organs and the number of organs resected during the operation were those factors that had a significant impact on the prognosis. A formula was obtained for the formal evaluation of the duration of patients. The results are preliminary. Conclusions. As a result of the regression analysis, a polynomial (formula) was obtained, which can be used to predict the survival of patients who underwent surgery for gastric cancer. There is a need to create clearer gradations of survival dependencies of cancer patients from different clinical and morphological situations. A mathematical apparatus with many variables can be used to create similar models for the analysis of survival in other types of pathology.*

Keywords: *stomach cancer, mathematical formula, immunohistochemistry, survival.*

Many factors affect the quantitative and qualitative analysis of the life span of the treated patients. However, relying on the mathematical analysis of many factors at once, it is possible to predict the expected life expectancy of a patient with stomach cancer. It depends on many factors that are variable, i.e. they can vary depending on the particular patient. Our task was to trace such tendencies, when knowing the age of the patient, the stage of the disease, the indicators of the degree of aggressiveness of the biology of the tumor, we can make an approximate prediction of the duration of the life after the operation. At the same time, the possible other causes of the patient's death not related to the oncological disease (myocardial infarction, stroke, other causes) were not taken into account. The need for such a mathematical tool is obvious. In the world, similar models already exist and are used for other cancers [2,3,4].

Materials and methods of research. A total of 221 patients operated on for gastric cancer in the period 2007-2013 were included in the study conducted on the basis of the abdominal oncosurgical department of the Odessa Regional Oncology Center. The study was retrospective, single-center, non-randomized, including only radically or conditionally radically operated patients. The average age is

60.88 ± 10.5 years, men - 180, women - 41. A total of 143 gastrectomies and 78 distal subtotal resections were performed. Gastrectomy was performed according to the method of Bondar with the formation of a termino-lateral esophagoejuno-anastomosis. Distal subtotal resections ended in most cases with the formation of gastroenteroanastomosis according to Billroth-2 by the modification of Hofmeister-Finsterer. Mortality rate was 1.2 %, operability was 84 %. The survivability of this group of patients by stages and by type of operation was taken into account.

Mathematical justification of the performed calculations. Regression analysis – is a powerful and effective statistical method for constructing mathematical models [1], describing the relationship between the indicator of the functioning of the system being analyzed and determine *at* its explanatory independent variables (factors) $F_1, F_2, ..., F_m$. To identify this connection, a series of experiments is conducted in which each experience $F_{j1}, F_{j2}, ..., F_{jm}$ its result is the value of the dependent variable y_j, $j = 1, 2, ..., n$. The required connection is usually described by the Kolmogorov-Gabor polynomial, which in the simplest case has the form:

$$y_j = x_0 + F_{j1}x_1 + F_{j2}x_2 + ... + F_{jm}x_m + \varepsilon_j \,. \qquad (1)$$

Here, F_{ji} - the value of **i** in the jth independent variable experience, $i = 0, 1, 2, ..., m$, $j = 1, 2, ..., n$;

y_j - the meaning of the explained variable in j experience, $j = 1, 2, ..., n$;

ε_j - the value of a random error in j experience, $j = 1, 2, ..., n$.

In the matrix form, the above relation has the form $FX = Y$, where

$$F = \begin{pmatrix} 1 & F_{11} & F_{12} & ... & F_{1m} \\ 1 & F_{21} & F_{22} & ... & F_{2m} \\ ... & ... & ... & ... & ... \\ 1 & F_{n1} & F_{n2} & ... & F_{nm} \end{pmatrix}, \; X = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ ... \\ x_m \end{pmatrix}, \; Y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix}. \qquad (2)$$

Finding unknown coefficients X polynomial (1) is carried out by the method of least squares in the following way. Using (1) - (2), the predicted values of the explained variable $\hat{y}$ in each experiment, which are compared with real values. The sum of the squares of the resulting deviations is a criterion for the adequacy of the desired set of coefficients X. The formula for calculating the vector X has the form

$$X = \left(H^T H\right)^{-1} H^T Y \,. \qquad (3)$$

The accuracy of the solution of the problem depends on the correctness (adequacy) of the model (1) and the quality of the original statistical material concentrated in the matrix F. In turn, the adequacy of the model is determined by the correctness of the choice of explanatory factors (indicators) $F_1, F_2, ..., F_m$. In the real task, the following factors were selected, presumably influencing the resulting indicator y - the remaining life expectancy of the patient who underwent the operation:

F_1 - the age of patient;
F_2 - part of stomach;
F_3 - degree of differentiation of tumor ;
F_4 - invasion of wall of the stomach;
F_5 - expression of the protein of VEGFR (neoangiogenesis);
F_6 - immunohistochemical expression of p 53;

F_7 - immunohistochemical expression of her2\new ;

F_8 - microscopic invasion in the nerves;

F_9 - microscopic invasion of vessels, incl. presence of tumor emboli;

F_{10} - invasion of neighboring organs;

F_{11} - volume of lymph node dissection;

F_{12} - quantity affected by metastases regional lymph nodes;

F_{13} - the number of organs resected during the operation;

F_{14} - genetic subtype of stomach cancer (microsatellite-unstable, epstein-barr virus-induced, genetically stable, chromosomal-unstable).

Preliminary analysis of the results of direct evaluation of factor values $F_1, F_2, ..., F_{14}$ allowed to remove from this set of little informative factors F_8 F_9, whose values are more than 95 % of the cases were the same and equal to zero.

Next, a correlation analysis was made of the relationship between the remaining factors and the resulting index. According to the results of this analysis, it is traditionally recommended to consider only those factors that have a correlation coefficient of at least 0.15 significant (affecting). The resulting set of correlation coefficients is summarized in Table 1.

Table 1. Correlation coefficients between the influencing factors and the resulting variable y

Factors	F_1	F_2	F_3	F_4	F_5	F_6
Correlation coefficient	**-0,15**	-0,03	0,09	-0,14	-0,03	0,08
Factors	F_7	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}
Correlation coefficient	0,01	-0,14	-0,04	-0,13	-0,12	-0,04

It can be seen from Table 1 that the only factor $k_{F_1,y}$ describes the manifestation of the dependence of the duration of the remaining life y from the factor F_1 (age). The observed effect of the absence of correlation for other factors appears to be the result of uncontrolled and unpredictable influence on the result of any unaccounted factors (for example, concomitant pathologies, differences in the nature and characteristics of residence in the postoperative period, etc.). After removal of factors whose values are known to be correlated with the values of the resultant factor was defined set of data suitable for the first regression analysis. The resultant polynomial, as before, forms an unsatisfactory quality prediction of the value of the remaining life. With this calculated using the model value I the resulting factor is significantly different from the observed, which, apparently, is also a consequence of extraneous uncontrolled influences. The only and natural way to reduce this effect is grouping patients who have the same values of monitored indicators and averaging the resulting indicator within each group. In this case, all the values of the factor F_1 (age) are divided into 10, and the result is rounded to the nearest whole, which provided the possibility of grouping. As a result, the number of groups turned out to be [2,4]. For the obtained grouped data set, a correlation analysis was again performed, which showed that the remaining factors significantly affect the average of the group in the group of the resultant index.

Table 2. Correlation coefficients between influencing factors and the resulting variable y

Factors	F_1	F_{10}	F_{13}
Correlation coefficient	-0,34	-0,40	-0,47

Thus, the final set of input data for regression analysis was obtained.

Table 3. Values of factors and the resultant indicator for each group

Age	Invasion of neighboring organs	Multiorganic resection	Time of life, months
6	1	1	40,15
5	0	0	52,22
7	1	1	35,29
4	1	1	36,00
5	1	1	49,25
8	1	1	33,80
6	2	1	20,00
6	3	1	45,00
5	3	1	38,33
6	0	0	49,90
7	1	0	43,33
7	0	0	48,40
4	0	0	55,17
9	1	1	14,50
8	1	0	31,00
8	0	0	40,00
4	1	0	48,00
7	2	1	31,50
3	0	0	46,00
6	1	0	43,00
5	4	0	26,00
6	2	0	31,80
7	3	1	28,00
8	0	0	41,00

Results of the study. As a result of the regression analysis, the following polynomial

$$y = 58.16 - 2.31F_1 - 2.27F_{10} - 6.93F_{13}. \tag{4}$$

We check the adequacy of the regression equation obtained by the Fisher criterion.

We calculate the total scattering of the observed values of the resulting variable of relative average their value.

$$Q = \sum_{j=1}^{n} \left(y_j - \bar{y}\right)^2, \quad \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j . \tag{5}$$

We then calculate the total scattering of the observed values with respect to the values predicted by the model

$$Q_{ост} = \sum_{j=1}^{n} \left(y_j - \hat{y}_j\right)^2 .$$

Value $Q_{ост}$ It reflects the impact of the scattering of reasons that cannot be explained in the equation of no regression.

Finally, calculate the parameter $F_{наблюдаемое}$, which determines the level of adequacy of the regression equation obtained

$$F_{наблюдаемое} = \frac{\left(Q - Q_{ост}\right)\left(n - m - 1\right)}{mQ_{ост}} . \tag{6}$$

In this relation $n - m - 1$ and m - the number of degrees of freedom in calculating the values Q and $Q_{ост}$.

The calculated value $F_{наблюдаемое}$ is compared with the value $F_{критическое}$, extracted from the Fisher distribution table with a given level of significance $\alpha = 0.05$ for a given number of degrees of freedom. If in this case $F_{наблюдаемое} > F_{критическое}$, then the hypothesis of the adequacy of the regression equation (4) is accepted, otherwise this hypothesis should be rejected.

In the problem under consideration

$$Q = 2426{,}31;\ Q_{ост} = 1074{,}34;\ m = 3;\ n = 24.$$

Wherein $F_{наблюдаемое} = 8{,}39$. Value $F_{критическое} = 5{,}93$.

Thus, the model obtained is adequate. It allows to calculate, with satisfactory accuracy, the expected average value of the duration of the remaining life for a given set of values of three significant indicators: age, invasion of neighboring organs, and the number of resected neighboring organs.

For various cancer localizations, there is very specific information on the survival of patients in stages of the disease. There is now a significant need for such *reference* information for more narrow clinical situations (the number of affected lymph nodes, the volume of the tumor, measured on CT, etc.). There should be solutions [2, 4], which can then be used as a consultative reference for the patient, his relatives, planning the number of courses of chemotherapy, the degree of aggressiveness of complex treatment, given the high cost of chemotherapy, etc. Such motivation can play a significant role in the personification / individualization / modification of the therapeutic approach.

Conclusions.

1. As a result of the regression analysis, a polynomial (formula) was obtained, which can be used to predict the survival of patients who underwent surgery for gastric cancer.
2. There is a need to create clearer gradations of survival dependencies of cancer patients from different clinical and morphological situations.
3. A mathematical apparatus with many variables can be used to create similar models for the analysis of survival in other types of pathology.

REFERENCES

1. STATISTICA 6. Статистический анализ данных. Второе издание. М.: Бином, 2009. Халафян А. А. ГЛАВА 14 Анализ выживаемости.

2. A Prognostic Model to Predict Mortality among Non-Small-Cell Lung Cancer Patients in the U.S. Military Health System. Lin J, Carter CA, McGlynn KA, Zahm SH, Nations JA, Anderson WF, Shriver CD, Zhu K. J Thorac Oncol. 2015 Dec;10(12):1694-702. doi: 10.1097/JTO.0000000000000691. PMID:26473644.

3. Development of individual survival estimating program for cancer patients' management. Chang MC.Healthc Inform Res. 2015 Apr;21(2):134-7. doi: 10.4258/hir.2015.21.2.134. Epub 2015 Apr 30. PMID:25995966.

4. A new scoring system for predicting survival in patients with non-small cell lung cancer. Schild SE, Tan AD, Wampfler JA, Ross HJ, Yang P, Sloan JA. Cancer Med. 2015 Sep;4(9):1334-43. doi: 10.1002/cam4.479. Epub 2015 Jun 23. PMID: 26108458.