

УДК 681.3

О.Н. Чопоров, А.И. Агарков, Г.Я. Клименко, Ю.Ю. Шуршуков  
**ПРОГНОЗИРОВАНИЕ РАЗВИТИЯ ОНКОЛОГИЧЕСКОЙ  
ЗАБОЛЕВАЕМОСТИ ПО ИНДИВИДУАЛЬНЫМ МЕДИКО-  
СОЦИАЛЬНЫМ ФАКТОРАМ РИСКА**

*Воронежская государственная медицинская академия им. Н.Н. Бурденко  
Воронежский институт высоких технологий*

*Проведен анализ значимости медико-биологических и социально-гигиенических характеристик онкологических больных и разработаны прогностические модели, описывающие их влияние на развитие злокачественного новообразования и состояние здоровья больных.*

**Ключевые слова:** медико-социальные факторы риска, онкологическая заболеваемость, прогнозирование.

Согласно данным научно-исследовательского онкологического института им. П.А. Герцена, в Российской Федерации ежегодно диагностируется более 500 тыс. новых случаев злокачественных новообразований. В 2012 году в Российской Федерации впервые в жизни выявлено 525931 случаев злокачественных новообразований [1]. Онкологические заболевания влекут за собой ряд неблагоприятных последствий, как социального, так и медицинского порядка. Во многих случаях заболевание носит затяжной характер с частыми рецидивами, с временной и стойкой потерей трудоспособности. На развитие онкологических заболеваний значительное влияние оказывают медико-социальные факторы риска, в связи с чем, для решения задачи повышения эффективности организации медико-профилактической помощи онкологическим больным, целесообразно провести анализ взаимосвязи медико-социальных характеристик онкологических больных, выделить ведущие факторы риска, и на их основе разработать модели для прогнозирования развития заболевания и изменения состояния здоровья больных.

Для решения поставленной задачи было проведено медико-социальное исследование в ходе которого собраны, обработаны, проанализированы и обобщены результаты анкетирования 320 больных злокачественными новообразованиями (основная группа) и 336 человек, у которых отсутствовала данная патология (контрольная группа). Анализировались такие показатели, как пол, возраст, рост, вес, уровень образования, социальное положение, место работы, общий трудовой стаж, режим труда, условия труда, наличие профессиональных вредностей и периодичность их влияния, семейное положение, жилищно-бытовые условия и материальное обеспечение, оценка питания, характер сна, занятия спортом и утренней гимнастикой, курение и употребление

алкоголя, наличие хронических заболеваний, оценка состояния здоровья, наследственная предрасположенность, удовлетворенность медицинской помощью и др. Для женщин также регистрировались показатели репродуктивной функции. Большинство показателей оценивались большими ретроспективно на момент, предшествующий заболеванию и на момент настоящего обследования. Врачом заполнялись такие показатели, как основной диагноз, сопутствующие заболевания, длительность заболевания, число обращений за медицинской помощью и число госпитализаций, длительность нахождения на учете, наличие инвалидности. При этом в качестве рабочей гипотезы было принято положение что «на состояние здоровья онкологических больных оказывает влияние комплекс медико-социальных факторов риска, связанных с их условиями жизни и труда».

Для статистической обработки с использованием моделирования все данные должны быть представлены в численном виде. Учитывая, что анкетные данные были представлены как в количественном, так и в качественном виде, мы сочли уместным преобразовать информацию, содержащую фиксированные смысловые (лингвистические) значения к численному виду [2, 3].

Алгоритм преобразования следующий.

1. Показатели, имеющие два возможных значения (типа «да», «нет» или «мужской», «женский»), преобразуются соответственно в 1 и 0.

2. Лингвистические значения показателя сортируются по возрастанию значимости (например, «удовлетворительно»-«хорошо»-«отлично»; «нет»-«затрудняюсь ответить»-«да»; «легкий»-«средний»-«тяжелый» и т.д.).

В случае затруднения или неоднозначности ситуации (например, при оценке семейного положения – «женат»-«холост»-«разведен»-«вдовец»), предлагается использование метода априорного ранжирования, позволяющего объективно оценить субъективное мнение специалистов (экспертов).

3. В случае, если отличия между двумя смежными парами значений показателей неравнозначны, то осуществляется следующий этап преобразования, при котором  $m$  ( $m > 7$ ) экспертам предлагается оценить по 5-балльной шкале значимость различий  $\Delta_{i-1,i}^j$  ( $i = \overline{2, n}$ ,  $j = \overline{1, m}$ ) между «соседними» градациями показателя, отсортированными в порядке возрастания их значимости.

В случае согласия мнений участников экспертизы ( $\chi_{рас}^2 > \chi_{табл}^2$ ), их оценки усредняются:

$$\Delta_{i-1,i} = \sum_{j=1}^m \Delta_{i-1,i}^j, \quad i = \overline{2,n}. \quad (1)$$

Численная оценка значений показателя вычисляется следующим образом:

$$z_1 = 0, \quad z_i = z_{i-1} + \Delta_{i-1,i}, \quad i = \overline{2,n}, \quad (2)$$

где  $n$  – число различных значений показателя, отсортированных в порядке возрастания их значимости.

Для удобства дальнейших расчетов предлагается использовать нормированные значения полученных в соответствии с (2) численных оценок:

$$z_i^n = \frac{z_i}{z_n}, \quad i = \overline{1,n}. \quad (3)$$

В результате данной нормировки все значения показателя будут находиться в интервале  $[0,1]$ , при этом, наименее значимое («наихудшее») значение будет соответствовать нулю, а наиболее значимое («наилучшее») – единице.

Для анализа значимости медико-социальных факторов риска развития злокачественных новообразований проводилось сравнение основной и контрольной группы по t-критерию Стьюдента. Так как t-статистика пропорциональна степени различия показателей в сравниваемых группах, данная величина была принята в качестве оценки значимости соответствующих факторов риска [4, 5]. Фрагмент расчетов представлен в табл. 1.

Для оценки степени влияния медико-социальных характеристик на состояние здоровья больных со злокачественными новообразованиями производилось вычисление коэффициентов парной корреляции, характеризующих степень взаимосвязи анализируемых характеристик с самооценкой состояния здоровья больными со злокачественными новообразованиями, длительностью заболевания, числом госпитализаций. На основе полученных данных производилось ранжирование медико-социальных характеристик по степени их влияния на ведущие показатели (табл. 2).

Для прогнозирования развития злокачественных новообразований и состояния здоровья онкологических больных по индивидуальным медико-социальным характеристикам построены модели, учитывающие взаимосвязь анализируемых показателей [6, 7]. Для включения в модели отбирались наиболее значимые и не связанные друг с другом показатели. В результате были отобраны следующие показатели.

1. Для прогнозирования развития злокачественных новообразований:  $X_1$  – возраст;  $X_2$  – пол;  $X_3$  – оценка условий труда;  $X_4$  – материальное

обеспечение;  $X_5$  – индекс массы тела;  $X_6$  – системность приемов пищи;  $X_7$  – число выкуриваемых сигарет;  $X_8$  – наличие хронических заболеваний;  $X_9$  – наличие онкологических заболеваний у отца;  $X_{10}$  – наличие онкологических заболеваний по материнской линии.

2. Для прогнозирования состояния здоровья больных со злокачественными новообразованиями:  $X_{1_1}$  – возраст;  $X_{1_2}$  – уровень образования;  $X_{1_3}$  – место работы;  $X_{1_4}$  – семейное положение;  $X_{1_5}$  – материальное обеспечение;  $X_{1_6}$  – питание;  $X_{1_7}$  – характер сна;  $X_{1_8}$  – число выкуриваемых сигарет в день;  $X_{1_9}$  – число переливаний крови.

3. Для прогнозирования числа обращений за медицинской помощью:  $X_{2_1}$  – возраст;  $X_{2_2}$  – место работы;  $X_{2_3}$  – режим работы;  $X_{2_4}$  – ИМТ;  $X_{2_5}$  – употребление алкоголя;  $X_{2_6}$  – число выкуриваемых сигарет в день;  $X_{2_7}$  – группа инвалидности.

4. Для прогнозирования группы инвалидности:  $X_{3_1}$  – возраст;  $X_{3_2}$  – место работы;  $X_{3_3}$  – физическая нагрузка;  $X_{3_4}$  – жилищно-бытовые условия;  $X_{3_5}$  – ИМТ;  $X_{3_6}$  – длительность сна;  $X_{3_7}$  – наличие хронических заболеваний;  $X_{3_8}$  – число переливаний крови;  $X_{3_9}$  – длительность заболевания.

Таблица 1

Результаты ранжирования медико-социальных факторов риска развития злокачественных новообразований (фрагмент)\*

Наименование медико-социального фактора риска*	t-статистика	Ранг
Возраст	11,1228	1
Наличие хронических заболеваний	9,2381	2
Большое число сигарет, выкуриваемых за день	8,1483	3
Наличие ожирения у матери	7,7968	4
Стаж работы на последнем месте работы	7,1192	5
Общий трудовой стаж	6,9154	6
Высокий ИМТ	5,8782	7
Потребность в приобретении медикаментов	5,5205	8
Потребность в материальной помощи	5,4573	9
Невысокий рост	-5,4508	10
Загазованность на рабочем месте	5,4009	11
Женский пол	-4,9736	12
Большое число родов	4,5561	13
Наличие онкологических заболеваний по материнской линии	4,5503	14
Гипертоническая болезнь у матери	4,4693	15
Наличие токсично-химических веществ на рабочем месте	4,4272	16
Материальное обеспечение	4,3982	17
Занятия утренней гимнастикой	4,1077	18

Наличие онкологических заболеваний у родственников	4,0740	19
Переливание крови	3,9053	20
Курение	-3,7427	21
Ранний возраст начала регулярного приема алкоголя	-3,3548	22
Наличие инсульта у матери	3,3491	23
Высокий вес	3,2630	24
Большое количество профессиональных вредностей	3,2308	25
Наличие онкологического заболевания у отца	3,0704	26
Наличие онкологического заболевания у матери	3,0151	27
Питание без системы, нерегулярное	-2,8040	28
Наличие ожирения у отца	2,7371	29
Наличие атеросклероза у отца	2,7160	30

\* - все представленные медико-социальные характеристики, достоверно связаны с развитием злокачественных новообразований (при  $p < 0,05$ ).

Таблица 2

Взаимосвязь медико-социальных характеристик больных злокачественными новообразованиями с состоянием их здоровья\* (фрагмент)

Наименование медико-социальной характеристики	Состояние здоровья		Число обращений за мед. помощью		Группа инвалидности	
	г	Ранг	г	Ранг	г	Ранг
Возраст	<b>-0,4588</b>	4	<b>-0,1765</b>	30	<b>0,1405</b>	33
Социальное положение	<b>0,3096</b>	17	0,0206	115	<b>-0,1135</b>	45
Место работы в настоящее время	<b>0,4523</b>	5	<b>-0,2756</b>	13	<b>-0,4139</b>	4
Общий трудовой стаж	<b>-0,4244</b>	8	-0,0969	75	0,0845	58
Физическая нагрузка в наст. время	-0,1415	54	-0,1744	31	<b>-0,3648</b>	5
Оценка условий труда до заб-я	<b>0,4148</b>	9	0,0324	106	<b>-0,1482</b>	29
Оценка условий труда в наст. время	<b>0,4102</b>	10	0,1125	61	-0,0867	56
Нервно-эмоциональное напряжение до заб-я	0,0308	116	<b>0,2592</b>	15	<b>0,1603</b>	24
Физическое напряжение до заб-я	-0,0875	74	<b>0,1945</b>	21	<b>0,1220</b>	38
Зрительно-напряженный труд до заб-я	<b>-0,1867</b>	34	<b>-0,1875</b>	23	<b>-0,1606</b>	23
Наличие токсично-химических веществ в настоящее время	-0,0135	127	<b>-0,1720</b>	33	<b>-0,2581</b>	11
Температурный фактор в наст. время	<b>0,2512</b>	22	<b>0,1789</b>	28	0,0080	128
Жилищно-бытовые условия в наст. вр.	<b>0,3696</b>	13	0,0987	74	<b>-0,1168</b>	42
Материальное обеспечение в наст. вр.	<b>0,4922</b>	2	0,0317	108	<b>-0,1141</b>	44
ИМТ в настоящее время	0,0117	128	<b>-0,3867</b>	6	<b>-0,1892</b>	19
Число приемов пищи в день до заб-я	<b>-0,2448</b>	24	<b>0,1771</b>	29	-0,0469	89
Питаются в настоящее время	<b>0,2253</b>	30	-0,1028	71	0,0736	67

Пища в кулинарной обработке до заб-я	<b>-0,2376</b>	26	<b>-0,3231</b>	10	<b>-0,2756</b>	10
Употребляете ли алкоголь до заб-я	-0,0398	103	<b>0,3086</b>	11	<b>0,1480</b>	30
Употребляете ли алкоголь в наст. вр.	<b>0,1596</b>	45	<b>0,3470</b>	8	0,0679	72
С какого возраста курите	<b>0,2551</b>	21	0,2124	20	0,1879	20
Сколько сигарет выкуриваете в день в настоящее время	<b>-0,4312</b>	7	0,1718	34	-0,0710	69
Сколько раз переливали кровь	<b>-0,4444</b>	6	-0,1105	63	0,3275	7
Онкологическое заболевание у матери	0,0629	87	<b>-0,1816</b>	25	<b>-0,2548</b>	12
Онкологическое заболевание у отца	<b>0,2307</b>	28	-0,0632	86	-0,0640	75
Число госпитализаций в течение последних 3 лет по заболеванию	-0,0323	113	<b>0,5950</b>	2	<b>0,6166</b>	2
Наличие инвалидности по онкозаболев-ю	<b>0,1259</b>	59	<b>0,3035</b>	12	<b>0,8005</b>	1

\* - жирным шрифтом выделены коэффициенты парной корреляции, характеризующие достоверные взаимосвязи (при  $p < 0,05$ ).

Выбранные параметры использовались для построения прогностических моделей. При этом, для построения модели развития злокачественных новообразований использовалось уравнение логистической регрессии, так как зависимая переменная в этом случае может принимать только два возможных значения: «есть заболевание» и «нет заболевания». Для прогнозирования состояния здоровья, числа обращений за медицинской помощью и группы инвалидности использовалось уравнение линейной регрессии [6, 7].

В результате проведенных расчетов были получены следующие модели:

1. Вероятность развития злокачественных новообразований (Y):

$$Y = \frac{e^{\text{reg}}}{1 + e^{\text{reg}}}$$

$$\text{reg} = -6,30127 + 0,074076 * X_1 - 2,07511 * X_2 - 1,99562 * X_3 + 4,52249 * X_4 + 0,0492517 * X_5 - 0,138805 * X_6 + 0,0544903 * X_7 + 1,05885 * X_8 + 1,45151 * X_9 + 0,858707 * X_{10}$$

Шкала для оценки:

0 - 0,29 – низкая вероятность развития злокачественных новообразований;

0,30 - 0,69 – средняя вероятность развития злокачественных новообразований;

0,70 - 1,00 – высокая вероятность развития злокачественных новообразований.

2. Оценка состояния здоровья больного со злокачественными новообразованиями ( $Y_1$ ):

$$Y_1 = 0,415384 - 0,00627148 * X_{1\_1} + 0,165945 * X_{1\_2} + 0,199263 * X_{1\_3} - \\ 0,0394057 * X_{1\_4} + 0,329452 * X_{1\_5} - 0,0574219 * X_{1\_6} + \\ 0,124349 * X_{1\_7} - 0,00393056 * X_{1\_8} + 0,0353085 * X_{1\_9}$$

Шкала для оценки:

- 0 - 0,20 – неудовлетворительное состояние здоровья;
- 0,21 - 0,50 – удовлетворительное состояние здоровья;
- 0,51 - 1,00 – хорошее состояние здоровья.

3. Число обращений за медицинской помощью ( $Y_2$ ):

$$Y_2 = 4,14626 - 0,0247552 * X_{2\_1} - 3,04021 * X_{2\_2} + 4,15333 * X_{2\_3} - 0,0567784 * X_{2\_4} \\ + 4,41239 * X_{2\_5} + 0,045939 * X_{2\_6} + 6,11346 * X_{2\_7}$$

Шкала для оценки:

- 0 - 0,50 – нет обращений;
- 0,51 - 1,50 – одно обращение;
- 1,51 - 2,50 – два обращения;
- 2,51 - 3,50 – три обращения;
- 3,51 и более – более 3-х обращений.

4. Группа инвалидности ( $Y_3$ ):

$$Y_3 = 0,870471 + 0,00688315 * X_{3\_1} - 0,355974 * X_{3\_2} - 0,120115 * X_{3\_3} + \\ 0,324834 * X_{3\_4} - 0,0128372 * X_{3\_5} - 0,0368302 * X_{3\_6} + 0,00617405 * X_{3\_7} + \\ 0,0511249 * X_{3\_8} - 0,048119 * X_{3\_9}$$

Шкала для оценки:

- 0 - 0,24 – нет инвалидности;
- 0,25 - 0,49 – III-я группа инвалидности;
- 0,50 - 0,74 – II-я группа инвалидности;
- 0,75 - 1,00 – I-я группа инвалидности.

На основе построенных моделей возможно прогнозирование показателей, характеризующих зависимость развития злокачественных новообразований и состояния здоровья онкологических больных от выделенных медико-социальных факторов риска.

Апробация построенных моделей проводилась на тестовой выборке, в которую вошло 60 пациентов, у 30 из них были зарегистрированы злокачественные новообразования. Результаты верификации модели прогнозирования развития злокачественных новообразований приведены в табл. 3.

Таблица 3

Результаты верификации прогностической модели  
развития злокачественных новообразований

Группа обследованных	Прогнозируемая вероятность развития злокачественных новообразований	
	>50 %	≤50 %
со злокачественными новообразованиями (n=30)	28 чел. (90,0 %)	3 чел. (10,0 %)
без злокачественных новообразований (n=30)	5 чел. (16,7 %)	26 чел. (83,3 %)

Как видно из таблицы, вероятность правильного прогноза, в среднем, составляет 86,7 %, причем факт развития злокачественного новообразования не был спрогнозировано всего в трех случаях из тридцати (10,0 %), что вполне достаточно для применения в практике.

*Пример прогнозирования:* пациентка ( $X_2 = 0$ ) Н., 55 лет ( $X_1 = X_{1_1} = X_{2_1} = X_{2_2} = 55$ ), имеет незаконченное высшее образование ( $X_{1_2} = 0,6727$ ), в общественном производстве не занята ( $X_{1_3} = X_{2_2} = X_{3_2} = 0$ ), оценила выполняемую физическую нагрузку как «легкая» ( $X_{3_3} = 0$ ), условия труда до заболевания оценила как «хорошие» ( $X_3 = 0$ ), разведена ( $X_{1_4} = 0,3502$ ), жилищно-бытовые условия и материальное обеспечение как «до», так и «после» заболевания оценила как «хорошие» ( $X_{3_4} = X_4 = X_{1_5} = 0,7874$ ), индекс массы тела (ИМТ) равен 32,3 ( $X_5 = X_{2_4} = X_{3_5} = 32,3$ ), в настоящее время свое питание оценила как «удовлетворительное» ( $X_{1_6} = 0,5781$ ), до заболевания питалась «без системы» ( $X_6 = 0$ ), длительность ежедневного сна составляет в среднем 9 часов ( $X_{3_6} = 9$ ), страдает бессонницей ( $X_{1_7} = 0$ ), алкоголь не употребляет ( $X_{2_5} = 0$ ), до заболевания выкуривала до 40 сигарет в день ( $X_7 = 40$ ), в настоящее время не курит ( $X_{1_8} = X_{2_6} = 0$ ), имеет хронические заболевания ( $X_8 = X_{3_7} = 1$ ), было два переливания крови ( $X_{1_9} = X_{3_8} = 2$ ), у отца и по материнской линии онкологических заболеваний не было ( $X_9 = X_{10} = 0$ ), длительность заболевания составляет 2 года ( $X_{3_9} = 5$ ), имеет первую группу инвалидности ( $X_{2_7} = 1$ ).

Подставляем представленные исходные данные в разработанные модели

$$reg = -6,30127 + 0,074076*55 - 2,07511*0 - 1,99562*1 + 4,52249*0,7874 + 0,0492517*32,3 - 0,138805*0 + 0,0544903*40 + 1,05885*1 + 1,45151*0 + 0,858707*0 = 4,1676$$

$$Y = e^{4,1676} / (1 + e^{4,1676}) = 0,9847$$

$$Y_1 = 0,415384 - 0,00627148*55 + 0,165945*0,5909 + 0,199263*0 - 0,0394057*0,3502 + 0,329452*0,7874 - 0,0574219*0,5781 + 0,124349*0 - 0,00393056*0 + 0,0353085*2 = 0,452$$



$$Y_2 = 4,14626 - 0,0247552*55 - 3,04021*0 + 4,15333*0,6992 - 0,0567784*32,3 + 4,41239*0 + 0,045939*0 + 6,11346*1 = 9,968$$

$$Y_3 = 0,870471 + 0,00688315*55 - 0,355974*0 - 0,120115*0 + 0,324834*0,7874 - 0,0128372*32,3 - 0,0368302*9 + 0,00617405*1 + 0,0511249*2 - 0,048119*2 = 0,7709$$

Согласно расчетам, вероятность развития злокачественного новообразования у пациентки Н. составила 98,5 %. Прогнозируется состояние здоровья больной в 0,452 балла, что в соответствии с предложенной шкалой соответствует оценке «удовлетворительное» (от 0,21 до 0,5 баллов). Согласно расчетам, число случаев обращения за медицинской помощью по основному заболеванию составляет 10 случаев. Прогноз группы инвалидности составил 0,7709, что согласно предложенной шкале соответствует первой группе (от 0,75 до 1,0). Так как реально, была взята пациентка, имеющая онкологическое заболевание, можно было сравнить прогноз с реальными данными: оценка состояния здоровья «удовлетворительно», 12 обращений за медицинской помощью, первая группа инвалидности. То есть, приведенный пример наглядно подтверждает адекватность разработанных моделей.

## ЛИТЕРАТУРА

1. Состояние онкологической помощи населению России в 2012 году / Под ред. В.И. Чиссова, В.В. Старинского, Г.В. Петровой. – М.: ФГУ «МНИОИ им. П.А. Герцена Росмедтехнологий», 2013. – 232 с.
2. Чопоров О.Н. Оптимизация функционирования медицинских систем на основе интегральных оценок и классификационно-прогностического моделирования: дис. ... д-ра техн. наук. – Воронеж, 2001. – 325 с.
3. Чопоров О.Н. Методика преобразования качественных характеристик в численные оценки при обработке результатов медико-социального исследования / О.Н. Чопоров, А.И. Агарков, Л.А. Куташова, Е.Ю. Коновалова // Вестник Воронежского института высоких технологий. – Воронеж, 2012. - №9. – С. 96-98.
4. Чопоров О.Н. Методы анализа значимости показателей при классификационном и прогностическом моделировании / О.Н. Чопоров, А.Н. Чупеев, С.Ю. Брегада // Вестник Воронежского государственного технического университета. - Т4, №9. – 2008. – С. 92-94.

5. Choporov O.N. Infobase formation technology for medical systems analysis and modeling / O.N. Choporov, S.V. Bolgov, L.A.Kutashova, E.Y.Konovalova // Modern informatization problems in economics and safety: Proceedings of the XVIII-th International Open Science Conference (Lorman, MS, USA, January 2013). – p. 157-162.

6. Choporov O.N. Development system for data analysis, modeling and integral estimation in medical research / O.N. Choporov, S.V. Bolgov, V.A. Kutashov // Modern informatization problems in economics and safety: Proceedings of the XVIII-th International Open Science Conference (Lorman, MS, USA, January 2013). – p. 365-367.

7. Choporov O.N. Classification and prognostic modeling in medical and social research / O.N. Choporov, N.V. Naumov, N.N. Kudinova, A.I. Agarkov // Modern informatization problems in economics and safety: Proceedings of the XVIII-th International Open Science Conference (Lorman, MS, USA, January 2013). – p. 90-93.

O.N. Choporov, A.I. Agarkov, G.J. Klimenko, Y.Y. Shurshukov  
**PREDICTION ONCOLOGICAL MORBIDITY INDIVIDUAL HEALTH  
AND SOCIAL RISK FACTORS**

*Voronezh State Medical Academy of N.N.Burdenko  
Voronezh Institute of High Technologies*

*The analysis of the significance of biomedical and socio- hygienic characteristics of cancer patients and developed predictive models that describe their influence on the development of cancer and health status of patients.*

**Keywords:** medical and social risk factors , oncological morbidity, forecasting