

ЕСТЕСТВЕННЫЕ НАУКИ

УДК 681.3

ИССЛЕДОВАНИЕ МЕДИКО-СОЦИАЛЬНЫХ ХАРАКТЕРИСТИК БОЛЬНЫХ НА ОСНОВЕ МЕТОДОВ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ И ПРОГНОСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

© 2017 О. Н. Чопоров, Н. В. Некрасова, Р. Е. Джавахадзе

Воронежский институт высоких технологий (г. Воронеж, Россия)

Воронежский государственный медицинский университет

им. Н. Н. Бурденко (г. Воронеж, Россия)

Воронежский государственный технический университет (г. Воронеж, Россия)

Анализируются задачи, возникающие при исследовании медико-социальных характеристик больных с различными заболеваниями. Предложены методы их решения, в частности: сравнительный анализ медико-социальных характеристик лиц групп сравнения на основе построения таблиц и использования хи-квадрат критерия Пирсона, преобразование качественных показателей к численному виду с последующим исследованием их взаимосвязи на основе параметрических и непараметрических критериев; исследование значимости медико-социальных факторов риска развития заболевания на основе t-статистики Стьюдента; построение классификационных моделей с целью для выявления лиц, имеющих высокий риск развития заболевания; построение регрессионных моделей для прогнозирования развития заболевания на индивидуальном уровне.

Ключевые слова: медико-социальные характеристики, факторы риска, прогностические модели, кластерный анализ, медицинская статистика.

Одним из основных направлений исследований в области общественного здоровья и здравоохранения является изучение медико-социальных характеристик больных с целью выявления особенностей их образа жизни и организация медицинской помощи больным, состоящим на диспансерном учете при различных заболеваниях, а также факторов, влияющих на возникновение этих заболеваний [1-10, 13, 15-19, 21].

При проведении таких исследований возникает ряд задач, среди которых:

1) сравнительный анализ медико-социальных характеристик лиц основной (больные) и контрольной групп (лица без анализируемой патологии) [4, 8, 15, 18, 19];

2) исследование взаимосвязи анализируемых медико-социальных характеристик [1, 5, 17];

3) исследование значимости медико-социальных факторов риска развития анализируемого заболевания [2, 6, 9, 16];

4) построение классификационных моделей с целью для выявления лиц, имеющих высокую вероятность развития анализируемого заболевания [13-14];

5) построение моделей для прогнозирования, на индивидуальном уровне, развития заболевания по медико-социальным факторам риска [7, 13, 21].

При проведении сравнения индивидуальных медико-социальных характеристик у лиц, включенных в основную и контрольную группы исследования, для показателей, представленных в численном виде, используется t-критерий Стьюдента. Для данных, представленных в качественном виде используется χ^2 -критерий Пирсона.

При использовании t-критерия Стьюдента в группах сравнения для каждого анализируемого показателя осуществляется вычисление выборочного среднего и 95-ти процентных доверительных интервалов для генеральных средних [22]:

Чопоров Олег Николаевич – ВИВТ-АНОО ВО, ВГТУ, д. т. н., профессор, choporov_oleg@mail.ru.

Некрасова Наталья Васильевна – ВГМУ им. Н. Н. Бурденко, к. м. н., and.gridin2012@yandex.ru.

Джавахадзе Роман Елгуджаевич – ВГМУ им. Н. Н. Бурденко, dzhavakhadze@mail.ru.

$$\tilde{x} - \Delta_{\tilde{x}} \leq \bar{x} \leq \tilde{x} + \Delta_{\tilde{x}}, \quad (1)$$

$$\Delta_{\tilde{x}} = t \mu_{\tilde{x}}, \quad \mu_{\tilde{x}} = \frac{s}{\sqrt{n}},$$

где \bar{x} – среднее значение, \tilde{x} – выборочное среднее, $\Delta_{\tilde{x}}$ – предельная ошибка выборочного среднего значения, $\mu_{\tilde{x}}$ – стандартная среднеквадратическая ошибка, t – t-статистика Стьюдента или коэффициент доверия (при уровне значимости $\alpha=0,05$ t-статистика равна $t=1,96$), s – значение среднеквадратического отклонения в выборке, n – объем анализируемой выборки.

При расчете критерия, который используется для проверки гипотезы о выполнении условия равенства выборочных средних, применяется следующая статистика [11]:

$$t = \frac{\tilde{X}_1 - \tilde{X}_2}{s \sqrt{\frac{1}{n} + \frac{1}{m}}}, \quad (2)$$

которая имеет распределение Стьюдента с числом степеней свободы $n+m-2$.

При использовании χ^2 -критерия Пирсона выполняется вычисление с использованием следующей формулы [11]:

$$\chi^2_{расч} = \sum_{i=1}^n \sum_{j=1}^k \frac{(f_{ij} - EXP_{ij})^2}{EXP_{ij}}, \quad (3)$$

где f_{ij} – частота (наблюдаемое количество значений) в ячейке таблицы сопряженности с координатами i, j ; EXP_{ij} – количество значений, ожидаемое, в ячейке таблицы сопряженности с адресом i, j ; i – порядковый номер строки в таблице сопряженности, j – порядковый номер столбца в таблице сопряженности, n – число строк в таблице сопряженности, k – число столбцов в таблице сопряженности.

Под ожидаемой частотой понимается частота, которая в среднем ожидается для указанной ячейки, в том случае, когда признаки являются статистически независимыми. При вычислении ожидаемой частоты применяется следующая формула:

$$EXP_{ij} = \frac{f_{i \bullet} \cdot f_{\bullet j}}{f}, \quad \text{где } f_{i \bullet} - \text{сумма наблюдаемых частот для } i\text{-ой строки таблицы сопряженности, } f_{\bullet j} - \text{сумма наблюдаемых частот для } j\text{-го столбца таблицы сопряженности. Следует отметить, что при увеличении значения } \chi^2_{расч}, \text{ возрастает степень различия между фактическими и ожидаемыми частотами, и, соответственно, увеличивается шанс гипотезы, высказанной в пользу наличия достоверной статистической связи между анализируемыми признаками. Далее, с использованием статистических таблиц определяется значение } \chi^2_{табл} \text{ при } (n-1) \cdot (k-1) \text{ степеней свободы с заданным уровнем значимости } \alpha. \text{ Если } \chi^2_{расч} > \chi^2_{табл}, \text{ то гипотеза о наличии статистической независимости анализируемых признаков должна быть отвергнута при заданном значении уровня ошибки первого рода } \alpha, \text{ иначе – нет оснований, чтобы отвергнуть гипотезу. В большинстве случаев при определении допустимой ошибки первого рода, устанавливается порог } 0,05.$$

Для обеспечения возможности построения сравнительной оценки при оценке значимости факторов риска заболевания, для показателей, которые изначально представлены как в численном, так и в качественном виде, а также для решения задачи построения классификационно-прогностических моделей целесообразно данные, представленные в качественном виде преобразовать к численному [14, 20].

Численная оценка степени взаимосвязи исследуемых медико-социальных характеристик больных обычно осуществляется на основе вычисления значений коэффициентов парной корреляции. Если коэффициент парной корреляции находится в диапазоне $0 - 0,3$, то связь оценивается как «слабая», для диапазона $0,3 - 0,7$ – как связь «средней степени», для диапазона $0,7 - 1,0$ – как «полная» связь. Направленность выявленной связи определяется по знаку, стоящему перед вычисленным коэффициентом парной корреляции. Если знак положительный (+), то связь прямая, если знак отрицательный (-), то связь обратная. При исследовании вычисленных коэффициентов парной корреляции учитывались лишь те, для которых их достоверность статистически доказана.

При исследовании значимости медико-социальных факторов риска развития анализируемого заболевания проводится сравнение анализируемых показателей лиц основной и контрольной групп с использованием t-критерия Стьюдента. Учитывая обстоятельство, что t-статистика возрастает при увеличении степени различия значений анализируемых показателей в сравниваемых группах, данная величина может использоваться в качестве оценки значимости исследуемых факторов риска.

Для оценки влияния исследуемых характеристик больных на самооценку состояния их здоровья вычисляются коэффициенты парной корреляции, на основе которых определяется степень взаимосвязи анализируемых характеристик с самооценкой состояния своего здоровья больными, а также длительностью их лечения и числом случаев заболеваемости с ВУТ (временной утратой трудоспособности) на протяжении последнего года. На основе полученных данных осуществляется процедура ранжирования медико-социальных характеристик по степени их воздействия на ведущие показатели.

При выделении лиц с повышенным риском развития заболевания предложено применение классификационных моделей, позволяющих по набору индивидуальных медико-социальных факторов риска отнести поступившего нового пациента к одной из построенных формализованных групп, каждая из которых соответствует разным прогнозам развития анализируемого заболевания.

В общем виде постановка задачи моделирования выглядит следующим образом.

Имеется в наличии обучающая выборка (исходное множество объектов) пациентов:

$$P = \{p_n\}, \quad n = \overline{1, N}. \quad (4)$$

Каждый пациент характеризуется набором индивидуальных медико-социальных факторов риска (показателей):

$$p_n \rightarrow F_n = \{f_n^1, f_n^2, \dots, f_n^i, \dots, f_n^I\}, \quad (5)$$

где $i = \overline{1, I}$ – индекс показателя (фактора риска), $n = \overline{1, N}$ – порядковый номер исследуемого пациента.

Следовательно, каждому отдельному пациенту соответствует некоторая точка f^i в гиперпространстве факторов риска, а каждому элементу p_n из множества P устанавливается в соответствие некая лингвистическая переменная l_n , которая принимает одно из двух возможных значений: «отсутствие заболевания» или «наличие заболевания».

Процесс построения классификационных моделей состоит в разбиении исходного множества объектов на однородные группы, при этом модель каждой группы M_j $j = \overline{1, 2}$ формируется следующим образом:

$$M_j = \{Z_j^i, L_j\}, \quad i = \overline{1, I}, \quad j = \overline{1, 2}, \quad (6)$$

где Z_j^i – эталон параметров (значение параметров модели), L_j – название группы (лингвистическое описание модели).

Для расчета близости между анализируемыми объектами используется евклидово расстояние

$$d_{Eab} = \left(\sum_{i=1}^I (f_a^i - f_b^i)^2 \right)^{1/2}, \quad (7)$$

Для проведения классификации установлено две группы: 1) больные исследуемым заболеванием; 2) лица без исследуемой патологии.

Классификация проводится на основе предлагаемого алгоритма:

1. С использованием евклидовой метрики (7), строится матрица взаимных расстояний между всеми исследуемыми объектами $p_n \in P$.

2. Выявляются объекты p_{k_1} и p_{k_2} , наиболее удаленные друг от друга, а значения параметров для этих объектов устанавливаются в виде эталона параметров для классов: $Z_1^i = f_{k_1}^i$, $Z_2^i = f_{k_2}^i$, $i = \overline{1, I}$, а объекты p_{k_1} и p_{k_2} считаются принадлежащими, соответственно, классу K_1 и классу K_2 .

3. Формируется матрицы расстояний между всеми объектами, не вошедшими ни в один класс, и абстрактными объектами p_{z_1} и p_{z_2} имеющими параметры, соответствующие текущим значениям эталонов параметров Z_1^i и Z_2^i .

4. Из объектов, еще не классифицированных, отбирается объект p_{z_x} , который наиболее близко расположен к объектам p_{z_1} или p_{z_2} . Далее объект p_{z_x} включается в соответствующий класс и пересчитываются значения эталонов параметров данного класса (Z_1^i или Z_2^i) как усредненные значения параметров объекта p_{z_x} и предыдущего эталона параметров. Далее переопределяются абстрактные объекты p_{z_1} и p_{z_2} .

5. Если классифицированы не все объекты, выполняется переход к п. 3, иначе рассчитанные эталоны параметров Z_1^i и Z_2^i включаются в модели M_1 и M_2 .

В качестве критерия адекватности (A) построенных моделей предлагается процент правильного отнесения пациентов основной группы и группы контроля из обучающей выборки в класс, который имеет аналогичное значение лингвистического описания

модели L_j («отсутствие заболевания», «наличие заболевания»):

$$A = \frac{100}{N} \cdot \sum_{n=1}^N a_n, \quad (8)$$

$$a_n = \begin{cases} 1, & \text{если } l_n = L_j \forall l_n \in P_j, \\ 0, & \text{если } l_n \neq L_j \forall l_n \in P_j, \end{cases}$$

где P_j – множество пациентов, вошедших в j -й класс (группу); $j = \overline{1, 2}$.

Построенные таким образом классификационные модели могут быть использованы для оценки величины риска развития заболевания. При решении этой задачи сначала у пациента регистрируется множество отобранных факторов риска:

$$X = \{f_x^1, f_x^2, \dots, f_x^i, \dots, f_x^l\}$$

Для оценки риска развития заболевания используется следующее выражение:

$$P_{др} = 1 - \frac{d_{Exz1}}{d_{Exz1} + d_{Exz2}} \quad (9)$$

где d_{Exz1} – близость, вычисленная на основе формулы (7) между объектом X и «эталоном параметров» для модели «наличие заболевания»; d_{Exz2} – близость между объектом X и «эталоном параметров» для модели «отсутствие заболевания».

При прогнозировании состояния больных с анализируемым заболеванием предложено использовать регрессионные модели, на основе которых описывается взаимосвязь между показателями, характеризующими состояние здоровья больных, с отобранными их медико-социальными характеристиками (или факторами риска).

Следует отметить, что регрессионный анализ предоставляет возможность выявления зависимости, существующей между прогнозируемым показателем и факторами риска (независимыми переменными).

В обобщенном виде регрессионная модель может быть представлена следующим образом:

$$y = b_0 + \sum_{i=1}^n b_i x_i + \sum_{i,j=1}^n b_{ij} x_i x_j + \sum_{i=1}^n b_{ii} x_i^2$$

где y – зависимая переменная (прогнозируемая величина); x_i, x_j – независимые переменные (индивидуальные медико-социальные характеристики больного); b_0, b_i, b_{ij} – коэффициенты регрессионного уравнения; n – число медико-социальных характеристик больного, которые включены в модель путем отбора с использованием метода «дискретных корреляционных плеяд» [12].

Обычно сначала строится регрессионная модель, которая содержит только линейную комбинацию независимых переменных (так называемая «линейная модель»). Далее выполняется проверка значимости вычисленных коэффициентов регрессии с использованием t -критерия Стьюдента:

$$t_0 = b_0 \frac{\sqrt{n-2}}{\sigma_\varepsilon}, \quad (10)$$

$$t_i = b_i \cdot \frac{\sqrt{n-2}}{\sigma_\varepsilon} \cdot \frac{\sum (x_i - \bar{x})^2}{n}. \quad (11)$$

Значения t вычисленные на основе выражений (10)-(11) сравниваются с критическими (табличными) при заданном уровне значимости α (обычно 0,05) и числе степеней свободы $\nu = n-2$. Критические (табличные) значения $t_{кр}$ выбираются из таблицы, представляющей распределение Стьюдента.

Коэффициенты уравнения регрессии, для которых расчетное значение t оказалось меньше, чем табличное $t_{кр}$, считаются статистически незначимыми и далее исключаются из регрессионного уравнения, сама же модель заново перестраивается и выполняется проверка ее адекватности на основе F -критерия Фишера:

$$F = \frac{\sigma_{yx}^2}{\sigma_\varepsilon^2} \cdot \frac{n-m}{m-1}, \quad (12)$$

где m – количество независимых переменных, вошедших в регрессионное уравнение.

Расчетное значение F -критерия сравнивается с табличным значением (критическим) $F_{кр}$, определяемым с учетом выбранного уровня значимости (α) и числа степеней свободы $\nu_1 = m-1, \nu_2 = n-m$.

В случае, когда расчетное значение оказывается выше, чем критическое, считается, что построенная модель адекватно описывает имеющиеся экспериментальные данные, и как следствие, ее можно применять на практике. Иначе возникает потребность в усложнении модели за счет учета парных взаимодействий независимых переменных (так называемая «неполная квадратичная модель»). В случае, когда и эта модель является неадекватной, осуществляется построение полной модели в виде выражения (10) (так называемая «квадратичная модель»).

Следует отметить, что еще одной характеристикой, отражающей меру оценки точности аппроксимации эмпирических данных на основе построенной модели, является коэффициент детерминации, определяемый на основе формулы:

$$R^2 = 1 - \frac{\sigma_{y-\hat{y}}^2}{\sigma_y^2}, \quad (13)$$

где $\sigma_{y-\hat{y}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$ - дисперсия прогностических данных, полученных на основе использования модели;

$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}$ - дисперсия эмпирических данных.

При этом, чем ближе расчетное значение R^2 к единице, тем более точно разработанная модель описывает эмпирические данные.

В целом, алгоритм построения прогностических моделей, основанный на регрессионном анализе, включает 8 этапов.

1. На основе результатов экспертного опроса определяется набор факторов риска (индивидуальных медико-социальных характеристик) X_i ($i = \overline{1, N}$), которые позволяют полностью идентифицировать состояние обследуемого больного.

2. отбирается прогнозируемый показатель (которых может быть несколько) Y_j ($j = \overline{1, M}$).

3. Выполняется оценка значимости факторов медико-социальных факторов риска (индивидуальных характеристик больных) по степени их влияния на оценку состояния здоровья больного.

4. За счет исключения параметрической избыточности выполняется формирование оптимального признакового пространства.

5. Производится выбор вида регрессионной модели (линейная, неполная квадратичная, квадратичная).

6. Выполняется построение модели на основе пошагового регрессионного анализа.

7. Осуществляется оценка степени адекватности модели. В случае, когда модель адекватна исходным данным, осуществляется переход к окончанию алгоритма. В случае, если разработанная модель неадекватна, но не исчерпаны возможности ее усложнения, выполняется переход к пункту 4, иначе возникает необходимость корректировки исходной выборки (в частности, на основе использования алгоритмов, которые направлены на исключение ошибок, повышение качества информации, увеличение объема базы данных).

8. Определение параметров моделей.

Следует отметить, что регрессионные модели достаточно просты в построении. При их использовании имеется возможность исследовать зависимость выделенного контролируемого показателя от всех анализируемых факторов риска (признаков-факторов), включенных в модель. Следовательно, с использованием прогностических моделей, построенных на основе представленного алгоритма, появляется возможность «проигрывания» различных вариантов принимаемых решений и выбора оптимального воздействия, направленного на улучшение состояния здоровья больного.

Таким образом, в общем можно выделить ряд этапов анализа персональных медико-социальных характеристик больных с исследуемой патологией: 1) формирование программы исследования; 2) сбор первичных данных в соответствии с программой исследования; 3) разработка и наполнение информационной базы данных; 4) начальная статистическая обработка и сравнительный анализ характеристик в исследуемых группах; 5) исследование взаимосвязи медико-социальных характеристик больных; 6) исследование значимости медико-социальных факторов риска; 7) выбор оптимального пространства признаков для построения моделей; 8) разработка классификационных моделей; 9) формирование прогностических моделей; 10) прогнозирование и выбор оптимальных лечебно-профилактических мероприятий. Сформированные этапы анализа и соответствующие методы приведены в таблице.

Таблица

Этапы и методы анализа медико-социальных характеристик больных

№ п/п	Название этапа	Используемые методы и средства
1	Формирование программы исследования	Проведение опроса экспертов
2	Сбор первичных данных в соответствии с программой исследования	Анкетирование больных Изучение историй болезней
3	Разработка и наполнение информационной базы данных	Разработка структуры базы данных и экранных форм с использованием средств СУБД Алгоритм преобразования качественных характеристик в нормированные численные оценки

Продолжение таблицы

4	Начальная статистическая обработка и сравнительный анализ характеристик в исследуемых группах	Методы описательной статистики Построение таблиц сопряженности Сравнение по критерию χ^2
5	Исследование взаимосвязи медико-социальных характеристик больных	Регрессионный анализ Корреляционный анализ
6	Исследование значимости медико-социальных факторов риска	Априорное ранжирование Корреляционный анализ Сравнение по t-критерию Стьюдента
7	Выбор оптимального пространства признаков для построения моделей	Метод «дискретных корреляционных плед»
8	Разработка классификационных моделей	Методы кластерного анализа
9	Формирование прогностических моделей	Регрессионный анализ
10	Прогнозирование и выбор оптимальных лечебно-профилактических мероприятий	Создание компьютерной программы включающей построенные модели

В целом, можно сделать вывод, что для анализа персональных медико-социальных характеристик исследуемых больных, необходимо использование ряда алгоритмических процедур:

- проведение сравнения медико-социальных характеристик лиц основной (больные) и контрольной групп (лица, не имеющие данной патологии) с использованием χ^2 – критерия Пирсона и t-критерия Стьюдента;

- исследование взаимосвязи анализируемых медико-социальных характеристик больных на основе расчета коэффициентов парной корреляции Пирсона или коэффициентов ранговой корреляции Спирмена;

- исследование значимости медико-социальных факторов риска развития заболевания;

- построение классификационных моделей с целью выявления лиц, имеющих высокую вероятность заболевания;

- построение прогностических моделей развития заболевания.

ЛИТЕРАТУРА

1. Анализ демографической ситуации в Воронежской области за период 2010-2013 гг. / В. П. Косолапов, Г. В. Сыч, М. В. Фролов, Г. В. Ласточкина // Системный анализ и управление в биомедицинских системах. – 2014. – Т. 13. – № 3. – С. 694-701.

2. Антоненков Ю. Е. Роль медико-социальных факторов в формировании здоровья молодежи России / Ю. Е. Антоненков, И. Э. Есауленко, В. П. Косолапов // Прикладные информационные аспекты медицины. – 2015. – Т. 18. – № 3. – С. 19-22.

3. Влияние социально-экономических факторов и образа жизни на здоровье населения в Воронежской области / В. П. Косолапов, Г. В. Сыч, Н. П. Куприна [и др.] // Гигиена и санитария. – 2016. – Т. 95. – № 5. – С. 445-449.

4. К вопросу о медико-биологических и социально-гигиенических характеристиках родильниц Воронежской области / П. Е. Чесноков, Г. Я. Клименко, В. П. Косолапов, Г. В. Сыч // Врач-аспирант. – 2013. – Т. 57. – № 2.2. – С. 360-368.

5. Косолапов В. П. Медико-социальная взаимосвязь показателей репродуктивного здоровья женского населения с деятельностью и ресурсным обеспечением системы родовспоможения Воронежской области / В. П. Косолапов, Г. В. Сыч, М. В. Фролов // Системный анализ и управление в биомедицинских системах. – 2014. – Т. 13. – № 2. – С. 367-371.

6. Косолапов В. П. Медико-социальные аспекты состояния здоровья женского и детского населения в Воронежской области / В. П. Косолапов, Г. В. Сыч, М. В. Фролов // Системный анализ и управление в биомедицинских системах. – 2014. – Т. 13. – № 2. – С. 405-411.

7. Косолапов В. П. Медико-социальные аспекты управления состоянием здоровья беременных, рожениц и родильниц на региональном уровне (на примере Воронежской области) / В. П. Косолапов, Г. В. Сыч, Е. В. Енькова // Системный анализ и управление в биомедицинских системах. – 2014. – Т. 13. – № 3. – С. 742-747.

8. Косолапов В. П. Медико-социальные особенности образа жизни и здоровья детей школьного возраста / В. П. Косолапов, И. Э. Есауленко, П.Е. Чесноков // Вестник Волгоградского государственного медицинского университета. – 2010. – № 4. – С. 45-48.

9. Косолапов В.П. Исследование проблем охраны материнства и детства центрально-черноземного региона и пути их решения в современных условиях: автореф. дис. ... д-ра. мед. наук / В. П. Косолапов. – Воронеж, 2011. – 47 с.
10. Математическое и алгоритмическое обеспечение расчета медико-социальных признаков заболеваний тканей пародонта у взрослого населения / Е. А. Шлыкова, И. Э. Есауленко, В. П. Косолапов, Н. А. Гладских // Системный анализ и управление в биомедицинских системах. – 2014. – Т. 13. – № 4. – С. 947-951
11. Медик В. А. Математическая статистика в медицине / В. А. Медик, М. С. Токмачев // Учеб. пособие. – М.: Финансы и статистика, 2007. – 800 с.
12. Методика формирования информационной базы данных для проведения многоуровневого мониторинга и классификационно-прогностического моделирования / О. Н. Чопоров, О. В. Золотухин, И. И. Манакин, С. В. Болгов // Вестник Воронежского института высоких технологий. – 2015. – № 14. – С. 19-24.
13. Совершенствование профилактики алкогольной зависимости на основе анализа медико-социальных факторов риска и прогностического моделирования / В. П. Косолапов, Г. В. Сыч, И. И. Манакин, Я. Е. Львович, О. Н. Чопоров // Проблемы социальной гигиены, здравоохранения и истории медицины. – 2016. – Т. 24. – № 5. – С. 266-272.
14. Сыч Г. В. Анализ значимости индивидуальных медико-социальных факторов риска и прогностическое моделирование развития онкологических заболеваний / Г. В. Сыч, В. П. Косолапов, О. Н. Чопоров // Проблемы социальной гигиены, здравоохранения и истории медицины. – 2016. – Т. 24. – № 6. – С. 366-370.
15. Сыч Г. В. Анализ профилактических мероприятий, связанных с заболеваемостью, смертностью от злокачественных новообразований, в Воронежской области / Г. В. Сыч, В. П. Косолапов, В. П. Гулов // Системный анализ и управление в биомедицинских системах. – 2017. – Т. 16. – № 1. – С. 175-179.
16. Сыч Г. В. Влияние медико-социальных факторов риска на онкологическую заболеваемость населения Воронежской области / Г. В. Сыч, В. П. Косолапов // Системный анализ и управление в биомедицинских системах. – 2016. – Т. 15. – № 2. – С. 349-358.
17. Сыч Г. В. Исследование взаимосвязи медико-социальных характеристик онкологических больных с состоянием их здоровья / Г. В. Сыч, В. П. Косолапов, А. И. Агарков // Вестник Воронежского института высоких технологий. – 2016.0 – № 3 (18). – С. 14-19.
18. Сыч Г. В. К вопросу о медико-социальной характеристике населения воронежской области, страдающего злокачественными новообразованиями / Г. В. Сыч, В. П. Косолапов // Врач-аспирант. – 2016. – Т. 78. – № 5. – С. 87-94.
19. Сыч Г. В. Медико-биологическая характеристика женщин, больных новообразованиями на примере Воронежской области / Г. В. Сыч, В. П. Косолапов, М. В. Фролов // Системный анализ и управление в биомедицинских системах. – 2014. – Т. 14. – № 2. – С. 306.
20. Чопоров О. Н. Особенности применения методов интеллектуального анализа данных и многоуровневого мониторинга при решении задачи рационализации медицинской помощи / О. Н. Чопоров, С. В. Болгов, И. И. Манакин // Моделирование, оптимизация и информационные технологии. – 2015. – № 1 (8). – С. 2.
21. Чопоров О. Н. Прогнозирование развития алкогольной зависимости по индивидуальным медико-социальным факторам риска / О. Н. Чопоров, И. И. Манакин, В. П. Косолапов // Системный анализ и управление в биомедицинских системах. – 2015. – Т. 14. – № 3. – С. 614-617.
22. Юнкеров В. И. Математико-статистическая обработка данных медицинских исследований / В. И. Юнкеров, С. Г. Григорьев. – СПб.: ВМедА, 2002. – 266 с.

**STUDY OF MEDICAL-AND-SOCIAL CHARACTERISTICS OF PATIENTS
THROUGH THE METHODS OF MATHEMATICAL STATISTICS
AND PROGNOSTIC MODELING**

© 2017 O. N. Choporov, N. V. Nekrasov, R. E. Javakhadze

*Voronezh Institute of High Technologies (Voronezh, Russia)
Voronezh State Medical University of N.N. Burdenko (Voronezh, Russia)*

The article analyzes some problems arising in the course of researching medical and social characteristics of patients with various diseases. The authors suggest such solutions as comparative analysis of the medical-and-social indicators of persons compared on the basis of the chi-square Pierson's criterion, conversion of qualitative indicators to a numerical form, and the subsequent investigation of their interrelation on the basis of parametric and nonparametric criteria; study of the importance of medical-and-social risk factors on the basis of Student t-statistics; construction of classification models for the purpose of revealing individuals with a high risk of disease; construction of regression models for predicting the development of diseases at the individual level.

Key words: medical-and-social characteristics, risk factors, prognostic models, cluster analysis, medical statistics.