

**А.А. Большаков, Н.Е. Казимилова, А.М. Златорев, И.Н. Золотко,
И.Н. Иноходова, Л.А. Горбунова**

**ВЫЯВЛЕНИЕ И АНАЛИЗ ФАКТОРОВ, ОКАЗЫВАЮЩИХ НАИБОЛЬШЕЕ
ВЛИЯНИЕ НА РЕЗУЛЬТАТЫ ЛЕЧЕНИЯ БОЛЬНЫХ ИНФИЛЬТРАТИВНЫМ
ТУБЕРКУЛЕЗОМ ЛЕГКИХ МЕТОДАМИ СТАТИСТИЧЕСКОГО АНАЛИЗА**

Проводится определение и анализ наличия, направления, силы связи и степени влияния независимых признаков на результаты лечения у больных инфильтративным туберкулезом с помощью корреляционного анализа (построение корреляционной матрицы) и регрессионного анализа с пошаговым включением и исключением переменных в уравнение регрессии.

Инфильтративный туберкулез легких, результат лечения, корреляционный анализ, регрессионный анализ, корреляционная матрица, пошаговое включение и исключение

**A.A. Bolshakov, N.E. Kazimirova, A.M. Zlatorev, I.N. Zolotko
I.N. Inohodova, L.A. Gorbunova**

**REVEALING AND THE ANALYSIS OF FACTORS MAKING THE GREATEST IMPACT
ON RESULTS OF TREATMENT SICK INFILTRATIVE TUBERCULOSIS OF LUNGS
METHODS OF THE STATISTICAL ANALYSIS**

On results of treatment at patients infiltrative tuberculosis by means of the correlation analysis (construction of a correlation matrix) and regression analysis with step-by-step inclusion and an exception of variables definition and the analysis of presence, a direction, force of communication and degree of influence of independent signs is spent to the regress equation.

Infiltrative tuberculosis of lungs, result of treatment, the correlation analysis, regression analysis, a correlation matrix, step-by-step inclusion and an exception

Актуальность исследования. Анализ применения компьютерных систем в медицинских учреждениях показывает, что информационные технологии используются в основном для создания и форматирования текстовых документов или в диагностических и лечебных приборах, что является экономически не обоснованным и не затрагивает все функциональные возможности вычислительной техники. В то же время создание информационной системы, способной прогнозировать конечный результат применения стандартных схем лечения уже на начальном этапе обследования пациента, позволит проводить адекватную коррекцию лечебно-реабилитационных мероприятий. Такой индивидуальный подход к лечению больного будет способствовать благоприятному течению патологического процесса и повышению эффективности лечения, что открывает для пользователей новые возможности компьютерных технологий в медицине и подчеркивает актуальность поставленной задачи.

Цель исследования: выявление групп риска больных по неблагоприятному течению туберкулеза легких на основании определения характера взаимодействия и степени влияния

социально-демографических и клинико-лабораторных параметров больных инфильтративным туберкулезом легких на результаты лечения.

Материалы и методы: обследовано 118 пациентов, больных инфильтративным туберкулезом легких, по 22 признакам, статистическая обработка данных проводилась с использованием ППП Statistica 7.0.

Каждый пациент обследован по 22 признакам, которые закодированы соответствующим образом: результат лечения (качественный, порядковый): 1 – ухудшение, 2 – без перемен, 3 – улучшение, 4 – значительное улучшение (x_1); пол (качественный, номинальный, бинарный): 1 – мужской, 2 – женский (x_2); возраст (кол-во лет) (количественный, непрерывный, интервальный) (x_3); трудовая занятость (качественный, порядковый): 0 – не занят, 1 – занят (x_4); образование (качественный, порядковый): 1 – среднее, 2 – высшее (x_5); давность заболевания (мес.) (количественный, непрерывный, интервальный) (x_6); объем поражения (качественный, порядковый): 1 – до доли, 2 – легкое, 3 – оба легких (x_7); выраженность остаточных изменений (качественный, порядковый): 1 – не выражены, 2 – выражены (x_8); непереносимость препаратов (качественный, порядковый): 1 – низкая, 2 – высокая (x_9); выраженность интоксикационных жалоб (качественный, порядковый): 1 – выражены, 0 – не выражены (x_{10}); выраженность грудных жалоб (качественный, порядковый): 1 – выражены, 0 – не выражены (x_{11}); эритроциты $\cdot 10^{12}/л$ (количественный, непрерывный, интервальный) (x_{12}); лейкоциты $\cdot 10^9/л$ (количественный, непрерывный, интервальный) (x_{13}); семейное положение (качественный, порядковый): 1 – холост/не замужем, 2 – женат/замужем (x_{14}); СОЭ, мм/ч (количественный, непрерывный, интервальный) (x_{15}); наличие детей (количественный, непрерывный, интервальный) (x_{16}); регион проживания (качественный, порядковый): 1 – южный, 2 – центральный, 3 – дальний (x_{17}); фаза распада при выявлении (качественный, порядковый): 0 – нет, 1 – есть (x_{18}); адаптивный показатель по крови (качественный, порядковый): 1 – стресс, 2 – тренировка, 3 – СА, 4 – ПА (x_{19}); Количество сопутствующих заболеваний (количественный, дискретный) (x_{20}); длительность лечения (количественный, непрерывный, интервальный) (x_{21}); выраженность нарушения режима (качественный, порядковый): 1 – выражен, 0 – не выражен (x_{22}).

Таблица 1

Корреляционная матрица

Переменная	Gamma Correlations MD pairwise deleted Marked correlations are significant at $p < 0,05000$				
	Рез-тат леч.	Адап. показ. по крови	Фаза распада при выявл.	Регион прожив. (удален. от)	Дети
Рез-тат леч.	1,000000	-0,142857	0,287356	0,059349	0,02162
Адап. показ. по крови	-0,142857	1,000000	-0,071672	0,149843	0,11636
Фаза распада при выявл.	0,287356	-0,071672	1,000000	-0,304752	-0,30973
Регион прожив. (удален. с)	0,059349	0,149843	-0,304752	1,000000	0,08432
Дети	0,021615	0,116364	-0,309735	0,084317	1,000000
Семейное положение	-0,097784	0,072405	-0,384615	0,130268	0,88687
Образование	-0,222222	-0,186441	0,584000	-0,119403	-0,38235
Труд. занят.	-0,004733	0,044950	-0,414671	-0,142285	0,29319
Пол	-0,106808	0,060902	-0,456044	0,263249	0,38158
Возраст, лет	-0,055874	-0,016738	-0,204486	0,040738	0,46762
Давность забол. (мес.)	0,261588	0,023112	0,052083	0,040792	-0,17846
Объем пораж.	-0,151163	-0,181590	0,517761	-0,332231	-0,44499
Выраж. ост. изм.	-0,958333	0,371134	0,700000	-0,414634	-1,00000
Неперен. препарат.	-0,004721	-0,033479	0,101124	0,033103	0,10488
Выраж. жалоб интокс.	0,647590	-0,281407	0,423488	-0,212871	-0,28358
Выраж. жалоб груд.	0,505005	-0,438572	0,303571	-0,079412	-0,37161
Эритр.	-0,112768	-0,092880	0,211030	-0,137110	-0,14075
Лейкц.	-0,068831	-0,094646	0,153703	-0,113347	-0,28169
СОЭ	0,044842	-0,016238	-0,110329	0,210427	0,04441
Кол. соп. забол.	0,075323	-0,092222	-0,117155	0,054029	-0,00157
Длит. леч. дн.	0,415709	0,010526	0,212955	0,246322	-0,10205
Выражен. наруш. режима	-0,351399	0,210072	0,365385	0,007813	-0,67664

Таблица 2

Коэффициенты корреляции

Признак	Коэффициент корреляции
Фаза распада при выявлении	0,28
Давность заболевания	0,26
Выраженность остаточных изменений	-0,95
Выраженность интоксикационных жалоб	0,64
Выраженность грудных жалоб	0,50
Длительность лечения	0,41
Выраженность нарушения режима	-0,35

Для установления наличия, направления и силы линейной статистической связи между результатом лечения пациента и остальными признаками в выборке используем корреляционный анализ. Исследуемая выборка является однородной и количество наблюдений в ней в 5 раз превышает количество признаков (118 и 22 соответственно), следовательно, она отвечает всем рекомендациям для проведения статистического анализа.

Так как исследуем взаимосвязь качественного признака с другими количественными и качественными признаками и в нашей матрице исходных данных имеется много совпадающих значений, применяется метод непараметрического корреляционного анализа – гамма-корреляция. В табл. 1 приведена часть корреляционной матрицы, полученной при анализе таблицы исходных данных больных инфильтративным туберкулезом.

Из корреляционной матрицы для $p < 0,05$ видно, что следующие признаки имеют значимую степень корреляции с результатом лечения (табл. 2).

Из полученных данных (табл.2) видно, что такие признаки, как длительность лечения (x_{21}) и выраженность нарушения режима (x_{22}), имеют значимую степень корреляции с результатом лечения, но оценка этих признаков не может быть выполнена на начальном этапе обследования пациентов. В связи с этим на основе корреляционной матрицы (табл. 1), определим те признаки, которые не только непосредственно взаимосвязаны с данными параметрами и коррелируют с результатом лечения, но и с медицинской точки зрения могут найти обоснованное объяснение своего влияния на конечный результат лечения.

Из корреляционной матрицы для $p < 0,05$ видно, что следующие признаки имеют значимую степень корреляции с длительностью лечения (табл. 3).

Таблица 3

Коэффициенты корреляции клинических признаков с длительностью лечения

Признак	Коэффициент корреляции
Фаза распада при выявлении	0,21
Регион проживания	0,24
Семейное положение	-0,24

Из корреляционной матрицы для $p < 0,05$ видно, что следующие признаки имеют значимую степень корреляции с нарушением режима (табл. 4).

Выявленные особенности взаимосвязи отдельных социально-демографических и медицинских признаков позволяют не исключать их при проведении последующих этапов математического анализа.

Построенная корреляционная матрица (табл. 1) показывает корреляционную зависимость свободных признаков, на

Таблица 4

Коэффициенты корреляции клинических признаков с нарушением режима

Признак	Коэффициент корреляции
Фаза распада при выявлении	0,36
Дети	-0,67
Семейное положение	-0,59
Образование	-0,44
Трудовая занятость	-0,70
Пол	-0,54
Возраст	-0,21
Давность заболевания	0,30
Объем поражения	0,50
Выраженность остаточных изменений	0,30
Непереносимость препаратов	-0,54
Количество лейкоцитов	0,25

основании которой можно построить статистическую модель, что в дальнейшем позволит выявить дополнительные признаки, в значительной мере влияющие на результат лечения.

Для выявления степени влияния независимых признаков на зависимый (результат лечения) используем регрессионный анализ с пошаговым включением и исключением переменных в уравнение регрессии.

Общий вид уравнения множественной линейной регрессии имеет вид

$$y = a_0 + a_1x_1 + \dots + a_mx_m(2),$$

где y – зависимая переменная, $x_1..x_m$ – независимые факторы, $a_0..a_m$ – искомые коэффициенты.

При пошаговом включении независимая переменная включается в уравнение в случае, если она существенно увеличивает значение множественного коэффициента корреляции, т.е. значение t-критерия оценок коэффициентов выше критического уровня. В первую очередь в уравнение множественной линейной регрессии включаем длительность лечения (x_{21}), т.е. ту переменную, которая наиболее сильно влияет на результат лечения (коэффициент множественной корреляции 0,406, коэффициент множественной детерминации = 0,164, изменение коэффициента множественной детерминации = 0,164), далее в уравнение включаем переменную (давность заболевания (x_6)), которая вместе с первой формирует максимальное изменение значения множественного коэффициента корреляции, и т.д. После включения всех необходимых переменных отбираем независимые признаки, для которых $p < 0,05$.

В табл. 5 представлены результаты регрессионного анализа с пошаговым включением переменных в уравнение регрессии. где, Step +in/-out – шаг включения/исключения, Multiple R – коэффициент множественной корреляции, Multiple R-square – коэффициент множественной детерминации, R-square change – изменение коэффициента множественной детерминации, F- to entr/rem – критерий F , p-level – уровень значимости, Variables included – количество включенных переменных.

Таблица 5

Результаты регрессионного анализа с пошаговым включением переменных

Переменная	Summary of Stepwise Regression; DV: Рез-тат леч.						
	Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variables included
Давность забол.(мес.)	1	0,295530	0,087338	0,087338	8,708324	0,004028	1
Выраж. жалоб груд.	2	0,371694	0,138156	0,050818	5,306794	0,023546	2
Лейкц.	3	0,430137	0,185018	0,046862	5,117563	0,026119	3
Объем пораж.	4	0,469622	0,220545	0,035527	4,010993	0,048282	4
Выраж. жалоб интокс.	5	0,495225	0,245248	0,024702	2,847432	0,095102	5
Выраж.ост.изм.	6	0,529637	0,280515	0,035267	4,215506	0,043095	6
СОЭ	7	0,550226	0,302749	0,022234	2,710434	0,103387	7
Фаза распада при выявл.	8	0,558021	0,311387	0,008639	1,053793	0,307582	8

Из полученных результатов (табл. 5) видно, что при условии $p < 0,05$ наибольшую степень влияния на результат лечения оказывают такие факторы как давность заболевания (x_6), выраженность грудных жалоб (x_{11}), количество лейкоцитов в крови (x_{13}), объем поражения легочной ткани (x_7), выраженность остаточных изменений в легких и плевре (x_8). Следовательно, уравнение множественной линейной регрессии примет вид

$$y = a_0 + a_6x_6 + a_{11}x_{11} + a_{13}x_{13} + a_7x_7 + a_8x_8.$$

При регрессионном анализе с пошаговым исключением после построения уравнения регрессии и оценки значимости всех множественных коэффициентов регрессии из уравнения исключаем ту переменную (количество лейкоцитов), значение t-критерия которой ниже критического уровня, т.е. наименьшим образом влияет на коэффициент множественной корреляции

($R = 0,705$; $R^2 = 0,497$; изменение $R^2 = -0,0002$, $p = 0,867$). Аналогично исключаются всех остальных переменных, с последующим отбором тех, для которых значение $p < 0,05$. Из полученных результатов видно, что при условии $p < 0,05$ наибольшую степень влияния на результат лечения оказывают такие факторы как скорость оседания эритроцитов (СОЭ) (x_{15}), выраженность грудных жалоб (x_{11}), выраженность остаточных изменений (x_8) и давность заболевания (x_6). Следовательно, уравнение множественной линейной регрессии примет вид

$$y = a_0 + a_{15}x_{15} + a_{11}x_{11} + a_8x_8 + a_6x_6.$$

Таким образом, по итогам проведенного исследования поставленная цель достигнута, т.е. определены группы признаков, наиболее взаимосвязанных с результатом лечения больных инфильтративным туберкулезом. Среди социально-демографических показателей выделены такие признаки, как пол, трудовая занятость, семейное положение, количество детей, уровень образования, регион проживания и возраст больного. Из клинико-лабораторных параметров отмечены выраженность остаточных изменений, интоксикационных и грудных жалоб, наличие фазы распада при выявлении, давность заболевания, объем поражения, непереносимость препаратов, количество лейкоцитов и скорость оседания эритроцитов. Использование математических методов статистического анализа позволило выявить ряд наиболее информативных медицинских факторов, определение которых на первоначальном этапе обследования больных туберкулезом легких позволит выделить пациентов с высокой вероятностью неблагоприятного течения реабилитационного процесса для коррекции лечебных мероприятий.

Большаков Александр Афанасьевич –

доктор технических наук, профессор кафедры «Системы искусственного интеллекта», декан факультета электронной техники и приборостроения Саратовского государственного технического университета

Казмирова Наталья Евгеньевна –

доктор медицинских наук, профессор, заведующая кафедрой «Фтизиопульмонология» Саратовского государственного медицинского университета

Златорев Артем Михайлович –

аспирант кафедры «Фтизиопульмонология» Саратовского государственного медицинского университета

Золотько Илья Николаевич –

аспирант кафедры «Системы искусственного интеллекта» Саратовского государственного технического университета

Иноходова Ирина Николаевна –

аспирант кафедры «Фтизиопульмонология» Саратовского государственного медицинского университета

Горбунова Людмила Анатольевна –

аспирант кафедры «Фтизиопульмонологии» Саратовского государственного медицинского университета

Статья поступила в редакцию 25.09.09, принята к опубликованию 25.11.09