

УДК 519.2:61

Определение статистически значимых факторов эмболии при тромбозе вен нижних конечностей

*А.А. Шайдуров¹, А.В. Акатов², В.В. Щербинин^{1,3}, Н.Г. Хорев⁴,
А.П. Момот⁵, М.И. Неймарк⁴, А.В. Беллер²*

¹ Алтайский государственный университет (Барнаул, Россия)

² Отделенческая клиническая больница на станции Барнаул ОАО «РЖД» (Барнаул, Россия)

³ Институт физики им. Л.В. Киренского Сибирского отделения Российской академии наук (Красноярск, Россия)

⁴ Алтайский государственный медицинский университет (Барнаул, Россия)

⁵ Алтайский филиал гематологического научного центра (Барнаул, Россия)

Identification of Statistically Significant Factors of Embolism Due to Vein Thrombosis of Lower Limbs

*A.A. Shaidurov¹, A.V. Akatov², V.V. Scherbinin^{1,3}, N.G. Horev⁴,
A.P. Momot⁵, M.I. Neimark⁴, A.V. Beller²*

¹ Altai State University (Barnaul, Russia)

² Regional Clinical Hospital at Barnaul railway station, OJSC «RZHD» (Barnaul, Russia)

³ L.V. Kirensky Institute of Physics, Siberian Branch of the Russian Academy of Sciences (Krasnoyarsk, Russia)

⁴ Altai State Medical University (Barnaul, Russia)

⁵ Altai Branch of Hematology Research Center (Barnaul, Russia)

Рассматривается возможность оценки факторов эмболизации при тромбозе системы нижней полой вены. В качестве инструментария используются статистические однопараметрические и многопараметрические методы анализа данных. Несмотря на то, что современная терапия позволяет предупредить возникновение тромботического события, ее проведение сопряжено с опасностью развития кровотечений. Хирургическое лечение тромбоза вен нижних конечностей также является сложным по технике выполнения оперативным вмешательством, с далеко не однозначными показаниями и противопоказаниями. Предполагается, что у больных имеются предрасполагающие факторы к возникновению этого осложнения. В силу этого выявление факторов, способствующих возникновению эмбологенности, явилось целью нашего исследования. Для выявления факторозависимости тромбоэмболии легочной артерии и показателей конституции, возраста, пола, анализов венозной крови, гемостаза и сопутствующей патологии в случайной выборке был проведен корреляционный анализ. Для установления возможных комбинаций параметров, оказывающих влияние на наличие тромбоэмболии легочной артерии, осуществлен дискриминантный анализ данных.

In the paper, the identifiability of embolization factors due to thrombosis of inferior vena cava using one-parameter and multi-parameter statistical methods is investigated. Despite the fact that modern therapy can prevent the occurrence of thrombotic events, its delivery entails a hemorrhage risk. Surgical treatment of lower limbs venous thrombosis is complicated due to the technique of surgical intervention with indications and contraindications being far from unambiguous. It is assumed that patients have predisposing factors to such complications. Therefore, the goal of the study is to identify the embologenic factors. To reveal the dependencies between pulmonary embolism and bodily constitution, age, sex, test results of venous blood, hemostasis, and comorbidity, a correlation analysis has been conducted in a random sample. Further, a discriminant analysis has been performed for identification of the potential combination of parameters that characterize the pulmonary embolism.

Ключевые слова: статистические методы анализа данных, корреляционный анализ, дискриминантный анализ, тромбоз системы нижней полой вены, тромбоэмболия легочной артерии.

DOI 10.14258/izvasu(2015)1.2-31

Введение. Проблема нахождения функциональной зависимости по эмпирическим данным традиционно является актуальной для различных отраслей науки. Обычно эти задачи решаются при помощи методов интерполяции или экстраполяции. Кроме того, нередко применяются различные статистические методы, в большинстве случаев сводящиеся к построению линейных регрессионных уравнений. В случае диагностики заболеваний постановка диагноза затруднена постоянным изменением объективных данных о состоянии здоровья пациента и большими объемами этой информации. Поэтому одним из путей повышения эффективности диагностики является применение математических методов анализа информации и разработка проблемно ориентированных систем обработки данных [1].

Одной из важных проблем в современной медицине является сложность выявления факторов эмболизации при тромбозе системы нижней полой вены. Частота возникновения тромботических поражений вен нижних конечностей постоянно увеличивается [2]. Это состояние чревато опасностью развития тромбоза легочной артерии (ТЭЛА) и требует профилактических мероприятий.

В настоящее время наиболее используемым методом лечения тромбозов бассейна нижней полой вены остается антикоагулянтная терапия. Однако даже использование этого метода не может предотвратить в 100% случаев миграцию тромба из бассейна нижней полой вены в легочную артерию [3, 4, 5].

Не все тромбозы бассейна нижней полой вены осложняются ТЭЛА. По данным разных исследований, тромбоз легочной артерии возникает лишь в 10-15% случаев тромбоза вен нижних конечностей [6]. По-видимому, у ряда больных имеется предрасположенность к возникновению этого осложнения. Выявление факторов, способствующих возникновению эмбологенности, явилось целью нашего исследования.

В ходе проводимого эксперимента были использованы данные 201 истории болезни пациентов с флеботромбозами, поступивших в отделение сосудистой хирургии НУЗ ОКБ на ст. Барнаул ОАО РЖД в период 2010–2014 гг. включительно.

В качестве учитываемых факторов были выбраны сезонность поступления пациентов в стационар, индекс массы тела, пол пациента, уровень гемоглобина, уровень гематокрита, количество эритроцитов, уро-

Keywords: statistical methods of data analysis, correlation analysis, discriminant analysis, thrombosis of the inferior vena cava, pulmonary embolism.

вень протомбина, уровень фибриногена, количество тромбоцитов в венозной крови. Кроме того, анализировалось наличие следующих сопутствующих патологий: ишемическая болезнь сердца, постинфарктный кардиосклероз, гипертоническая болезнь, хроническая обструктивная болезнь легких.

Общая характеристика исследуемых факторов. Все пациенты находились в стационаре с диагнозом «флеботромбоз». Из 201 человек у 30 была зарегистрирована ТЭЛА. Средний возраст пациентов составил $61,19 \pm 13,98$ года. Количество женщин составило 144, мужчин — 57 человек. Средний индекс массы тела — $32,56 \pm 5,02$. Поступление весной — 102 человека, осенью — 99 человек. Средний уровень гемоглобина в группе составил $129,4 \pm 20,5$ г/л, гематокрита — $39,72 \pm 5,35$, эритроцитов — $4,49 \pm 0,55$, уровень тромбоцитов — $194,35 \pm 84,92$, фибриноген — $3,51 \pm 1,46$, протромбиновое время — $16,20 \pm 3,42$ с.

Сопутствующая патология: гипертонической болезнью страдало 64,8% пациентов, хронической обструктивной болезнью легких — 2,6%, ишемической болезнью сердца — 21,4%, постинфарктный кардиосклероз был у 4,6% пациентов, нарушения ритма присутствовали у 12,8% пациентов.

Статистический анализ данных. На первом этапе с целью выявления факторозависимости тромбоэмболии легочной артерии и показателей конституции, возраста, пола, анализов венозной крови, гемостаза и сопутствующей патологии в исходной выборке был проведен корреляционный анализ.

В результате выявлено, что в исходной выборке исходные данные не оказывали влияния на наличие тромбоэмболии легочной артерии.

Для установления возможных комбинаций параметров, оказывающих влияние на наличие тромбоэмболии легочной артерии, проведен дискриминантный анализ данных. Выбор такого метода статистической обработки данных был обусловлен следующими причинами. Дискриминантный анализ предназначен для отнесения неклассифицированных объектов к заранее известным группам. Метод дискриминантного анализа вырабатывает некоторые решающие правила, позволяющие отнести предлагаемые объекты к заданным группам. При использовании дискриминантного анализа Фишера решающие правила получаются в виде простых линейных классифицирующих функций. Это позволяет оценить качество классифи-

кации при помощи тестирования на другой выборке с использованием дополнительных статистических параметров.

При анализе данных использовались часто применяемые в медицинской статистике параметры, такие как чувствительность и специфичность. Их аналогами в технике являются ошибки первого и второго рода.

Чувствительность определяется как доля лиц с положительным результатом теста в популяции с изучаемым заболеванием. Чувствительный тест редко пропускает пациентов, у которых имеется болезнь, т. е. его следует выбирать, если имеется риск пропустить опасную, но излечимую болезнь.

Специфичность – это доля лиц с отрицательным результатом теста в популяции без изучаемой болезни. Специфичный тест, как правило, не относит здоровых к категории больных. Специфичные тесты нужны для подтверждения (установления) диагноза, предложенного на основании других данных.

Исходя из таблицы 1 чувствительность (Se) и специфичность (Sp) можно рассчитать следующим образом:

$$Se = \frac{a}{a + c}; \quad (1)$$

$$Sp = \frac{d}{b + d}. \quad (2)$$

Высокочувствительные тесты рекомендуются на ранних стадиях диагностического поиска для сужения его рамок, когда возможных вариантов много и диагностические тесты позволят исключить некоторые из них, т.е. сделать вывод, что эти заболевания маловероятны. А высокоспецифичный тест не должен дать положительный результат в отсутствие заболевания. Высокоспецифичные тесты особенно необходимы, если ложноположительный результат может нанести пациенту вред.

Целью проведения статистической обработки данных с помощью дискриминантного анализа являлось определение списка дискриминирующих параметров для определенного диагноза и системы линейных уравнений относительно исходных данных, определяющих диагноз.

При применении дискриминантного анализа обычно имеется несколько переменных, и задача состоит

в том, чтобы установить, какие из переменных вносят свой вклад в дискриминацию между совокупностями. Другими словами, должна быть построена модель, позволяющая лучше всего предсказать, к какой совокупности будет принадлежать тот или иной образец. При этом окончательный критерий значимости того, разделяет переменная две совокупности или нет, дает F-критерий. F-статистика вычисляется, как отношение межгрупповой дисперсии к объединенной внутригрупповой дисперсии. Если межгрупповая дисперсия оказывается существенно больше, тогда это должно означать различие между средними значениями. Таким образом, основная идея дискриминантного анализа заключается в том, чтобы определить, отличаются ли совокупности по среднему какой-либо переменной (или линейной комбинации переменных), и затем использовать эту переменную, чтобы предсказать для новых членов их принадлежность к той или иной группе.

Наиболее общим применением дискриминантного анализа является включение в исследование переменных или их исключение, с целью определения тех из них, которые наилучшим образом разделяют совокупности между собой. В пошаговом анализе дискриминантных функций модель дискриминации строится по шагам.

При проведении исследований использовались следующие критерии отбора:

1. Межгрупповая F-статистика учитывает выборочные размеры классов. Расстояния для малых классов получают меньшие веса, чем расстояния для больших классов. Таким образом, этот критерий стремится увеличить различия между парами, содержащими большие группы.

2. Статистика F-включения представляет собой частную F-статистику, оценивающую улучшения различения от использования рассматриваемой переменной по сравнению с различием, достигнутым с помощью других уже отобранных переменных. Если величина статистики F-включения мала, то такая переменная не дает достаточно большого вклада в различения и она не отбирается.

3. Статистика F-удаления также является частной F-статистикой. Однако она оценивает значимость ухудшения различения после удаления переменной из списка уже отобранных переменных. Эта процедура проводится в начале каждого шага, чтобы про-

Таблица 1

Соотношение между результатами теста и наличием заболевания

Тест	Болезнь	
	Присутствует	Отсутствует
Положительный	Истинно положительный (a)	Ложноположительный (b)
Отрицательный	Ложноотрицательный (c)	Истинно отрицательный (d)

верить, имеется ли какая-нибудь переменная, уже не вносящая достаточно большого вклада в различения, поскольку отобранные позже переменные дублируют ее вклад.

Для оценки исходных параметров параметр толерантности был выбран равным 0.01, значение статистики удаления — нулевым. В процессе работы подстраивался параметр статистики включения. Первоначально он задавался равным 0.01 и увеличивался до тех пор, пока число выделенных дискриминирующих параметров не оказывалось слишком малым. На каждом этапе рассчитывались значения чувствительности и специфичности. В итоге из всего количества полученных наборов параметров выбирался тот набор, который имел отношение чувствительности и специфичности, близкое к оптимальному.

В результате получен оптимальный набор статистически значимых исходных данных, обладающий наилучшими значениями чувствительности и специфичности. Таким образом, на основе дискриминантного анализа была построена статистическая модель, позволяющая определять наличие тромбоэмболии легочной артерии у пациента с флеботромбозом.

В ходе проводимого исследования был выявлен комплекс параметров, оказывающих влияние на наличие тромбоэмболии легочной артерии, представленный в таблице 2.

В таблице 3 представлена полученная классификационная матрица для выявленных дискриминирующих параметров.

Таким образом, построенная статистическая модель, позволяющая определять наличие тромбоэмболии легочной артерии, представляет собой высокочувствительный и высокоспецифичный тест с показателями, рассчитанными по формулам (1) и (2):

$$Se = 129/132 = 0,98$$

$$Sp = 24/25 = 0,96$$

Выводы. В результате проведенных исследований была определена прямая и обратная зависимость между индексом массы тела, временем поступления пациента, и наличием или отсутствием ТЭЛА, а также присутствием у пациента постинфарктного кардиосклероза и ТЭЛА в анализируемой выборке. Также в данной выборке были выявлены прямые зависимости ТЭЛА от уровня гемоглобина, количества эритроцитов и тромбоцитов в венозной крови. Однако обратная зависимость между этими показателями отсутствует.

Зависимость между наличием сопутствующей патологии в виде гипертонической болезни, хронической обструктивной болезни легких, ишемической болезни сердца, нарушениями сердечного ритма,

Таблица 2

Список выявленных статистически значимых параметров

Параметр	Показатель Wilks'	F-значение	p-уровень	Толерантность
Фибриноген	0,20	29,94	0,00	0,80
Возраст	0,17	5,03	0,03	0,87
Индекс массы тела	0,17	2,10	0,15	0,90
Постинфарктный кардиосклероз	0,17	2,46	0,12	0,94
Тромбоциты	0,17	4,24	0,04	0,83
Время поступления: весна — 0, осень — 1	0,16	1,41	0,24	0,93
Гемоглобин	0,17	3,03	0,08	0,35
Эритроциты	0,16	1,74	0,19	0,38

Таблица 3

Классификационная матрица

ТЭЛА	Процент совпадений	G_1:0	G_2:1
		количество записей	
Отсутствует (G_1:0)	99,23	129	1
Присутствует (G_2:1)	88,88	3	24
Итоговый процент совпадений	97,45	132	25

а также уровнем гематокрита венозной крови и эмболизацией легочной артерии при флеботромбозе не найдена.

По результатам исследований можно говорить об определении тенденции, распространяемой на всю группу анализируемых пациентов с флеботромбозом.

Тем не менее для более достоверного статистического анализа и выявления более четких зависимостей необходима репрезентативная выборка, которая позволит говорить уже не о тенденциях, а о достоверных взаимосвязях вышеуказанных факторов с ТЭЛА в группе пациентов с острыми тромботическими со-

бытиями в бассейне нижней полой вены. Таким образом, к основным выводам можно отнести:

1. Обнаружена тенденция к развитию ТЭЛА при флеботромбозах у пациентов с наиболее высоким индексом массы тела, количеством тромбоцитов, уровнем фибриногена и гемоглобина в венозной крови.

2. Установлена тенденция зависимости ТЭЛА при флеботромбозах от сезонности, возраста.

3. Наибольшая тенденция к развитию ТЭЛА определяется у пациентов, имеющих в анамнезе постинфарктный кардиосклероз.

Библиографический список

1. Юнкеров В.И., Григорьев С.Г. Математико-статистическая обработка данных медицинских исследований. — СПб., 2002.

2. Gavorník P., Dukát A., Gašpar E., Gavorníková E. Present and future in the management of venous vascular diseases // *Vnitr. Lek.* — 2015. — Winter 61 (2):151-6.

3. Matsuo H, Prins M, Lensing AW, Fujinuma EW, Miyamoto Y, Kajikawa M. Shortened length of hospital stay with rivaroxaban in Japan with symptomatic venous thromboembolism: the J-EINSTEIN PE and DVT program // *Curr. Med. Res. Opin.* — 2015. — Apr. 8:1-15.

4. Schulman S., Zondag M., et. al. Recurrent venous thromboembolism in anticoagulated patients with cancer —

management and short-term prognosis // *J Thromb Haemost.* — 2015. — Apr. 7. DOI:10.1111/jth.12955.

5. Messerschmidt C., Friedman RJ. Clinical experience with novel oral anticoagulants for thromboprophylaxis after elective hip and knee arthroplasty // *Arterioscler. Thromb. Vasc. Biol.* — 2015. — Apr. 35 (4):771-8. DOI:10.1161/ATVBAHA.114.303400.

6. Lim W., Meade M., et. al. Failure of anticoagulant thromboprophylaxis: risk factors in medical-surgical critically ill patients // *Crit. Care. Med.* — 2015. — Feb. 43 (2):401-10. DOI:10.1097/CCM.0000000000000713.