

УДК 519.216.3:616.12

А.В. Яковенко

Національний технічний університет України «КПІ», Київ

## АНАЛІЗ МЕТОДІВ ПРОГНОЗУВАННЯ ДЛЯ ВИЯВЛЕННЯ ФАКТОРІВ РИЗИКУ ПІСЛЯОПЕРАЦІЙНИХ УСКЛАДНЕНЬ

*Методами логістичної регресії та дискримінантного аналізу виявлені фактори ризику розвитку гострої серцевої недостатності. Проведено порівняльний аналіз ефективності цих методів. Показано, що на основі їх спільного застосування може бути створена інформаційна система, яка може використовуватися для корекції лікувального процесу.*

**Ключові слова:** бінарна логістична регресія, дискримінантний аналіз, інтелектуальний аналіз, модель прогнозу.

### Вступ

На сьогоднішній день в кардіохірургії існує велика кількість систем оцінки ризику післяопераційного розвитку ускладнень, які досить успішно застосовуються в багатьох центрах світу. Так, однією з них є система EuroScore [1].

Однак, вона заснована на усереднюванні даних різних кардіохірургічних центрів, що пов'язано з похибками прогнозування при оцінці хворих високого ризику конкретної клініки зі своїми індивідуальними особливостями.

Оскільки в умовах кардіохірургічного стаціонару лікувальний процес постійно вдосконалюється, актуальним є створення інформаційних технологій підтримки прийняття рішень, в основі яких лежать моделі оцінки ризику розвитку післяопераційних ускладнень на основі інтелектуального аналізу даних [2].

Для інтелектуального аналізу даних використовується достатньо широкий спектр різних методів. Саме методи математичної статистики виявилися корисними для розвідувального аналізу.

**Метою** роботи є аналіз ефективності застосування методів бінарної логістичної регресії (БЛР) та дискримінантного аналізу (ДА) для індивідуальної оцінки ризику розвитку в ранньому післяопераційному періоді гострої серцевої недостатності (ГСН) при хірургічному лікуванні ішемічної хвороби серця (ІХС) в умовах конкретного кардіохірургічного підрозділу і вивчення можливості створення інформаційних технологій, спрямованих на динамічну корекцію та вдосконалення лікувального процесу.

### Матеріал і методи дослідження

У дослідження були включені 2077 пацієнтів з ІХС, яким було проведено аорто-коронарне шунтування (АКШ) на працюючому серці і з застосуванням штучного кровообігу в 2009–2011 роках. Більшість операцій проводилася на працюючому серці, але 20% хворих екстрено довелося перевести на штучний кровообіг. Всі втручання виконані в Наці-

ональному інституті серцево-судинної хірургії ім. М.М. Амосова НАМН України.

Для виявлення впливу факторів ризику на післяопераційні ускладнення вся вибірка випадковим чином була розділена на дві приблизно рівні групи: навчальну і тестову. На навчальній групі було побудовано прогностичні моделі розвитку ускладнень, а на тестовій проводилася перевірка ефективності роботи моделей. Група 1 включає в себе 1062 хворих, групу 2 складають 1015 хворих, оперованих у 2009–2011 роках.

Ефективність методів БЛР [3] і ДА [4] оцінювали по виявленій структурі факторів ризику розвитку ГСН в ранньому післяопераційному періоді у пацієнтів з ІХС після АКШ.

З метою оцінки впливу факторів ризику на післяопераційні ускладнення у пацієнтів з ІХС використовувалися критерій Пірсона та кореляція Спірмена і Кендалла. Статистична обробка матеріалу проводилася з використанням методів БЛР за допомогою модуля автоматичного включення Binary Logistic та методом ДА за допомогою модуля крокового відбору Discriminant Analysis ПП IBM SPSS Statistics 17.0. В модель включені ознаки, для яких рівень значущості  $p \leq 0,05$ .

Оскільки фактори не є абсолютними – різні медичні установи мають свою структуру, нами була зроблена спроба спільного застосування цих методів для виявлення факторів ризику розвитку ГСН в ранньому післяопераційному періоді після АКШ для корекції їх в умовах конкретного кардіохірургічного підрозділу.

### Результати і обговорення

Для того щоб оцінити ефективність застосування методів БЛР та дискримінантного аналізу нами була зроблена спроба виявити з їх допомогою структуру найбільш вагомих факторів, що впливають на розвиток післяопераційних ускладнень, що дає змогу визначити ймовірність, тобто прогнозувати виникнення конкретних можливих ускладнень під дією факторів ризику.

За допомогою методу БЛР було створено математичні моделі прогнозу розвитку ускладнень за статистично значимими факторами.

Модель, представлена рівнянням БЛР, визначає структуру факторів, що зумовлюють розвиток післяопераційних ускладнень в ранньому післяопераційному періоді у пацієнтів з ІХС після АКШ, має наступний вигляд:

$$p = 1 / (1 + e^{-z}), \quad (1)$$

$$\text{де } z = a_n \cdot x_n + a_{n-1} \cdot x_{n-1} + \dots + a_0; \quad (2)$$

p – вірогідність розвитку післяопераційних усклад-

нень в ранньому післяопераційному періоді; e – експонента, основа натуральних логарифмів; z – показник, що визначає ступінь впливу прогностичних факторів на розвиток післяопераційних ускладнень після АКШ;  $a_{1...n}$  – вагові коефіцієнти рівняння регресії;  $x_{1...n}$  – фактори, що впливають на розвиток післяопераційних ускладнень.

Значення показників що ввійшли в математичну модель  $a_{1...n}$  та  $x_{1...n}$  частоти розподілу, їх статистична значущість та відношення шансів їх впливу на ГСН у досліджуваній групі наведено у табл. 1.

Таблиця 1

Фактори ризику, що ввійшли в модель прогнозу розвитку ГСН у пацієнтів після АКШ в ранньому післяопераційному періоді, визначені методом БЛР (навчальна група, N=1062)

Групи за факторами ризику		Число пацієнтів n		Відношення шансів OR (95%CI)	Коеф. рівняння регресії $a_{1...n}$	Коефіцієнт кореляції з розвитком атеросклерозу	p*
Умов. познач.	Фактори ризику	Не було ГСН	ГСН				
1	2	3	4	5	6	7	8
X <sub>1</sub>	<b>ССП**</b> ІХС** ізольована	746	50	1,0	-0,67	0,065	0,035
	ІХС** +та/або АЛШ**, та/або ІКМП**, та/або аортальний стеноз	239	27	1,7 (1,0–2,8)			
X <sub>2</sub>	<b>Емболія судин ГМ**</b> немає	943	69	1,0	1,23	0,075	0,015
	є	42	8	2,6 (1,2–5,8)			
X <sub>3</sub>	<b>Задихка</b> немає	591	32	1,0	0,61	0,104	0,001
	є при навантаженні	374	40	2,0 (1,2–3,2)			
	є в спокої	20	5	4,6 (1,6–13,1)			
X <sub>4</sub>	<b>Аритмія</b> немає	872	59	1,0	0,79	0,094	0,002
	є	113	18	2,4 (1,3–4,1)			
X <sub>5</sub>	<b>Стенокардія</b> немає	184	28	1,0	-0,85	-0,083	0,007
	напруги	673	39	0,4 (0,2–0,6)			
	напруги/спокою	128	10	0,5 (0,2–1,1)			
X <sub>6</sub>	<b>Характер ІМ**</b> не було ІМ	399	40	1,0	-0,49	-0,056	0,003
	інтрамуральний	113	7	0,6 (0,3–1,4)			
	трансмуральний	473	30	0,6 (0,4–1,0)			
X <sub>7</sub>	<b>Застосування антиагрегантів</b> не застосовувалися	405	33	1,0	0,26	0,033	0,021
	аспірин	449	23	0,6 (0,4–1,1)			
	варфарин	13	2	1,9 (0,4–8,7)			
	плавікс	77	13	2,1 (1,0–4,1)			
	фраксипарин	41	6	1,8 (0,7–4,5)			
X <sub>8</sub>	<b>АК**</b> норма	819	64	1,0	-0,47	0,000	0,049
	недостатність	53	4	1,0 (0,3–2,8)			
	стеноз	113	9	1,0 (0,5–2,1)			
X <sub>9</sub>	<b>Застосування ШК**</b> ні	804	31	1,0	3,03	0,262	0,000
	так	181	46	6,6 (4,1–10,7)			

Закінчення табл. 1

1	2	3	4	5	6	7	8
X <sub>10</sub>	Резекція аневризми не проводилася	849	62	1,0	-0,6	0,042	0,006
	проводилася з накладанням швів	136	15	1,5 (0,8–2,7)			
a <sub>0</sub> = -5,59							

\*) a<sub>0</sub> – вільний член рівняння БЛР

\*\*) позначення: ССП – супутня серцева патологія; ІХС – ішемічна хвороба серця; АЛШ – аневризма лівого шлуночка; ІКМП – ішемічна кардіоміопатія; ГМ – головний мозок; ІМ – інфаркт міокарду; АК – аортальний клапан; ШК – штучний кровообіг.

Математична модель прогнозування імовірності розвитку ГСН в ранньому післяопераційному періоді після АКШ на основі формули (1), має наступний вигляд:

$$p = 1 / (1 + e^{-z}),$$

де

$$z_1 = -x_1 \cdot 0,67 + x_2 \cdot 1,23 + x_3 \cdot 0,61 + x_4 \cdot 0,79 - x_5 \cdot 0,85 - x_6 \cdot 0,49 + x_7 \cdot 0,26 - x_8 \cdot 0,47 + x_9 \cdot 3,03 - x_{10} \cdot 0,6 - 5,59.$$

З даних табл. 1 (стовпчик 5) видно, що найбільші шанси розвитку післяопераційної ГСН в ранньому післяопераційному періоді мають хворі в яких присутня задишка в стані спокою і ті, яким було застосовано екстрений перехід на штучний кровообіг (ШК).

Одним із способів відбору інформативних дискримінантних змінних є покроковий ДА. Відібрані

найбільш інформативні фактори ризику використовуються для розрахунку дискримінантних функцій.

Дискримінантна функція є лінійною комбінацією дискримінантних змінних і має вигляд:

$$Y = b_n \cdot x_n + b_{n-1} \cdot x_{n-1} + \dots + b_0; \quad (2)$$

де Y – показник, що визначає ступінь впливу прогностичних факторів на розвиток післяопераційних ускладнень після АКШ; b<sub>1...n</sub> – вагові коефіцієнти регресії; x<sub>1...n</sub> – фактори, що впливають на розвиток післяопераційних ускладнень.

За допомогою ДА були створені моделі прогнозування ризику розвитку ГСН по статистично значущим факторам.

Значення показників що ввійшли в математичну модель b<sub>1...n</sub> та x<sub>1...n</sub>, їх статистична значущість та значення стандартизованих дискримінантних коефіцієнтів, у досліджуваній групі наведено у табл. 2.

Таблиця 2

Фактори ризику, що ввійшли в модель прогнозу розвитку ГСН у пацієнтів після АКШ в ранньому післяопераційному періоді, визначені методом ДА (навчальна група, N=1062)

Умов. познач.	Фактори ризику	Коеф. ЛДФ** b <sub>1...n</sub>	СДК**	p*
1	2	3	4	5
X <sub>1</sub>	ССП**	-0,51	-0,22	0,035
X <sub>2</sub>	Емболія судин ГМ**	0,90	0,19	0,015
X <sub>3</sub>	Задишка в стані спокою	0,37	0,20	0,001
X <sub>4</sub>	Аритмія	0,67	0,22	0,002
X <sub>5</sub>	Стенокардія	-0,65	-0,37	0,007
X <sub>6</sub>	Трансмуральний ІМ**	-0,25	-0,23	0,003
X <sub>7</sub>	Застосування антиагрегантів	0,17	0,19	0,021
X <sub>8</sub>	Рівень холестерину	-0,08	-0,19	0,010
X <sub>9</sub>	АК**	-0,23	-0,15	0,049
X <sub>10</sub>	Застосування ШК**	3,04	1,20	0,000
X <sub>11</sub>	Резекція аневризми	-0,61	-0,42	0,006
b <sub>0</sub> = -3,06				

\*) b<sub>0</sub> – вільний член ЛДФ

\*\*) позначення: ЛДФ – лінійна дискримінантна функція; СДК – стандартизовані дискримінантні коефіцієнти; ССП – супутня серцева патологія; ГМ – головний мозок; ІМ – інфаркт міокарду; АК – аортальний клапан; ШК – штучний кровообіг.

Математична модель прогнозування імовірності розвитку ГСН в ранньому післяопераційному періоді після АКШ на основі формули (19), має наступний вигляд:

$$Y = -x_1 \cdot 0,51 + x_2 \cdot 0,9 + x_3 \cdot 0,37 + x_4 \cdot 0,67 - x_5 \cdot 0,65 - x_6 \cdot 0,25 + x_7 \cdot 0,17 - x_8 \cdot 0,08 - x_9 \cdot 0,23 + x_{10} \cdot 3,04 - x_{11} \cdot 0,61 - 3,06.$$

За величиною стандартизованих коефіцієнтів (табл. 2, стовпчик 4) видно, що найбільш можливий розвиток ГСН в ранньому післяопераційному періоді у хворих, яким було застосовано екстрений перехід на ШК.

Порівнюючи фактори ризику виявлені методом покрокових ДА та БЛР, бачимо що вони різняться лише одним фактором, виявленим ДА. Тобто ДА є більш чутливим методом і відносить більше ознак до факторів ризику.

Для верифікації виявлених закономірностей була використана перехресна перевірка (cross validation test), тобто метод оцінки моделі і її поведінки на незалежних даних.

Обидві моделі є статистично значимі і обладують достатньо високою прогностичною здатністю. Чутливість моделі БЛР становить 69,2%, специфічність – 93,8%.

При перевірці на тестовій вибірці – 80,0 і 96,9% відповідно.

Чутливість моделі ДА становить 82,2%, специфічність – 96,6%, при загальній перевірці по тестовій вибірці точність прогнозування склала 98,1%.

Так як ДА є більш чутливий та має вищу прогностичну оцінку отриманої моделі, доцільним буде використовувати, в подальшому, саме цей метод для виявлення факторів ризику розвитку післяопераційних ускладнень та побудови інформаційної технології підтримки прийняття рішень.

## Висновки

Проведений аналіз продемонстрував ефективність застосування методів БЛР і ДА для виявлення інформативних ознак, що впливають на розвиток ускладнень в умовах конкретного кардіохірургічного підрозділу.

Отримані моделі, можуть використовуватися для побудови інформаційної технології підтримки прийняття рішень при оцінці ризику розвитку післяопераційних ускладнень.

Така інформаційна технологія інтелектуального аналізу даних може бути корисним інструментом для вдосконалення та корекції лікувального процесу.

## Список літератури

1. Gogbashian A. EuroSCORE: a systematic review of international performance / A. Gogbashian, A. Sedrakyan, T. Treasure // *Eur. J. Cardiothorac. Surg.* – 2004. – Vol. 25. № 5. – P. 695-700.
2. Дюк В. *Data Mining: учебный курс* / В. Дюк, А. Самійленко. – СПб.: Питер, 2001. – 564 с.
3. Григорьев, С.Г. *Логистическая регрессия. Многомерные методы статистического анализа категориальных данных медицинских исследований* / С.Г. Григорьев, В.И. Юнкеров, Н.Б. Клименко // *Уч. пособие. ВМедА.* – СПб, 2001. – С. 10-21.
4. Abdi, H. (2007) "Discriminant correspondence analysis." In: N.J. Salkind (Ed.): *Encyclopedia of Measurement and Statistic. Thousand Oaks (CA): Sage.* – P. 270-275.

Надійшла до редколегії 3.10.2012

**Рецензент:** д-р біол. наук, Є.А. Настенко, Національний технічний університет України «КПІ», Київ.

## АНАЛИЗ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ПО ВЫЯВЛЕНИЮ ФАКТОРОВ РИСКА ПОСЛЕОПЕРАЦИОННЫХ ОСЛОЖНЕНИЙ

А.В. Яковенко

*Методами логистической регрессии и дискриминантного анализа выявлены факторы риска развития острой сердечной недостаточности. Проведен сравнительный анализ эффективности этих методов. Показано, что на основе их совместного применения может быть создана информационная система, которая может использоваться для коррекции лечебного процесса.*

**Ключевые слова:** бинарная логистическая регрессия, дискриминантный анализ, интеллектуальный анализ, модель прогноза.

## ANALYSIS OF THE FORECASTING METHODS FOR THE IDENTIFICATION OF RISK FACTORS FOR POSTOPERATIVE COMPLICATIONS

A. V. Yakovenko

*Risk factors for acute heart failure bypass grafting were identified by logistic regression and discriminant analysis. The comparative analysis of the efficiency of these methods was carried out. It was shown that On the basis of the above methods the information system can be created, which can be used for the correction of the direct care.*

**Key words:** binary logistic regression, discriminant analysis, predictive analysis, forecasting model.