

PRACA ORYGINALNA  
ORIGINAL ARTICLE

# ИСПОЛЬЗОВАНИЕ ДИСКРИМИНАНТНОГО АНАЛИЗА С ЦЕЛЬЮ ПРОГНОЗИРОВАНИЯ РАЗВИТИЯ ТРОМБОФИЛИЙ У БОЛЬНЫХ ХБП VД СТАДИИ

## DISCRIMINANT ANALYSIS FOR PROGNOSIS OF THROMBOPHILIA DEVELOPMENT IN PATIENTS WITH STAGE VD CKD

Алексей Б. Сторожук<sup>1</sup>, Алексей Б. Панасенко<sup>2</sup>, Борис Г. Сторожук<sup>1</sup>, Татьяна Н. Платонова<sup>3</sup>, Лариса А. Сторожук<sup>4</sup>, Эдуард В. Луговской<sup>3</sup>, Ирина Б. Селезнёва<sup>4</sup>, Татьяна В. Довгальук<sup>4</sup>

<sup>1</sup>ВИННИЦКИЙ НАЦИОНАЛЬНЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ ИМ. Н.И. ПИРОГОВА, ВИННИЦА, УКРАИНА

<sup>2</sup>ВИННИЦКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. М.М. КОЦЮБИНСКОГО, ВИННИЦА, УКРАИНА

<sup>3</sup>ИНСТИТУТ БИОХИМИИ ИМ. О.В. ПАЛЛАДИНА НАН УКРАИНЫ, КИЕВ, УКРАИНА

<sup>4</sup>НИИ РЕАБИЛИТАЦИИ ИНВАЛИДОВ ВИННИЦКОГО НАЦИОНАЛЬНОГО МЕДИЦИНСКОГО УНИВЕРСИТЕТА ИМ. Н.И. ПИРОГОВА, ВИННИЦА, УКРАИНА

Oleksii B. Storozhuk<sup>1</sup>, Oleksii B. Panasenko<sup>2</sup>, Borys G. Storozhuk<sup>1</sup>, Tetiana N. Platonova<sup>3</sup>, Larysa A. Storozhuk<sup>4</sup>, Eduard V. Lugovskoy<sup>3</sup>, Iryna B. Seleznyova<sup>4</sup>, Tetiana V. Dovgalyuk<sup>4</sup>

<sup>1</sup>VINNYTSIA NATIONAL PIROGOV MEMORIAL MEDICAL UNIVERSITY, VINNYTSIA, UKRAINE

<sup>2</sup>VINNYTSIA STATE KOTSYUBYNSKI MEMORIAL PEDAGOGICAL UNIVERSITY, VINNYTSIA, UKRAINE

<sup>3</sup>PALLADIN INSTITUTE OF BIOCHEMISTRY, NAS OF UKRAINE, KYIV, UKRAINE

<sup>4</sup>RESEARCH INSTITUTE OF REHABILITATION OF VINNYTSIA NATIONAL PIROGOV MEMORIAL MEDICAL UNIVERSITY, VINNYTSIA, UKRAINE

### РЕЗЮМЕ

**Введение:** Хроническая болезнь почек (ХБП) является значительной медико-социальной проблемой. 7% больных ХБП ежегодно нуждаются в почечнозаместительной терапии, одним из самых угрожающих осложнений которой является нарушение гемостаза с развитием тромбофилических состояний, являющихся составной осложненных коморбидных заболеваний.

Прогнозирование возможного тромбоза у данной категории больных требует комплексного подхода, который включал бы всестороннее лабораторное исследование гемостаза с одной стороны и создание математической модели на базе предикторов тромбофилии – другой.

**Цель:** С помощью дискриминантного анализа и данных биохимических переменных гемостаза разработать модель прогнозирования развития тромбофилии у больных ХБП VД стадии, лечащихся программным гемодиализом.

**Материалы и методы:** Из 12 показателей гемостаза (факторов влияния), которые исследовались у 85 больных ХБП VД стадии, были сформированы две группы переменных - предтромбоза (растворимый фибрин (рФ), Д-димер (Д-д), протеин С (РС)) и посттромбоза (Д-д, Н (максимальная конечная мутность сгустка плазмы) и рС).

**Результаты:** Построенные на базе двух групп переменных линейные дискриминационные функции (две модели) показали достаточно высокую частоту совпадения (92,2–93,5%) распределения по классам. При этом больные, отнесенные к классу с тромботическими осложнениями, 100% соответствовали проспективному (в течение года) исследованию.

**Выводы:** По результатам корреляционного анализа определены наиболее значимые молекулярные биохимические показатели гемостаза (рФ, Д-д, рС, Н). По определенным биохимическим показателям разработаны дискриминантные модели для больных ХБП VД стадии, лечащихся программным гемодиализом, взаимодополняющие друг друга и способные с высокой вероятностью ( $p < 0,0001$ ) прогнозировать тромбофилические состояния в 92,2–93,5% случаев.

**КЛЮЧЕВЫЕ СЛОВА:** дискриминантный анализ, математическое моделирование, гемостаз, тромбофилии, хроническая почечная недостаточность, гемодиализ

### ABSTRACT

**Introduction:** Chronic kidney disease (CKD) is a critical medical and social problem. Every year 7% of such patients require renal replacement therapy associated with one of the most life threatening complications – disturbed hemostasis and development of thrombophilic conditions which are typical of complicated comorbid diseases.

Prediction of potential thrombogenesis in that category of patients requires multidimensional approach, including comprehensive laboratory investigation of hemostasis on the one hand, and creation of mathematical model based on thrombophilia predictors on the other.

**The aim:** To create the prognostic model of thrombophilia development in patients with stage VD CKD, treated by long term hemodialysis, using discriminant function analysis and biochemical variables of hemostasis.

**Materials and methods:** Of 12 hemostatic parameters (factors of influence), studied in 85 patients with stage VD CKD, two groups of variables were formed – pre-thrombosis (soluble fibrin (sF), D-dimer (D-d), protein C (PC)) and post-thrombosis (D-d, H (maximal end turbidity of plasma clot) and pC).

**Results:** Linear discriminant functions (two models), based on two groups of variables, demonstrated rather high coincidence rate (92.2–93.5%) in group distribution. The patients referred to those with thrombotic complications, 100% matched the prospective (for a year) study.

**Conclusions:** By the results of correlation analysis, the most significant molecular biochemical hemostatic parameters (SF, D-d, pC, H) were determined. Using definite biochemical parameters, discriminant models complementing each other and having high prediction ( $p < 0.0001$ ) of thrombophilic states in 92.2–93.5% of cases, were developed for patients with stage VD CKD, treated by long term hemodialysis.

**KEY WORDS:** discriminant analysis, mathematical modeling, hemostasis, thrombophilia, chronic renal failure, hemodialysis

## ВВЕДЕНИЕ

Хроническая болезнь почек (ХБП) VD стадии составляет значительную медицинскую, экономическую и социальную проблему, так как согласно статистическим данным ХБП встречается у 10% взрослого населения, из которых до 7% ежегодно переходят к VD стадии и нуждаются в лечении методами заместительной почечной терапии [1]. При этом темпы роста количества больных с ХБП превышают в пять раз темпы прироста населения [2].

Современное развитие медицинской науки позволило внедрить современные диализные технологии, за счет которых значительно удлинилась продолжительность жизни данной категории больных. Вместе с тем, все более актуальной становится проблема эффективной лабораторной диагностики, которая позволила бы адекватно оценивать течение заболевания и прогнозировать осложнения [3]. Одним из самых угрожающих осложнений программного гемодиализа является нарушение в звеньях системы гемостаза с развитием тромбофилических состояний, сопровождающих основные осложненные коморбидные заболевания [4].

Одним из вариантов прогноза развития тромбофилических состояний является создание математической модели, позволяющей объяснить и выявить взаимосвязи между лабораторными показателями гемостаза и правильно оценить прогноз заболевания.

Тромбофилический синдром как процесс внутрисосудистого тромбогенеза характеризуется системной активацией свертывания крови [4, 6]. Поэтому выявление ранних и доступных биохимических маркеров диагностики данного осложнения призвано предотвратить процесс генерализации тромбообразования за счет своевременной коррекции.

## ЦЕЛЬ РАБОТЫ

Цель работы: с помощью дискриминантного анализа и данных биохимических переменных гемостаза разработать модели прогнозирования развития тромбофилий у больных ХБП VD стадии.

## МАТЕРИАЛЫ И МЕТОДЫ

Создание математической модели, в которой прогностический показатель определен как тромбоз (да - 1) или

его отсутствие (нет - 0) (переменные Label), построено на базе совокупности основных биохимических показателей (растворимый фибрин, Д-димер, протеин С, протромбиновый и екамулиновый индекс, фибриноген, креатинин, гемоглобин, оптическая плотность плазмы, общий гемостатический потенциал, растущий потенциал и фибринолитический потенциал) с целью по данным биохимических переменных гемостаза предсказать значение переменной Label. Из указанных показателей были определены те, что имели связь с развитием проспективно доведенного прогнозируемого осложнения (1 год наблюдения).

На первом этапе в качестве возможных факторов влияния у 85 больных с ХБП VD стадии, лечащихся программным гемодиализом, определялись все 12 показателей. Затем на основе корреляционного анализа (первичное сжатие исходной информации) были определены те, что имели наибольшую силу по значимости связей между собой. В результате для создания модели выбраны пары признаков с достаточной информационной способностью. Для построения математической модели использован аппарат дискриминантного анализа, наиболее адекватно отвечающий поставленной нами задаче и позволяющий распознавать объекты наблюдения по разделяющим (дискриминирующим) переменным (многомерный анализ). Для вычислений использовалась программа SPSS, а математическая обработка результатов проводилась с использованием программ Microsoft Excel 7.0. [5].

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Полученные с помощью дискриминантного анализа классификационные линейные функции позволяют решить задачу отнесения больных к тому или иному классу. Используя матрицу результатов корреляционного анализа в различных группах больных, определено, что наиболее значимые связи наблюдаются между молекулярными биохимическими показателями - растворимым фибрином (рФ) и Д-димером (Д-д) ( $r=0,56 \div 0,70$ ), а также между рФ и протеином С (рС) ( $r=-0,47$ ).

Указанную группу биохимических маркеров возможно эмпирически отнести к компоненте предтромбоза, тогда как другую группу, которую составляют Д-д, рС и Н (максимальная конечная мутность сгустка плазмы), - к компоненте посттромбоза, поскольку показатель Н имел достаточно высокую корреляционную связь с общим гемостатическим потенциалом ( $r=0,86 \div 0,97$ ).

**Таблица I.** Критерий М Бокса проверки нулевой гипотезы равенства ковариационных матриц.

№ п/п	Label	Ранг матрицы	Натуральный ln определителя	Результат теста	Значение F-критерия	Ст. св. 1*	Ст. св. 2*	Значимость P
1	«0»	3	13,109					
2	«1»	3	13,915	4,262	0,610	6	1083,01	0,722
3	Вместе	3	13,207					

Примечание: \* - степени свободы.

**Таблица II.** Критерии равенства групповых средних (сгенерировано SPSS).

№ п/п	Показатели гемостаза	Лямбда Уилкса	Значение F-критерия	Ст. св. 1*	Ст. св. 2*	Значимость P
1	pФ	0,857	14,037	1	84	0,0001
2	Д-д	0,552	68,073	1	84	0,0001
3	pC	0,962	3,319	1	84	0,072

Примечание: \* - степени свободы.

Примем первую комбинацию показателей (pФ, Д-д, pC) и проведем классификацию по классам, а для проверки равенства ковариационных матриц применим критерий М Бокса (сравнение дисперсий различных групп - автокорреляция), результаты которого свидетельствуют о том, что «0» гипотеза о равенство матриц не отклоняется (значимость = 0,722) (табл. I).

Кроме того, используя одномерный дисперсный анализ, проверим гипотезу о равенстве средних значений показателей по каждой из групп (табл. II).

Из приведенных данных табл. II видно, что гипотеза о равенстве групповых средних на уровне значимости 0,1 отклоняется для каждой из переменных, что дает основания надеяться на возможность эффективного распределения классов.

Дискриминантный анализ, проведенный в системе SPSS, обеспечил определение линейных дискриминантных функций (ЛДФ) Фишера:

$$d_0 = (1,159 \times pФ) + (0,053 \times Д-д) + (0,348 \times pC) - 18,621$$

$$d_1 = (1,567 \times pФ) + (0,143 \times Д-д) + (0,312 \times pC) - 26,881$$

Необходимо отметить, что при этом были использованы одинаковые априорные вероятности принадлежности объекта исследования к каждому из классов. В дальнейшем по правилу классификации для заданных показателей вычисляются значения  $d_0$  и  $d_1$ , и если  $d_0 > d_1$ , то объект относится к классу «0», а в другом случае – к классу «1». Результаты классификации подтвердили отсутствие ошибок первого рода (все пациенты класса «1» классифицированы правильно). При этом, 93,5% пациентов классифицированы правильно к классу «1», что свидетельствует о невысоком уровне ошибки второго рода. Проведенная кросс-проверка не изменила результаты классификации.

Для заданных показателей определялись также коэффициенты (ненормированные) канонической дискриминантной функции (КДФ), каноническая корреляция (коэффициент корреляции Пирсона) между значениями КДФ и классами и значение КДФ в центроидах классов (табл. III).

В таблице III λ Уилкса указывает, насколько хорошо функция разделяет выборку на классы, а статистика  $\chi^2$  в данном случае проверяет нулевую гипотезу о том, что эта гипотеза отвергается. Полученные значения КДФ в центроидах классов еще раз доказывают значительные различия в группах.

Проверим на примерах совпадение прогностического результата ЛДФ с данными проспективного исследования пациентов (один год наблюдения).

Пример 1. Пациент Б., 45 лет, на программном гемодиализе 11 лет, тромботических осложнений в течение года наблюдения не зафиксировано. Показатели гемостаза: pФ – 4,6 мкг/мл, Д-д – 76 нг/мл, pC – 64%.

$$ЛДФ_0 = (1,159 \times 4,6) + (0,053 \times 76) + (0,348 \times pC) - 18,621 = 13,01$$

$$ЛДФ_1 = (1,567 \times 4,6) + (0,143 \times 76) + (0,312 \times 64) - 26,881 = 11,16$$

Пример 2. Пациент Г., 53 года, на программном гемодиализе 2 года, тромбоз шунта в течение первого месяца наблюдения. Показатели гемостаза: pФ – 5,3 мкг/мл, Д-д – 211 нг/мл, pC – 39%.

$$ЛДФ_0 = (1,159 \times 5,3) + (0,053 \times 211) + (0,348 \times 39) - 18,621 = 12,28$$

$$ЛДФ_1 = (1,567 \times 5,3) + (0,143 \times 211) + (0,312 \times 39) - 26,881 = 23,77$$

Вывод: у пациента Г., где  $d_0 < d_1$ , что относит его к «1» классу и прогностическая вероятность развития тромботических осложнений достаточно высокая, что подтверждено результатами проспективного исследования.

Примем определенную нами вторую комбинацию показателей (Д-д, Н, pC).

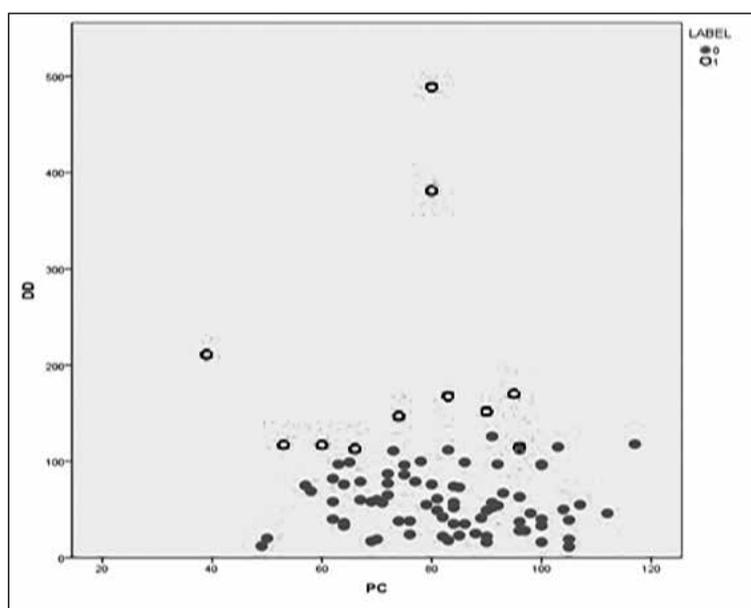
Отметим, что определение переменных, которые лучше разделяют выборку на два класса по отношению к значениям Label, проведено по высшему коэффициенту корреляции с Label. Первой такой переменной стал Д-д, но, как выяснилось, в системе координат Д-д – Н происходит улучшение расположения выборок. Вместе с тем, и в этой системе координат можно выделить переменные, принадлежащие к разным

**Таблица III.** Коэффициенты КДФ и показатели разделения выборки на классы с помощью КДФ.

№ п/п	Показатели гемостаза	Коэффициент КДФ	$\lambda$ Уилкса	$\chi^2$	Коэффициент корреляции Пирсона	Ст. св.	Р	Значение КДФ в центроидах
1	рФ	0,134						
2	Д-д	0,03	0,531	52,208	0,685	3	0,0001	«0» - 0,317 «1» - 2,716
3	рС	-0,012						
4	Константа	-1,523						

**Таблица IV.** Показатели КДФ и показатели разделения второй выборки на классы с помощью КДФ.

№ п/п	Показатели гемостаза	Коэффициент КДФ	$\lambda$ Уилкса	$\chi^2$	Коэффициент корреляции Пирсона	Ст. св.	Р	Значение КДФ в центроидах
1	Д-д	0,032						
2	Н	1,600	0,535	57,597	0,682	3	0,0001	«0» - 0,315 «1» - 2,695
3	рС	-0,010						
4	Константа	-1,927						

**Рис. 1.** Разделение пациентов по классам в системе координат рС – Д-д.

классам, но имеющие близкие показатели. Последнее побудило провести поиск третьей характеристики, которая окончательно отделила бы классы больных. Такой характеристикой оказался протеин С при слабой корреляции с переменными Д-д и Н (рис. 1).

Таким образом, можем выдвинуть гипотезу, что по данным переменных Д-д, Н, рС возможно классифицировать данные. Затем, как и в случае с первой комбинацией переменных, для проведения классификации применим дискриминантный анализ.

Как показывает корреляционная матрица, выбранные переменные не имеют сильных корреляционных связей, что соответствует одному из ограничений на исследуемые данные при применении дискриминантного анализа. При этом, также важным условием является равенство ковариационных матриц для различных классов и различие средних значений переменных в каждом классе.

Проверку равенства ковариационных матриц по разным классам, как и в первой комбинации, проведем по критерию М Бокса, результаты которого показали, что нулевая гипотеза о равенстве ковариационных матриц не отклоняется (значимость = 0,518), а гипотеза о равенстве групповых средних на уровне значимости 0,1 отклоняется для каждой из переменных. Как и ожидалось, средние в разных классах наиболее отличаются для предикторной переменной Д-д.

Дискриминантный анализ, проведенный в системе SPSS, определил следующие линейные дискриминационные функции Фишера:

$$d_0 = (0,065 \times \text{Д-д}) + (41,591 \times \text{Н}) + (0,407 \times \text{рС}) - 27,249$$

$$d_1 = (0,160 \times \text{Д-д}) + (46,406 \times \text{Н}) + (0,377 \times \text{рС}) - 36,629$$

При вычислении значений  $d_0$  та  $d_1$  для заданных показателей (классификация), определяем отсутствие ошибки первого рода, и верно классифицированных 93,5% пациентов класса «0» (проведенная кросс-проверка показала результат 92,2%).

Как свидетельствуют данные таблицы IV, ненормированный коэффициент канонической корреляции между значениями КДФ и классами достаточно весомый, а показатели функции разделения выборки на классы ( $\lambda$  Уилкса,  $\chi^2$ ) показали эффективность их применения.

Полученные данные указывают на высокую дискриминационную способность функции.

Как и для первой группы переменных, проведено совпадение прогностического результата ЛДФ данным проспективного исследования, которое показало высокую вероятность развития тромботических осложнений у пациента Г.,  $d_0 < d_1$ , ( $d_0 = 25,20$ ;  $d_1 = 41,12$ ).

## ВЫВОДЫ

1. По результатам корреляционного анализа определены наиболее значимые молекулярные биохимические показатели гемостаза (рФ, Д-д, рС, Н).
2. По определенным биохимическим показателям разработаны дискриминантные модели для больных ХБП VD стадии, лечащихся программным гемодиализом, взаимодополняющие друг друга и способные с высокой вероятностью ( $p < 0,0001$ ) прогнозировать тромбофилических состояния в 92,2-93,5% случаев. Дальнейшее внедрение более эффективных лабораторно-биохимических маркеров тромбоза и использования доступных математических моделей позволит улучшить диагностику, прогнозирование осложнений и оптимизировать процессы лечения и реабилитации данной категории больных.

## REFERENCES

1. Belov V. V., Il'icheva O. V. Dinamika pokazateley KHSN u bol'nykh khronicheskoy bolezn'yu pochek pri pochechno-zamestitel'noy terapii [Nefrologiya. Dynamics of CHF indices in patients with chronic kidney disease in renal replacement therapy]. Nephrology. 2006;10(1):35-39.
2. Dudar I. O., Driyans'ka V. YE., Savchuk V. M., Loboda O. M., Honchar YU. I., Shyfris I. M., Krot V. F., Krasnyuk E. K. Khronichne zapalennya ta anemiya u khvorykh na KHKHNVD st. [Chronic inflammation and Anemia in patients with CKD VD st.]. Ukrainian Journal of Nephrology and Dialysis. 2017;55(3):84-85.
3. Volkov G. L., Platonova T. N., Savchuk A. M., Gornitskaya O. V., Chernyshenko T. M., Krasnobryzha Ye. M. Sovremennyye predstavleniya o sisteme gemostaza [Contemporary ideas about the hemostatic system]. Kiev: Naukova Dumka; 2005.
4. Shyfris I. M., Dudar I. O. Komorbidnist' ta vyzhyvannya khvorykh na khronichnu khvorobu nyrok VD stadiyi [Comorbidity and survival of patients with chronic kidney disease in the VD stage]. Ukrainian Journal of Nephrology and Dialysis. 2015;4(48):30-39.
5. Byul' A., Tsofel' P. SPSS: Iskusstvo obrabotki informatsii. Analiz statisticheskikh dannykh i vosstanovleniye skrytykh zakonomernostey [SPSS: The Art of Information Processing. Analysis of statistical data and the restoration of hidden patterns]. SPb.: DiaSoft Yu.P. 2005.

*Исследование проведено в рамках научно-исследовательской работы «Обосновать научные подходы к определению составляющих реабилитационного потенциала у больных хронической болезнью почек I-V Д, Т», государственный регистрационный номер 0116U001421.*

### Вклад авторов:

*В порядке очередности авторства*

### Конфликт интересов:

*Авторы не заявляют о конфликте интересов*

## АВТОР ДЛЯ КОРРЕСПОНДЕНЦИИ

**Татьяна В. Довгалюк**

просп. Космонавтов 26, кв. 22, 21021, Винница, Украина

тел: +380500636522

e-mail: postboxt05@gmail.com

**Прислана:** 30.07.2018

**Утверждена:** 15.12.2018