

---

## ПРОГНОЗИРОВАНИЕ РАЗВИТИЯ РАКА ПРЕДСТАТЕЛЬНОЙ ЖЕЛЕЗЫ С ПОМОЩЬЮ ОБЩИХ МОДЕЛЕЙ ДИСКРИМИНАНТНОГО АНАЛИЗА

**А.А. Костин, А.Т. Асратов**

Кафедра урологии, онкологии, радиологии ФПК  
Медицинский институт  
Российский университет дружбы народов  
*ул. Миклухо-Маклая, 6, Москва, Россия, 117198*

**Н.Г. Кульченко**

Кафедра гистологии, цитологии и эмбриологии  
Медицинский институт  
Российский университет дружбы народов  
*ул. Миклухо-Маклая, 6, Москва, Россия, 117198*

**А.О. Толкачев**

Московский научно-исследовательский  
онкологический институт им. П.А. Герцена  
филиал ФГБУ «НМИРЦ» Минздрава России  
*2-й Боткинский проезд, 3, Москва, Россия, 125284*

В данном исследовании представлен ретроспективный анализ факторов риска рака предстательной железы у 122 мужчин. Сначала мы проводили сравнительный анализ множества факторов: возраст, вес, вредные привычки, уровень тестостерона, уровень простат-специфического антигена (ПСА), сопутствующая патология, группа крови, резус-фактор, пальцевого ректального исследования, результаты ультразвукового исследования предстательной железы. Далее с помощью общих моделей дискриминантного анализа мы выявили пять наиболее значимых факторов риска рака простаты: воспалительные заболевания почек и мочевого пузыря, повышение уровня ПСА крови, группа крови, избыточный вес больного.

Примененное дискриминантное уравнение, содержащее пять наиболее информативных признаков, позволяет с точностью до 97,5% прогнозировать вероятность развития рака предстательной железы.

**Ключевые слова:** рак предстательной железы, доброкачественная гиперплазия предстательной железы, факторы риска рака предстательной железы, простат-специфический антиген.

**Введение.** Заболеваемость онкологическими заболеваниями мужского населения в Европе на 2012 г. составила 1,8 млн (53%), из них рак предстательной железы (РПЖ) — 417 000 (22,8%) [9]. Таким образом, после рака толстой кишки РПЖ является наиболее распространенной формой рака у европейских мужчин.

Наиболее высокий уровень заболеваемости РПЖ выявлен в Северной и Западной Европе, таких странах, как Норвегия (193) и Франции (187), а самый низкий — в странах Центральной и Восточной Европы — Республика Молдова (30) и Албания (25) [8; 10].

В Российской Федерации у мужчин в 2012 г. впервые выявлено 240 938 (45,8%) случаев злокачественных новообразований, причем заболеваемость РПЖ составила 93,7 на 100 000 населения [3]. По величине прироста показателя заболеваемости в России РПЖ занимает 1-е место за период 2000—2010 гг. (136,86%) [1; 7].

Летальность в течение года с момента установления диагноза, причиной которой явился РПЖ, на протяжении последних 5 лет в Российской Федерации неуклонно снижается, и в 2012 г. составила 10,4%, что в 1,5 раза меньше чем в 2007 г., и в два раза ниже чем в 2002 г. [3; 7]. В Европе данный показатель за 2012 г. составил 92 000 (9,5%) человек [8].

Таким образом, высокая смертность от РПЖ связана не только с преобладанием распространенных форм в структуре заболеваемости и недостаточной диагностикой ранних форм опухолевого процесса, но и с отсутствием способов первичной диагностики и профилактики данного онкологического процесса.

**Цель исследования** — улучшить результаты диагностики рака предстательной железы за счет выявления корреляционных связей между основными факторами риска и точностью дифференциальной диагностики между РПЖ и доброкачественной гиперплазией предстательной железы.

**Материалы и методы.** Нами был проведен ретроспективный анализ 122 пациентов, находившихся на обследовании по поводу РПЖ. Все пациенты были разделены на две группы: основная — 59 (48,3%) мужчин с верифицированным РПЖ и контрольная — 63 (51,7%) больных с доброкачественной гиперплазией предстательной железы (ДГПЖ).

Сначала мы проводили сравнительный анализ максимального количества факторов риска РПЖ с помощью многофакторного анализа. При этом мы учитывали следующие критерии.

1. Возраст.
2. Характер жалоб (резкое похудание, болезненное мочеиспускание, отвращение к пище, постоянная мышечная слабость, боль в промежности и над лобком, IPSS, QoL).
3. Антропометрические параметры (вес, рост).
4. Анамнез (травмы и операции на органах мочеполовой системы, вредные привычки, прием алкоголя, производственные вредности).
5. Сопутствующая патология (сахарный диабет, гипертоническая болезнь, ИБС, атеросклероз, вирус простого герпеса, анемия, простатит, инфекционные заболевания почек, мочевого пузыря, мошонки).
6. Группа крови и резус-фактор.
7. Уровень простат-специфического антигена (ПСА).
8. Пальцевое ректальное исследование (ПРИ).
9. Признаки возрастного гипогонадизма (наличие волос на теле, наличие волос на голове, определение уровня общего тестостерона).
10. Ультразвуковое исследование (объем простаты, объем остаточной мочи, гипер- и гипозоногенные включения в предстательной железе).

Многофакторный анализ мы проводили с помощью следующей модели: задавалась система переменных  $X_1, X_2, \dots, X_n$  [2; 5]. Например,  $X_1$  — возраст,  $X_2$  — вес больного,  $X_n$  — уровень ПСА. Мы представляли исходную информацию в виде матрицы  $X = x_{ji}$  размерности  $(n \times N)$ . Каждая строка состояла из значений одного показателя для каждого из объектов исследования ( $N$ ), и предполагали, что каждый

элемент этой матрицы  $x_{ji}$  являлся результатом воздействия некоторого числа  $m$ , гипотетических общих факторов и одного характерного фактора:

$$x_{ji} = a_{j1}f_{1i} + a_{j2}f_{2i} + \dots + a_{jr}f_{ri} + \dots + a_{jm}f_{mi} + d_jv_{ji}$$

где  $a_{jr}$  — весовой коэффициент  $j$ -й переменной на  $r$ -м общем факторе или нагрузка  $j$ -й переменной на  $r$ -м общем факторе;  $f_{ri}$  — значение  $r$ -го общего фактора на  $i$ -м объекте исследования;  $d_j$  — нагрузка, или весовой коэффициент  $j$ -й переменной на  $j$ -м характерном факторе;  $v_{ji}$  — значение  $j$ -го характерного фактора на  $i$ -м объекте исследования;  $j = 1, \dots, n$ ;  $i = 1, \dots, N$ ;  $r = 1, \dots, m$ ;  $m \ll n$ .

Так как массив данных  $X = x_{ji}$  представляет величины различной размерности, то для того чтобы перейти к безразмерным величинам, мы прогнозировали элементы массива:

$$Y_{ji} = (x_{ji} + x_j)/S_j,$$

где  $x_j$  — выборочное среднее  $j$ -й переменной (признака);  $S_j$  — выборочная дисперсия  $j$ -й переменной.

После этих преобразований получим:

$$Y_{ji} = a_{j1}v_{1i} + a_{j2}v_{2i} + \dots + a_{jr}v_{ri} + \dots + a_{jm}v_{mi} + d_jv_{ji},$$

где  $a_{jm}$  — неизвестные коэффициенты (3.3), называемые факторными нагрузками;  $d_jv_{ji}$  — остаточный специфический фактор.

Далее мы выбирали наиболее значимые факторы риска развития РПЖ с помощью общих моделей дискриминантного анализа (Generaldiscriminant Analysis Models; GDA) [4; 5]. GDA осуществляли по методу «обучающей выборки». Сначала из 59 больных РПЖ и 63 больных ДГПЖ случайным образом (с использованием датчика случайных чисел) было отобрано 39 больных с РПЖ и 43 больных с ДГПЖ и на этом массиве данных строилось дискриминантное уравнение классифицирующее больных, а затем правильность полученных уравнений проверялась на массиве из 20 больных РПЖ и 20 больных ДГПЖ. Было получено дискриминантное уравнение, содержащее пять наиболее информативных признаков, которые позволяют с 100% точностью разделить «обучающую выборку» на два класса: первый класс — больные с РПЖ, второй класс — больные с аденомой. Программа GDA позволяет производить оценки параметров, стандартизированные оценки параметров, их стандартные ошибки, уровни значимости ( $p$ ) и соответствующие статистики. Уровни значимости показывают, по значениям каких переменных существенно различаются наблюдения из разных классов, оценки параметров — силу и характер вклада переменных в определение принадлежности наблюдений к тому или иному классу.

**Результаты и их обсуждение.** В группе больных ДГПЖ были несколько старше (средний возраст 68,6 против 61,1;  $p = 0,0011$ ), среди них больше больных с вредными привычками (50 против 33, соответственно 79,4% против 55,9%,  $p = 0,0053$ ). Следовательно, такие показатели, как возраст, вредные привычки, не влияют на развитие РПЖ.

При проведении факторного анализа было выделено семь факторов, описывающих 59,1% всей дисперсии исследуемых данных. В таблице 1 для лучшего анализа различия групп были построены факторы отдельно по каждой группе анализируемых больных.

Таблица 1

## Корреляция факторов в группах

Больные РПЖ		Больные аденомой простаты	
Фактор и признаки	корреляция	Факторы и признаки	корреляция
<b>Фактор 1</b>		<b>Фактор 1</b>	
Вес	0,616	Вес	0,696
Наличие волос на теле	0,654	Наличие волос на теле	-0,802
Наличие волос на голове	-0,453	Наличие волос на голове	0,792
Вирус простого герпеса	0,832	Вирус простого герпеса	0,617
Анемия	0,774	Анемия	-0,864
Пальпируемая опухоль	-0,896	Пальпируемая опухоль	-0,911
Боль в промежности и над лобком	0,896	Боль в промежности и над лобком	-0,912
Болезненность простаты при ректальном исследовании	-0,797	Болезненность простаты при ректальном исследовании	-0,818
Инфекционные заболевания мочевого пузыря	0,879	Инфекционные заболевания мочевого пузыря	-0,707
Травмы и операции на органах мочеполовой системы	-0,586	Травмы и операции на органах мочеполовой системы	-0,637
Дисперсия	9,09	Дисперсия	6,30
Процент объясняемой дисперсии	25,2	Процент объясняемой дисперсии	17,7
<b>Фактор 2</b>		<b>Фактор 2</b>	
Болезненное мочеиспускание	-0,709	Болезненное мочеиспускание	-0,488
Отвращение к пище	-0,984	Отвращение к пище	0,879
Отрицательный резус-фактор	-0,984	Отрицательный резус-фактор	0,748
Дисперсия	4,96	Дисперсия	4,37
Процент объясняемой дисперсии	16,5	Процент объясняемой дисперсии	12,8
<b>Фактор 3</b>		<b>Фактор 3</b>	
Резкое похудание	0,538	Резкое похудание	0,591
Болезненное мочеиспускание	0,617	Болезненное мочеиспускание	0,615
Индекс IPSS	0,839	Индекс IPSS	0,827
Индекс QoL	0,912	Индекс QoL	0,819
Инфекционные заболевания почек	-0,878	Инфекционные заболевания почек	-0,568
Дисперсия	5,27	Дисперсия	3,83
Процент объясненной дисперсии	13,2	Процент объясненной дисперсии	11,9
<b>Фактор 4</b>		<b>Фактор 4</b>	
Концентрация ПСА	0,775	Концентрация ПСА	0,513
Объем остаточной мочи	0,751	Объем остаточной мочи	0,908
Группа крови	0,820	Группа крови	0,506
Дисперсия	5,82	Дисперсия	4,42
Процент объясненной дисперсии	11,6	Процент объясненной дисперсии	11,3
<b>Фактор 5</b>		<b>Фактор 5</b>	
Вредные производства	0,454	Вредные производства	0,475
Постоянная мышечная слабость	-0,876	Постоянная мышечная слабость	0,771
Дисперсия	3,44	Дисперсия	3,86
Процент объясненной дисперсии	7,5	Процент объясненной дисперсии	9,9

Больные РПЖ		Больные аденомой простаты	
Фактор и признаки	корреляция	Факторы и признаки	корреляция
<b>Фактор 6</b>		<b>Фактор 6</b>	
Объем простаты при УЗИ	0,648	Объем простаты при УЗИ	0,856
Камни в предстательной железе	-0,464	Камни в предстательной железе	0,586
Гипер- и гипозохогенные включения	-0,732	Гипер- и гипозохогенные включения	0,792
Дисперсия	3,11	Дисперсия	3,53
Процент объясненной дисперсии	7,2	Процент объясненной дисперсии	7,6
<b>Фактор 7</b>		<b>Фактор 7</b>	
Прием алкоголя	0,745	Прием алкоголя	-0,675
Уровень тестостерона	-0,709	Уровень тестостерона	0,636
Дисперсия	4,55	Дисперсия	4,29
Процент объясненной дисперсии	5,9	Процент объясненной дисперсии	5,2

Таким образом, с помощью проведенного факторного анализа рисков развития заболевания было объяснено дисперсии в группе пациентов с РПЖ 86,6%, в группе ДГПЖ — 76,5%.

При анализе результатов табл. 1 обращают на себя внимание следующие различия между группами:

а) в группе больных РПЖ увеличивается жесткость связей признаков с факторами, говорящая о более глубокой перестройке организма под воздействием РПЖ;

б) нарушается качество корреляционных связей с факторами за счет изменения пропорциональности. Например, в 1-м факторе при РПЖ анемия имеет корреляцию с фактором 0,774, а при ДГПЖ -0,847; боли в промежности и над лобком 0,896 и -0,912, соответственно.

Как показывают результаты многофакторного анализа, все параметры являются определяющими признаками для разделения больных на группы. При этом переменные для группы пациентов с РПЖ имеют преимущественно большие (характер имеет положительный знак) значения по сравнению с ДГПЖ.

С помощью общих моделей дискриминантного анализа мы из 7 групп факторов риска выявили 5 наиболее важных признаков риска РПЖ, представленных в табл. 2. При этом константа составила 33,6.

Таблица 2

**Информативность признаков, включенных в дискриминантное уравнение прогнозирования развития РПЖ**

Симптомы (признаки) ( $X_i$ )	Вес признака ( $W_i$ )	Информативность по F-критерию Фишера
Инфекционные заболевания почек	14,0	109,6
Уровень ПСА	0,32	92,5
Группа крови	-4,4	79,3
Вес больного	0,39	77,3
Инфекционные заболевания мочевого пузыря	-8,6	76,4

Мы установили: если выполняются условия  $G(x) = \sum Wi / X_i > 33,6$ , то такому больному прогнозируется развитие РПЖ. Если данное условие не выполняется, то прогнозируется развитие ДГПЖ. В таблице 3 отражено отношение параметров для ДГПЖ и РПЖ к признакам.

Таблица 3

Оценки параметров, входящих в дискриминантное уравнение

Признак	Знак параметра для РПЖ	Доверительный интервал	$p$	Знак параметра для ДГПЖ	Доверительный интервал	$p$
Вес пациента	0,0145	6,98	0,0000001	-0,0145	6,97	0,000001
Уровень ПСА	0,0123	9,58	0,000001	-0,0123	9,58	0,000001
Инфекционные заболевания почек	-0,275	8,06	0,000001	0,275	8,06	0,000001
Инфекционные заболевания мочевого пузыря	0,227	6,47	0,000001	-0,227	6,47	0,000001
Группа крови	-0,201	2,37	0,201	0,201	2,37	0,0205

Из табл. 3 видно, что такие признаки, как инфекционные заболевания почек и группа крови, имеют меньшее (параметр имеет отрицательный знак) значение у больных с РПЖ, чем по сравнению с ДГПЖ, и являются менее значимыми, чем уровень ПСА, группа крови и вес пациента.

Следовательно, применение дискриминантного анализа позволило с достаточной точностью (97,5%) и эффективностью 99,2% классифицировать больных и сформировать при скрининге РПЖ группу риска.

**Выводы.** Примененное дискриминантное уравнение, содержащее пять наиболее информативных признаков, позволяет с точностью до 97,5% прогнозировать вероятность развития рака предстательной железы.

Наиболее значимыми факторами в определении развития рака предстательной железы являются воспалительные заболевания почек и мочевого пузыря, повышение уровня ПСА крови, группа крови, малоподвижный образ жизни, повышенная масса тела.

## ЛИТЕРАТУРА

- [1] Воробьев А.В. Рак предстательной железы: эволюция взглядов // Вопросы онкологии. 2009. Т. 55. С. 241—250.
- [2] Дубров А.М., Мхитарян В.С., Троцкин Л.И. Многомерные статистические методы: Учебник. М.: Финансы и статистика, 2000.
- [3] Каприн А.Д., Старинский В.В., Петрова Г.В. Состояние онкологической помощи населению России в 2012 году. М.: МНИОИ им. П.А. Герцена, 2013.
- [4] Таганов Д.Н. Статистический анализ в маркетинговых исследованиях. СПб.: Питер, 2005. С. 132—141.
- [5] Халафян А.А. STATISTICA 6. Статистический анализ данных. 3-е издание. Учебник. М.: ООО «Бином-Пресс», 2008.
- [6] Чиссов В.И., Старинский В.В., Петрова Г.В. Злокачественные новообразования в России в 2011 году (заболеваемость и смертность). М.: ФГБУ «МНИОИ им. П.А. Герцена» Минздрава России, 2013.

- [7] Ferlay J., Steliarova-Foucher E., Lortet-Tieulent J. et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012 // *Eur J Cancer*. 2013 Apr;49(6):1374—403.
- [8] Roobol M.J., Schröder F.H., Hugosson J. et al. Importance of prostate volume in the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk calculators: results from the prostate biopsy collaborative group // *World J Urol*. 2012 Apr;30(2):149—55.
- [9] Schroder F.H., Hugosson J., Roobol M.J. et al. Screening and prostate-cancer mortality in a randomized European study // *N. Engl. J. Med*. 2009;360:1320—1328.

## PREDICTION OF PROSTATE CANCER USING GENERAL DISCRIMINANT ANALYSIS MODELS

A.A. Kostin<sup>1,3</sup>, A.T. Asratov<sup>1</sup>,  
N.G. Kulchenko<sup>2</sup>, A.O. Tolkachev<sup>3</sup>

<sup>1</sup>Department of urology, oncology and radiology

<sup>2</sup>Department of histology, cytology, embryology  
Medical Institute of People's Friendship University  
*Miklukho-Maklaya str., 8, Moscow, Russia, 117198*

<sup>3</sup>P.A. Herzen Moscow Oncology Research Institute  
*2<sup>nd</sup> Botkinskiy passway, 3, Moscow, Russia, 125284*

The study presents a retrospective analysis of risk factors for prostate cancer in 122 men. First, a comparative analysis of many factors: age, weight, bad habits, levels of testosterone and of prostate-specific antigen (PSA), comorbidities, blood group, RH factor, manual rectal investigation, the results of the prostate ultrasound was conducted. Next, using general models of discriminant analysis, five of the most significant risk factors for prostate cancer were identified: inflammatory diseases of the kidneys and bladder, increased PSA blood test, blood group, overweight patient.

Applied discriminant equation containing the five most informative features, allows up to 97.5% to predict the likelihood of developing prostate cancer.

**Key words:** prostate cancer, benign prostatic hyperplasia, the risk factors for prostate cancer, prostate-specific antigen.

### REFERENCES

- [1] Vorob'iev A.V. Prostate cancer: evolution of views. *Questions of Oncology*. 2009. Vol. 55. P. 241—250.
- [2] Dubrov A.M., Mkhitarayan V.S., Troschin L.I. Multivariate statistical techniques. Textbook. Moscow: Finance and Statistics, 2000.
- [3] Kaprin A.D., Starinskiy V.V., Petrova G.V. Status of cancer care to the population of Russia in 2012. Moscow: MORI n.a. P.A. Herzen, 2013.
- [4] Taganov D.N. Statistical analysis in marketing research. St. Petersburg: Piter, 2005. P. 132—141.
- [5] Khalafyan A.A. STATISTICA 6. Statistical analysis of data: textbook, 3rd ed. M.: OLS "Bean-Press", 2008.
- [6] Chissov V.I., Starinskiy V.V., Petrova G.V. Malignancies in Russia in 2011 (morbidity and mortality). Moscow: FGBI MORI n.a. P.A. Herzen, 2013.

- [7] *Ferlay J., Steliarova-Foucher E., Lortet-Tieulent J. et al.* Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013 Apr; 49 (6): 1374—403.
- [8] *Roobol M.J., Schröder F.H., Hugosson J. et al.* Importance of prostate volume in the European Randomised Study of Screening for Prostate Cancer (ERSPC) risk calculators: results from the prostate biopsy collaborative group. *World J Urol*. 2012 Apr; 30 (2): 149—55.
- [9] *Schroder F.H., Hugosson J., Roobol M.J. et al.* Screening and prostate-cancer mortality in a randomized European study. *N. Engl. J. Med.* 2009; 360: 1320—1328.