

ЛИТЕРАТУРА

1. Буянова С. Н., Щукина Н. А., Добровольская Т. Б. // Российский вестник акушера-гинеколога. — 2004. — Т. 4, № 5. — С. 30—33.
2. Гзззян А. М. Роль аутоиммунного оофорита в развитии гипергонадотропной и нормогонадотропной недостаточности яичников: автореф. дис. ... канд. мед. наук. — СПб, 1995. — 21 с.
3. Козлова В. И., Пухнер А. Ф. Вирусные, хламидийные и микоплазменные заболевания гениталий. — М.: Триада-Х, 2003. — 439 С.
4. Краснопольский В. И., Серова В. А., Туманова В. А. // Российский вестник акушера-гинеколога. — 2004. — Т. 4, № 5. — С. 26—29.
5. Кулаков В. И., Анкирская А. С., Белобородов С. М. // Гинекология: Экспертные рекомендации по антибиотикотерапии. — 2005. — С. 3—5.
6. Потин В. В., Смагина Е. Е., Гзззян А. М. и др. // Журн. акушерства и женских болезней. — 2000. — Т. 49, № 2. — С. 59—66.
7. Серов В. Н., Тихомиров А. Л., Лубнин Д. М. Современные принципы терапии воспалительных заболеваний женских половых органов / Метод. пособ. для врачей. — М, 2003. — 23 с.
8. Серов В. Н., Царегородцева М. В., Кожин А. А. и др. // Журн. Рос. общества акушеров-гинекологов. — 2006. — № 4. — С. 32—34.
9. Смагина Е. Е. Роль аутоиммунных процессов в патогенезе нормогонадотропной недостаточности яичников: автореф. дис. ... канд. мед. наук. — СПб, 1995. — 19 с.
10. Beal A., Gerra F. // JAMA. — 1999. — Vol. 272. — P. 226—260.
11. Damewood M. D., Zacur H. A., Hoffman G. J., et al. // Obstet. Gynecol. — 1996. — Vol. 68, № 6. — P. 850—854
12. Haggerty C. L., Ness R. B. // Expert Rev. Infect. Ther. — 2006. — Vol. 4, № 2. — P. 235—247.
13. Luborsky J. L., Visintin I., Boyer S., et al. // Clin. Endocrinol. Metabol. — 1998. — № 70. — P. 69—75.
14. Tomer Y., Davies T.F. // Baillieres Clin. Endocrinol. Metab. — 1998. — Vol. 9, № 1. — P. 47—70.
15. Wiesenfeld H. C., Sweet R.B., et al. // J. Sex. Transm. Dis. — 2005. — Vol. 32, № 7. — P. 400—405.

УДК 616.62 — 006 — 089 — 036.8

ПРИМЕНЕНИЕ МНОГОФАКТОРНОГО КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА В ОПРЕДЕЛЕНИИ РИСКА РАЗВИТИЯ И В ПРОГНОЗИРОВАНИИ ИСХОДА РАКА МОЧЕВОГО ПУЗЫРЯ

Ш. Х. Ганцев, А. А. Зимичев, В. Н. Маклаков

Башкирский государственный медицинский университет, Самарский государственный медицинский университет, Самарский государственный технический университет

Корреляционно-регрессионный анализ позволяет выявить комплекс наиболее весомых независимых факторов, оказывающих влияние на возникновение и исход рака мочевого пузыря. Эти факторы могут быть положены в основу математических моделей риска развития и прогноза исхода рака мочевого пузыря. Разработаны компьютерные программы для оценки прогноза жизни больных, риска развития рака мочевого пузыря. Применение программ позволяет разработать комплекс мероприятий по первичной и третичной профилактике рака мочевого пузыря.

Ключевые слова: рак мочевого пузыря, корреляционно-регрессионный анализ, прогноз, заболеваемость.

USE OF MULTIFACTOR CORRELATION AND REGRESSION ANALYSIS IN DETERMINING THE RISK AND PROGNOSIS OF THE OUTCOME OF URINARY BLADDER CANCER

Sh. H. Gancev, A. A. Zimichev, V. N. Maklakov

Correlation and regression analysis reveals the most important and independent factors determining occurrence and outcome of the urinary bladder cancer. These factors may be used for the mathematical models of the risk assessment and prognosis of the urinary bladder cancer. Software for the estimation of life prognosis and risk assessment in patients with urinary bladder cancer was developed. Application of this software program enables to develop measures for the primary and tertiary cancer prophylactics.

Key words: urinary bladder cancer, correlation and regression analysis, prognosis, risk assessment.

Профилактика злокачественных новообразований является в настоящее время приоритетной с позиций современной онкологии. Для разработки комплекса мер по профилактике необходимо изучать основные факторы, приводящие к заболеванию. Наиболее перспективны в этом плане эпидемиологические факторы и, в пер-

вую очередь, экзогенные: профессия, вредные привычки, особенности питания и питьевого режима человека. На данный момент существующие методы лечения рака мочевого пузыря не всегда эффективны. Анатомо-физиологические особенности мочевого пузыря, биологические свойства опухолей, трудность определения стадии

и степени злокачественности, склонность к рецидивированию, множественность поражения, возраст больного и сопутствующие заболевания — вот далеко не все параметры, которые влияют на исход лечения. Совокупность перечисленных факторов создает большие трудности при выборе адекватного способа лечения.

ЦЕЛЬ РАБОТЫ

Улучшение диагностики и результатов лечения на основе выявления комплекса факторов, способствующих развитию рака мочевого пузыря и влияющих на прогноз злокачественного заболевания.

МЕТОДИКА ИССЛЕДОВАНИЯ

Для оценки риска заболевания и прогнозирования результатов лечения рака мочевого пузыря использована выборка из 510 больных раком мочевого пузыря. В качестве входящих переменных x_j использованы в обеих программах эпидемиологические информативные признаки: профессиональные особенности, курение, особенности питания, питьевой режим, пол и возраст пациента, генетическая детерминация опухолевых образований, наличие сопутствующих заболеваний мочеполовой системы, а в математической модели прогноза использованы клинико-морфологические информативные признаки: своевременность начала лечения, ультразвуковая структура рака мочевого пузыря, размер опухоли, ее локализация, характер роста рака мочевого пузыря, количество опухолей, функциональное состояние почек, степень распространенности заболевания по Tumor Node Metastasis (TNM), объем проведенного лечения, объем оперативного пособия, чувствительность рака мочевого пузыря к проводимой системной полихимиотерапии, гистологическая структура опухоли и ее гистопатологическая градация, степень инвазии рака в стенку мочевого пузыря, в лимфатические и венозные сосуды, рецидивный характер опухоли. В качестве выходящей переменной y принят уровень состояния здоровья пациента (наличие заболевания или рецидив опухоли, летальный исход). Каждой из входящих и выходящей переменных присвоено то или иное дискретное числовое значение натурального ряда в соответствии со степенью проявления переменной для каждого пациента.

Определены коэффициенты линейной корреляции R_{yx} , характеризующие степень взаимосвязи между величинами y , x . Однако по абсолютному значению коэффициента корреляции проблематично судить о тесноте связей с количественной точки зрения. Поэтому в данном случае были привлечены процедуры проверки статистических гипотез [1, 4]. Процедура проверки значимости начинается с формулировки проверяемой гипотезы или нулевой гипотезы H_0 . Например, в рассматриваемом случае $H_0: R = 1$ (корреляция имеет место). Тогда альтернативная гипотеза запишется как $H_1: R = 0$ (корреляция отсутствует). Оценка значимости коэффициента корреляции производится с помощью статистики F распределения. Если $F_{\text{набл}} > F_{\text{крит}}$, то нулевая гипотеза

H_0 принимается, а альтернативная H_1 отвергается, при этом вероятность отвергнуть правильную гипотезу H_0 равна уровню значимости α (0,01). Значение статистики F наблюдаемого вычисляется по имеющимся данным, а значение F критического находится по таблицам F распределения с учетом степеней свободы, зависящих от объема выборки и уровня значимости α (вероятность отвергнуть правильную гипотезу). Доверительный интервал для коэффициента корреляции строится с использованием \hat{z} преобразования Фишера [1].

Статистическая обработка проводилась на ЭВМ Intel Pentium 4, с использованием специально разработанной прогностической программы в среде Visual Basic 7.0 for Application (Свидетельство об официальной регистрации программ для ЭВМ №2005612134 от 19.08.05), позволяющей построить математическую модель как по риску развития рака мочевого пузыря, так и определению прогноза заболевания.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

В начале исследования вычислены парные коэффициенты корреляции для всех информационных признаков x_j и выходящей переменной y . При этом исключались факторы, для которых не получено достоверной корреляции с возникновением рака мочевого пузыря, и факторы, не влияющие на исход заболевания. Далее рассмотрен вопрос мультиколлинеарности. Явление мультиколлинеарности состоит в существовании линейной связи между объясняющими переменными. Наличие мультиколлинеарности вызывает технические трудности, связанные с уменьшением точности оценки тех или иных параметров или даже с невозможностью оценки вообще.

В результате из исследования были исключены факторы, не связанные с риском развития и исходом заболевания, а также связанные между собой, и всего было использовано 6 признаков ($m = 6$) для оценки риска развития рака мочевого пузыря (профессиональные особенности, курение, особенности питания, питьевой режим, генетическая детерминация опухолевых образований, наличие инфравезикальных заболеваний мочеполовой системы) и 6 признаков ($m = 6$) для прогноза исхода заболевания: курение сигарет, генетическая детерминация опухоли, глубина инвазии раковых клеток, рецидивная опухоль, локализация опухоли, характер лечения рака мочевого пузыря.

При исследовании влияния информативных факторов на развитие рака мочевого пузыря и выживаемость онкологических больных было использовано уравнение множественной линейной регрессии [2]. Но на практике в каждом конкретном случае следует предварительно убедиться, является ли зависимость между рассматриваемыми величинами $(x_j, j = \overline{1, 6}, y)$ линейной $y = ax + b$ [4]. Для всех независимых признаков $x_j, j = \overline{1, 6}$ и зависимой переменной y для уровня значимости $\alpha = 0,01$ была проверена статистическая гипотеза о том, является ли зависимость между величинами x_j, y линейной. В

случае, когда связь между величинами x , y не является линейной, можно воспользоваться квазилинейной зависимостью, так как известно, что при увеличении степени полинома точность аппроксимации увеличивается [4]

$$y = a_0 + a_1x + a_2x^2 + \dots + a_kx^k = a_0 + \sum_{j=1}^k a_jx^j.$$

Для всех перечисленных выше признаков, вплоть до $k = 15$, были вычислены коэффициенты соотношения по методу наименьших квадратов (МНК) [1, 4],

дисперсии остатков $s_u^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$, точность

$$L = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} * 100\% \quad [3], \text{ где } y_i \text{ — исходные данные,}$$

\hat{y}_i — результаты расчетов. Отметим, что дисперсию остатков s_u^2 можно рассматривать как относительную характеристику, а величину L — как абсолютную. Но увеличение порядка полинома k практически не привело к уменьшению дисперсии остатков и величин ошибок. Поэтому можно воспользоваться уравнением множественной линейной регрессии

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 = a_0 + \sum_{j=1}^6 a_jx_j.$$

Для определения прогноза заболевания необходимо найти значения коэффициентов уравнения множественной линейной регрессии, для которых сумма квадратов отклонений опытных и теоретических значений зависимой переменной минимальна. По имеющимся данным, помимо МНК-оценок коэффициентов соотношения были найдены доверительные интервалы.

Для характеристики тесноты связи между зависимой величиной y и несколькими независимыми величинами x_1, x_2, \dots, x_m вычислялся коэффициент множественной корреляции [1, 4]. Для рассматриваемых данных коэффициент множественной корреляции $R_{y,12\dots m} = 0,82$ ($F_{\text{наб}} = 32,0$; $F_{\text{крит}} = 1,49$) для модели прогноза исхода заболевания и $R_{y,12\dots m} = 0,70$ ($F_{\text{наб}} = 135,82$, что значительно превышает $F_{\text{крит}} = 2,02$) для оценки риска развития рака мочевого пузыря, что указывает на высокую тесноту связи между величинами y и x_1, x_2, \dots, x_m и в связи с этим может служить косвенным оправданием выбора функции в виде уравнения множественной линейной регрессии. Заметим, что для всех признаков коэффициент множественной корреляции оказался равным $R_{y,12\dots m} = 0,83$ и практически не отличается от указанного выше, что косвенно подтверждает правильность исключения некоторых признаков из рассмотрения.

Изложенная методика может быть использована для оценки риска и прогнозирования состояния здоровья отдельного пациента. Для нахождения риска и прогноза, если коэффициенты a_1, a_2, \dots, a_m найдены по МНК, достаточно в уравнение линейной регрессии подставить отдельные значения признаков пациента в виде вектора x_j и вычислить \hat{y}_i . Для вычисленного риска и

прогноза \hat{y}_i может быть построен доверительный интервал для уровня значимости $\alpha = 0,01$ с использованием неравенств $\hat{y} - t_{f,\alpha} s_e \leq \tilde{y} \leq \hat{y} + t_{f,\alpha} s_e$, где $t_{f,\alpha}$ — квантиль t -распределения при заданном уровне значимости α и числе степеней свободы и $f = n - m - 1$, $s_e^2 = s_u^2 x'(X'X)^{-1}x + s_u^2$, где X — матрица отдельных значений информативных признаков, символ ' означает операцию транспонирования, символ $^{-1}$ означает операцию вычисления обратной матрицы [1, 4]. Тогда с вероятностью $P = 0,99$ ($1 - \alpha$) можно утверждать, что истинное значение \tilde{y} при фиксированных значениях признаков x_j отдельного пациента находится в этом интервале.

С помощью специально разработанных программ рассчитаны величины риска развития и прогноза рака мочевого пузыря и доверительный интервал. Так как значения информативных признаков состоят из значений натурального ряда, а результат, в общем случае, не будет принимать целые значения, то можно говорить лишь о попадании в интервал, границы которого равно отстоят от целочисленных значений. Риск развития рака мочевого пузыря считается минимальным, если значение u менее 1,4; риск умеренный, когда число находится в интервале 1,5—2,4; высокий риск развития рака мочевого пузыря отмечается, если u превышает 2,5. Прогноз считается хорошим (безрецидивная выживаемость 10 и более лет), если число находится в интервале 0—2,5; прогноз считается удовлетворительным (безрецидивная выживаемость от 5 до 10 лет), если число находится в интервале 2,5—3,5; прогноз считается сомнительным (выживаемость больного более 5 лет, но возник рецидив заболевания), если число находится в интервале 3,5—4,5; прогноз считается неудовлетворительным (летальный исход в течение 5 лет), если число превышает 4,5.

ЗАКЛЮЧЕНИЕ

С помощью корреляционно-регрессионного анализа создана математическая модель оценки риска и прогноза рака мочевого пузыря. Использование математической модели и разработанной на ее основе компьютерной программы позволяет рассчитать индивидуальный риск заболевания и прогноз для каждого пациента проспективно и разработать комплекс мер по первичной и третичной профилактике заболевания, учитывающий индивидуальные особенности пациента.

ЛИТЕРАТУРА

1. Закс Л. Статистическое оценивание. — М.: Статистика. — 1976. — 598 с.
2. Иванов О. А., Сухарев А. Е., Егоров С. Н. // Рос. онкологический журн. — 1997. — № 5. — С. 35—38.
3. Мацкевич И. П., Свирид Г. П., Булдык Г. М. Теория вероятностей и математическая статистика. — Минск: Вишэйшая школа. — 1996. — 318 с.
4. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа. — М.: Финансы и статистика. — 1983. — 302 с.