

А. В. Дедов, Г. А. Попов, А. А. Панов

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СТЕПЕНИ ЗАВИСИМОСТИ ОСОБЕННОСТЕЙ КЛИНИЧЕСКОЙ КАРТИНЫ ОСТРОГО КОРОНАРНОГО СИНДРОМА ОТ НАЛИЧИЯ МАРКЕРОВ РЯДА ВИРУСНО-БАКТЕРИАЛЬНЫХ ИНФЕКЦИЙ

Оценка степени взаимосвязи между заболеванием и вирусно-бактериальными маркерами может являться основой как при принятии диагностических решений, так и при выборе методов лечения. Описан процесс формирования процедуры оценки, которая реализована на основе методов математической статистики применительно к влиянию вирусно-бактериальных маркеров на признаки, характеризующие клиническую картину хронических диффузных острых коронарных синдромов (ОКС). Приведены результаты обследований 136 пациентов с наличием направительного диагноза ОКС, подтвержденного с помощью общепринятого комплекса лабораторно-инструментальных методов исследования. Для решения проблемы использовался ранее модифицированный авторами применительно к многомерному случаю классический метод Манна – Уитни для проверки однородности двух выборок (многомерное обобщение). На основе предложенного метода была сформирована таблица оценок значимости модифицирующего действия 22 вирусно-бактериальных маркеров при ОКС. Анализ данных показал, что 4 маркера имеют высокий уровень связи с ОКС и могут быть эффективно использованы при проведении диагностических и лечебных мероприятий. Для 14 маркеров установлен средний уровень связи с ОКС, поэтому при их обнаружении необходимо проведение дополнительных диагностических процедур.

Ключевые слова: проверка гипотез на однородность, многомерные данные, принятие диагностических решений, метод Манна – Уитни, вирусно-бактериальные маркеры, острый коронарный синдром.

Введение

Проблема взаимосвязи различных заболеваний с составом и количеством различных вирусов и бактерий (как патогенных, так и нейтральных и полезных), а также антител к вирусно-бактериальным антигенам в организме пациента представляет большой интерес, поскольку понимание характера этой взаимосвязи позволит выбрать более эффективные методы лечения, способствующие нейтрализации опасных и негативно влияющих микроорганизмов и активизации микроорганизмов, поддерживающих процесс вылечивания. Кроме того, наличие или отсутствие этой зависимости может быть использовано при диагностировании заболеваний. Нами проводится анализ применительно к сердечно-сосудистым заболеваниям (ССЗ), относящимся к наиболее распространенным и опасным. Ранее подобный анализ проводился применительно к хроническим диффузным заболеваниям печени [1]. В настоящем исследовании методы и подходы, разработанные в [1], распространены на случай ССЗ. Как и при проведении исследований, выполненных в [1], имеется достаточно представительный набор результатов обследования пациентов по ССЗ, и в качестве основного метода анализа выбраны методы математической статистики.

Построение регрессионных моделей для установления диагноза по заболеваниям печени рассматривалось в [2], и полученные результаты используются ниже. Работ по анализу влияния различных микроорганизмов на процесс течения определенных заболеваний достаточно много, укажем прежде всего [3–5]. Однако методы анализа в этих работах отличаются от методов многомерного статистического анализа, которые предлагаются нами. *Целью исследований* являлось формирование формализованной процедуры оценки влияния вирусно-бактериальных маркеров на клиническую картину острого коронарного синдрома (ОКС) с точки зрения концепции «модификации» болезни.

Исходные данные обследования пациентов

В исследование были включены пациенты, которые находились на обследовании и лечении в отделении кардиологии (блоке интенсивной терапии) городской клинической больницы № 3 г. Астрахани в период 2004–2010 гг. Они составили группу из 136 лиц с ОКС в возрасте от 35 до 84 лет (в среднем $58,47 \pm 0,93$ года). Критериями включения в исследование

было наличие направительного диагноза ОКС, подтвержденного с помощью комплекса лабораторно-инструментальных методов исследования; отсутствие предшествующих нарушений ритма и клинически значимой сердечной недостаточности, не связанных с ишемической болезнью сердца, пороков сердца, кардиомиопатий, тяжелых конкурирующих и (или) сопутствующих заболеваний со стороны других органов и систем в фазе обострения или декомпенсации, а также отсутствие летальных исходов при нахождении в стационаре. Больных отбирали в случайном порядке. Все пациенты были обследованы согласно российским национальным стандартам в кардиологии. Кроме того, определяли сывороточные маркеры вирусных и бактериальных инфекций (*Chlamidophyla pneumoniae*, *Mycoplasma pneumoniae*, *Helicobacter pylori*, вирус простого герпеса (HSV-1), вирус Эпштейна – Барр (ВЭБ), цитомегаловирус (ЦМВ), вирус гепатита А (HAV) и антитела к патогенной и условно-патогенной микрофлоре (*Escherichia coli*, *Proteus mirabilis*, *Candida albicans*, *Streptococcus spp.*, *Bacteroides*, *Klebsiella pneumoniae*) классов IgG, IgM, IgA или суммарные антитела методом иммунофлюоресцентного анализа (качественным, полуколичественным или количественным). Была составлена база данных из 106 используемых параметров, включая результаты исследований, формализованные по «бинарному» принципу «0 или 1», т. е. «нет признака – есть признак» (жалобы пациентов, результаты физикального исследования, некоторые инструментальные и лабораторные показатели), а также количественные показатели – результаты определения биохимических параметров, данные инструментальных методов исследования (ультразвуковое, ЭКГ-исследование) и иные факторы (год рождения, возраст). В ряде случаев использовалась шкала признаков. Целью исследования была разработка алгоритма для оценки «вклада» в болезнь сывороточных маркеров вирусных и бактериальных инфекций, антител к патогенной и условно-патогенной микрофлоре.

Построение процедуры статистического анализа данных

Задача выявления зависимости между клиническими признаками заболевания и вирусно-бактериальными маркерами с точки зрения математической статистики, по аналогии с [1], может быть сведена к задаче проверки гипотезы на однородность двух многомерных выборок – для конкретного вирусно-бактериального маркера, данные по которому имеются в результатах обследования пациентов. Формируются две выборки. Первая – наборы результатов обследований по всем показателям больных, у которых зафиксировано либо наличие данного маркера (бинарный принцип – «есть – нет»), либо – при количественном определении (концентрация, превышающая таковую у здоровых лиц, т. е. норму). Вторая – аналогичные наборы у больных, у которых маркера либо не было, либо его концентрация не превышала пороговое значение (норму). В ряде случаев было возможно выделение дополнительной, третьей группы со сниженной, относительно «донорской», концентрацией антител, определяемых количественно.

Рассматривается гипотеза H_0 : можно ли считать, что обе выборки подчиняются одному и тому же вероятностному распределению, т. е. неразличимы с точки зрения их вероятностных свойств? Справедливость данной гипотезы будет означать, что наличие изучаемого конкретного вирусно-бактериального маркера не влияет на особенности протекания ОКС. Альтернативной является гипотеза H_1 – гипотеза о несовпадении распределений, которым подчиняются выборочные значения для пациентов по следующим двум выборкам: пациентов с высоким содержанием данного маркера и пациентов, у которых данный маркер практически отсутствует.

В исходных данных имеется ряд особенностей, которые также повлияли на выбор методов решения проблемы. Прежде всего, в исходных данных, которые формировались на протяжении многих лет, имелись большие пропуски, причем эти пропуски разбросаны по данным достаточно нерегулярно. Это связано с тем, что данные по ряду показателей стали собираться и фиксироваться относительно недавно, поэтому в материалах более ранних наблюдений информация по этим показателям отсутствует. Кроме того, анализ ряда показателей требует значимых усилий и финансовых затрат, что ограничило круг пациентов, для которых эти анализы были проведены. Наконец, по отдельным показателям данные отсутствуют по различным текущим причинам, связанным с конкретным пациентом, конкретными условиями работы медицинского учреждения. Существуют методы восстановления пропущенных данных (см., например, [5–7]). Однако эти методы эффективны для случаев, когда пропуски имеют негрупповой, разбросанный характер и их количество не очень велико, что не выполняется для использованных в нашем исследовании данных.

Далее, многие из исходных показателей имеют субъективное содержание – эти параметры были получены в основном при опросе и физикальном обследовании пациентов.

Перечисленные особенности наложили определенные ограничения на выбор методов проверки гипотез. Один из возможных подходов мог бы опираться на проверку одномерных гипотез об однородности по каждому показателю, и если хотя бы по одному из показателей гипотеза об однородности не выполнялась, то можно было бы считать, что и исходные наборы не являются однородными. Имеется достаточно много методов проверки гипотезы об однородности для одномерных выборок: для нормально распределенных наборов – критерии Стьюдента, F -критерий, критерий Крамера – Уэлча, непараметрические критерии Смирнова, типа ω -квадрат (Лемана – Розенблатта), Вилкоксона (Манна – Уитни), Ван-дер-Вардена, Сэвиджа, χ -квадрат и др. [5, 8]. Однако при таком подходе (на основе анализа одномерных выборок) игнорируется возможная взаимосвязь между признаками, т. е. они рассматриваются как независимые, что, вообще говоря, некорректно. Кроме того, поскольку в задаче, рассматриваемой нами, общее число признаков, или учтенных параметров болезни, равно 135, то практическое воплощение описанного подхода в рамках настоящей работы реализовать сложно ввиду ее трудоемкости. В связи с этим целесообразно рассмотреть многомерные методы анализа однородности. Воспользуемся предложенным в [1] вариантом многомерного обобщения метода Манна – Уитни, модифицированным с учетом особенностей имеющихся данных.

Анализ степени зависимости между признаками и маркерами

Описанная выше процедура была использована для оценки степени связи маркеров и признаков – как с отдельными, так и со всей их совокупностью. В качестве весового коэффициента w_j для каждого признака была выбрана величина, пропорциональная коэффициенту корреляции по Спирмену между ним и рассматриваемым маркером. В данной задаче коэффициент корреляции по Спирмену следует предпочесть классическому коэффициенту корреляции по Пирсону по двум причинам.

Во-первых, данные, приведенные в исходных таблицах наблюдений (прежде всего, данные субъективного характера), имеют по ряду признаков значительные погрешности, что, безусловно, отражается и на величине классического коэффициента корреляции. Но в отличие от классического, коэффициент корреляции по Спирмену вычисляется только на основе порядкового расположения оценок, без непосредственного учета их абсолютных значений – при искажении оценок их порядок обычно не изменяется.

Во-вторых, в отличие от классического коэффициента, коэффициент корреляции по Спирмену при линейном преобразовании признака может изменять свои значения. Указанное свойство в контексте нашего исследования важно при учете пола пациента (0 соответствует женскому полу, 1 – мужскому): классический коэффициент корреляции выдает одинаковые значения как для признака «пол мужской», так и для признака «пол женский», поскольку они линейно связаны – их сумма равна 1. В то же время значения коэффициента корреляции по Спирмену для мужского и женского пола различны, что, по-видимому, точнее передает содержательное понимание вопроса о взаимосвязи маркера с полом пациента.

Отметим, однако, следующее: при больших объемах выборок значения классического коэффициента корреляции и коэффициента корреляции по Спирмену отличаются в среднем на 5 %, что укладывается в рамки погрешности исходных данных. Это означает, что различия между значениями коэффициентов по Пирсону и Спирмену могут быть объяснены только погрешностями в исходных данных, и поэтому эти коэффициенты могут считаться равными (с точностью до погрешностей в исходных данных).

Результаты оценок рассмотренных выше коэффициентов корреляции по Спирмену приведены в таблице. В первом столбце таблицы оставлены только те маркеры, для которых выполняется предположение о связи их со всей совокупностью признаков ОКС при 5 %-м уровне значимости. Отметим, что каждый из этих маркеров оказался существенно взаимосвязанным с большим (135) числом признаков, т. е. имел большое абсолютное значение коэффициента связи признака с маркером.

**Оценки значимости
модифицирующего действия вирусно-бактериальных маркеров при ОКС**

| № | Маркер | Коэффициент связи набора признаков с маркером | Оценка модифицирующего действия маркера |
|----|---|---|---|
| 1 | <i>Chlamidophyla pneumoniae</i> IgM | 1,197 | Слабый модификатор |
| 2 | <i>Chlamidophyla pneumoniae</i> IgA | 1,563 | Средней силы модификатор* |
| 3 | <i>Chlamidophyla pneumoniae</i> IgG | 1,435 | Средней силы модификатор |
| 4 | <i>Mycoplasma pneumoniae</i> IgM | NS** | |
| 5 | <i>Mycoplasma pneumoniae</i> IgA | 1,367 | Средней силы модификатор |
| 6 | <i>Mycoplasma pneumoniae</i> IgG | 1,471 | Средней силы модификатор |
| 7 | <i>Helicobacter pylori</i> (суммарные антитела) | 1,419 | Средней силы модификатор |
| 8 | Вирус простого герпеса (HSV-1) IgM | 1,642 | Средней силы модификатор* |
| 9 | Вирус простого герпеса (HSV-1) IgG | NS** | |
| 10 | Вирус Эпштейна – Барр (ВЭБ) IgM | 1,290 | Средней силы модификатор |
| 11 | Вирус Эпштейна – Барр (ВЭБ) IgG | 1,362 | Средней силы модификатор |
| 12 | Цитомегаловирус (ЦМВ) IgM | 1,634 | Средней силы модификатор* |
| 13 | Цитомегаловирус (ЦМВ) IgG | NS** | |
| 14 | Вирус гепатита А (HAV) – суммарные антитела (КП – коэффициент позитивности) | 1,62 | Средней силы модификатор* |
| 15 | <i>Escherichia coli</i> | 1,62 | Средней силы модификатор* |
| 16 | <i>Candida albicans</i> | 1,071 | Слабый модификатор |
| 17 | <i>Streptococcus</i> spp. | NS** | |
| 18 | <i>Klebsiella pneumoniae</i> | 2,094 | Сильный модификатор |
| 19 | <i>Staphylococcus</i> | 2,911 | Сильный модификатор |
| 20 | <i>Pseudomonas aeruginosa</i> | 3,274 | Сильный модификатор |
| 21 | Гликолипин (эндотоксин) | NS** | |
| 22 | <i>Proteus mirabilis</i> | 3,274 | Сильный модификатор |

* $p = 0,055-0,08$, т. е. близко к статистической достоверности.

** Значение вероятности p не вычислено ввиду недостаточного объема необходимых данных.

Отметим, что 1,645 – табличный квантиль для 95 %-й статистической достоверности, 1,282 – для 90 %-й статистической достоверности; > 95 % – сильный модификатор, 90–95 % – средней силы модификатор, < 90 % – слабый модификатор

Как видно из значений коэффициентов связи набора признаков с маркером, при уровне значимости 0,95 значимыми являются маркеры-модификаторы под номерами 18, 19, 20, 22 (всего 4 из 22), указывающие на сильную связь с ОКС. При уровне значимости 0,9, что можно рассматривать как средний уровень связи, значимыми оказались 14 маркеров из 22 рассматриваемых, в том числе с номерами 2–8, 10–12, 14, 15. Два маркера под номерами 1 и 16 слабо влияли на показатели ОКС – коэффициенты соответствуют уровню значимости меньше 0,9. Наконец, по пяти маркерам с номерами 4, 9, 13, 17, 21 не удалось сделать конкретного вывода ввиду отсутствия требуемого объема данных одновременно по всей совокупности диагностических признаков.

Заключение

Таким образом, нами получены следующие результаты.

1. Исходная задача оценки влияния вирусно-бактериальных маркеров на клиническую картину ОКС сведена к задаче в области математической статистики – к задаче проверки гипотезы на однородность двух многомерных выборок.

2. Предложена алгоритмическая процедура проверки гипотезы на однородность двух многомерных выборок, обобщающая метод Манна – Уитни, предназначенный для проверки однородности одномерных выборок.

3. Описанная процедура проверки гипотезы об однородности использована для анализа указанных в исследовании данных по ОКС. Проведен анализ полученных результатов, сформулированы выводы и рекомендации. Предложена классификация вирусно-бактериальных маркеров по степени их влияния на многомерный массив данных по ОКС, а именно сильные, умеренные и слабые модификаторы болезни.

СПИСОК ЛИТЕРАТУРЫ

1. Дедов А. В. Определение диагноза по сердечно-сосудистым заболеваниям на основе использования набора формализованных моделей / А. В. Дедов, Г. А. Попов // Изв. Волгоград. гос. техн. ун-та. Сер.: Актуальные проблемы управления, вычислительной техники и информатики в технических системах. 2014. Вып. 22, № 3. С. 124–135.

2. Дедов А. В. Построение математических моделей предварительного диагноза заболеваний печени на основе методов регрессионного анализа / А. В. Дедов, Г. А. Попов // Вестн. Астрахан. гос. техн. ун-та. Сер.: Управление, вычислительная техника и информатика. 2014. № 4. С. 124–136.
3. Медик В. А. Руководство по статистике в медицине и биологии / под ред. Ю. М. Комарова. Т. 2. Прикладная статистика здоровья / В. А. Медик, Б. Б. Фишман, М. С. Токмачев. М.: Медицина, 2001. 352 с.
4. Леонов В. П. Статистика в кардиологии. 15 лет спустя / В. П. Леонов // Медицинские технологии. Оценка и выбор. 2014. № 1. С. 17–28.
5. Реброва О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTIKA / О. Ю. Реброва. М.: МедиаСфера, 2002. 312 с.
6. Bruijne M. H. J. de. On the use of Cox regression in the presence of an irregularly observed time-dependent covariate / M. H. J. de Bruijne, S. le Cessie, H. C. Kluine-Nelemas, H. C. van Houwelingen // Statistic in medicine. 2001. Vol. 20, no. 24, pp. 3817–3829.
7. Marshall G. Prospective prediction in the presence of missing data / G. Marshall, B. Warner, S. MaWhinney, K. Hammermeister // Statistics in Medicine. 2002. Vol. 21, no. 4. P. 561–570.
8. Холлендер М. Методы непараметрической статистики / М. Холлендер, Д. А. Вулф. М.: Финансы и статистика, 1983. 518 с.

Статья поступила в редакцию 12.12.2014

ИНФОРМАЦИЯ ОБ АВТОРАХ

Дедов Алексей Владимирович – Россия, 414000, Астрахань; Астраханская государственная медицинская академия; канд. мед. наук, доцент; доцент кафедры пропедевтики внутренних болезней; dedov1965.d@yandex.ru.

Попов Георгий Александрович – Россия, 414056, Астрахань; Астраханский государственный технический университет; г-р техн. наук; профессор; зав. кафедрой «Информационная безопасность»; popov@astu.org.

Панов Анатолий Анатольевич – Россия, 414000, Астрахань; Астраханская государственная медицинская академия, г-р мед. наук, профессор; зав. кафедрой пропедевтики внутренних болезней; agma@astranet.ru.



A. V. Dedov, G. A. Popov, A. A. Panov

MATHEMATICAL MODELING OF THE DEGREE OF THE DEPENDENCE OF THE PECULIARITIES OF THE CLINICAL PERFORMANCE OF THE ACUTE CORONARY SYNDROME ON THE MARKERS OF THE VIRAL AND BACTERIAL INFECTIONS

Abstract. The assessment of the relationship between a disease and viral-bacterial markers can be the basis as both in making diagnostic decisions and choosing the treatments. The procedure of assessment is implemented on the basis of the methods using mathematical statistics with regard to the influence of viral-bacterial markers on the features that characterize the clinical performance of chronic diffuse acute coronary syndromes (ACS). The results of the studies based on the examination of 136 patients with the existence of the guidance ACS diagnosis confirmed by a set of laboratory and instrumental methods, are presented. To solve this problem the authors used the previously modified in relation to the multidimensional case, classical Mann – Whitney method to test for homogeneity of two samples (multidimensional generalization). On the basis of the proposed method, a table of estimations of the importance of modifying actions of 22 viral and bacterial markers in ACS was formed. The analysis of the data in the table showed that 4 markers have a high level of dependence with ACS and can be successfully used for diagnostic and therapeutic measures. 14 markers have medium level of significance to ACS and their use in the treatment process requires additional diagnostic procedures.

Key words: hypothesis testing for homogeneity, multidimensional data, making of diagnostic decisions, Mann – Whitney method, viral and bacterial markers, acute coronary syndrome.

REFERENCES

1. Dedov A. V., Popov G. A. Opredelenie diagnoza po serdechno-sosudistym zabolevaniiam na osnove ispol'zovaniia nabora formalizovannykh modelei [Diagnosing cardiovascular diseases using a set of formalized models]. *Izvestiia Volgogradskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Aktual'nye problemy upravleniia, vychislitel'noi tekhniki i informatiki v tekhnicheskikh sistemakh*, 2014, iss. 22, no. 3, pp. 124–135.
2. Dedov A. V., Popov G. A. Postroenie matematicheskikh modelei predvaritel'nogo diagnoza zabolevanii pecheni na osnove metodov regressionnogo analiza [Construction of mathematical models of preliminary diagnosis of liver disease using the methods of regressive analysis]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravlenie, vychislitel'naia tekhnika i informatika*, 2014, no. 4, pp. 124–136.
3. Medik V. A., Fishman B. B., Tokmachev M. S. *Rukovodstvo po statistike v meditsine i biologii* / pod red Iu. M. Komarova. Vol. 2. Prikladnaia statistika zdorov'ia [Guidelines on statistics in medicine and biology. Ed. by Yu. M. Komarov. Vol. 2. Applied statistics of health]. Moscow, Meditsina Publ., 2001. 352 p.
4. Leonov V. P. Statistika v kardiologii. 15 let spustia [Statistics in cardiology. 15 years later]. *Meditsinskie tekhnologii. Otsenka i vybor*, 2014, no. 1, pp. 17–28.
5. Rebrova O. Iu. *Statisticheskii analiz meditsinskikh dannykh. Primenenie paketa prikladnykh programm STATISTIKA* [Statistical analysis of medical data. Application of the a set of applied software STATISTIKA]. Moscow, MediaSfera Publ., 2002. 312 p.
6. De Bruijne M. H. J., le Cessie S., Kluine-Nelemas H. C., van Houwelingen H. C. On the use of Cox regression in the presence of an irregularly observed time-dependent covariate. *Statistic in medicine*, 2001, vol. 20, no. 24, pp. 3817–3829.
7. Marshall G., Warner B., MaWhinney S., Hammermeister K. Prospective prediction in the presence of missing data. *Statistics in medicine*, 2002, vol. 21, no. 4, pp. 561–570.
8. Kholender M., Vulf D. A. *Metody neparametricheskoi statistiki* [Methods of nonparametric statistics]. Moscow, Finansy i statistika, 1983. 518 p.

The article submitted to the editors 12.12.2014

INFORMATION ABOUT THE AUTHORS

Dedov Alexey Vladimirovich – Russia, 414000, Astrakhan; Astrakhan State Medical Academy; Candidate of Medicine, Assistant Professor; Assistant Professor of the Department of Propaedeutics of Internal Diseases; dedov1965.d@yandex.ru.

Popov Georgiy Aleksandrovich – Russia, 414056, Astrakhan; Astrakhan State Technical University; Doctor of Technical Sciences, Professor; Head of the Department "Information Security"; popov@astu.org.

Panov Anatolii Anatolievich – Russia, 414000, Astrakhan; Astrakhan State Medical Academy; Doctor of Medicine, Professor; Head of the Department of Propaedeutics of Internal Diseases; admed@yandex.ru.

