

АНАЛИЗ АССОЦИАЦИЙ МЕЖДУ ВИРУСНО-БАКТЕРИАЛЬНЫМИ МАРКЕРАМИ И КЛИНИЧЕСКОЙ КАРТИНОЙ ХРОНИЧЕСКИХ ДИФФУЗНЫХ ЗАБОЛЕВАНИЙ ПЕЧЕНИ

Дедов А.В.¹, Попов Г.А.²

¹ФГБОУ ВО «Астраханский государственный медицинский университет» Минздрава России, Астрахань, e-mail: dedov1965.d@yandex.ru;

²ФГБОУ ВО «Астраханский государственный технический университет», Астрахань, e-mail: popov@astu.org

Статья посвящена формированию процедуры оценки степени взаимосвязи между заболеванием и вирусно-бактериальными маркерами, реализованной на основе методов математической статистики применительно к влиянию вирусно-бактериальных маркеров на признаки, характеризующие клиническую картину хронических диффузных заболеваний печени (ХДЗП). Исследования опираются на результаты обследований 165 пациентов хроническим гепатитом (ХГ) и циррозом печени (ЦП). Для решения проблемы авторами был модифицирован применительно к многомерному случаю классический метод Манна – Уитни для проверки однородности двух выборок (многомерное обобщение). На основе предложенного метода выявлены наиболее значимые маркеры при 5 % уровне статистической значимости. Для статистически значимых маркеров был проведен анализ их взаимосвязей с клиническими признаками ХДЗП. Предложен термин «коэффициент модификации».

Ключевые слова: метод Манна – Уитни, многомерное обобщение, вирусно-бактериальные маркеры, хронические диффузные заболевания печени, модификация болезни.

THE ANALYSIS OF THE ASSOCIATIONS BETWEEN THE VIRAL AND BACTERIAL MARKERS AND THE CLINICAL FEATURES OF THE CHRONIC DIFFUSE LIVER DISEASES

Dedov A.V.¹, Popov G.A.²

¹Astrakhan State Medical University, Astrakhan, e-mail: dedov1965.d@yandex.ru;

²Astrakhan State Technical University, Astrakhan, e-mail: popov@astu.org

The article is devoted to the creation of the procedure for the evaluation of the interrelation degree between the disease and the viral and bacterial markers, and was realised on the basis of mathematical, statistical methods with the special relation to the influence of viral and bacterial markers to the significant characteristics of clinical picture of chronic diffuse liver diseases (CDLD). The researches were based on the large diagnostic information collected during many years (165 patients). To solve this problem, the authors had used the multidimensional modification of the classical Mann - Whitney method for the checking the hypothesis of the uniformity of two choises. On the basement of the offered method the most significant markers were revealed with the 5 % significance level. For statistically significant markers the analysis of their interrelations with the clinical signs of CDLD was carried out. The term “coefficient of the modification” was proposed.

Keywords: Mann – Whitney method, multidimensional modification, viral and bacterial markers, chronic diffuse liver diseases (CDLD), modification of the disease.

Хронические диффузные заболевания печени (ХДЗП) являются одной из ведущих причин смерти лиц среднего возраста в России. Патология печени занимает 5-ое место в структуре заболеваемости среди всех видов болезней желудочно-кишечного тракта в Астраханской области. На особенности течения заболеваний печени в этом регионе оказывают влияние характер и качество пищевых продуктов и воды, климатические, социально-экономические факторы, вирусный фактор, а также склонность к злоупотреблению алкоголем, курение, психосоциальные стрессы [5]. Но лишь в немногих работах отмечается роль цитомегаловируса, вируса Эпштейн – Барр, герпес-вирусов, иных

вирусов при хроническом гепатите (ХГ) и циррозе печени (ЦП), причем чаще всего они рассматриваются как кофакторы, не имеющие самостоятельного значения [7]. Участие бактериального фактора (патогенной и условно-патогенной микрофлоры) в патогенезе заболеваний печени изучено в ещё меньшей степени. Микроорганизмы могут влиять на болезнь, выступая в роли этиологических факторов, провоцируя чрезмерный иммунный ответ, а также иными способами. Огромное разнообразие связей между маркером инфекции и патогенезом инфекционного процесса, которое CaronL. в статье 2003 г. называет «модификацией» [8]. Проблема взаимосвязи многообразных заболеваний, связанных с наличием в организме пациента различных вирусов и бактерий (как патогенных, так и нейтральных или даже полезных), а также антител к вирусно-бактериальным антигенам, представляет большой интерес. Выявление таких взаимосвязей позволяет выбирать более эффективные методы лечения, улучшать диагностический процесс.

Цель исследования: создание формализованной процедуры оценки влияния вирусно-бактериальных маркеров на течение хронических диффузных заболеваний печени (ХДЗП) с точки зрения концепции «модификации» клинической картины последних.

Материалы и методы исследования. За период 2000–2014 гг. у 42 здоровых лиц и у 96 больных ЦП и 69-ХГ, всего у 165 пациентов с ХДЗП, качественным, полуколичественным или количественным методом твёрдофазного иммуноферментного анализа (ИФА) были определены сывороточные маркеры: Ch. pneum., Muc. Pneum., H. pylori, вируса простого герпеса (HSV-1), вируса Эпштейна – Барр (EBV), цитомегаловируса (CMV), вируса гепатита А (HAV), антитела к сапрофитной и условно-патогенной микрофлоре, а именно к E. coli, Proteusspp., Candidaalbicans, Strept.spp., Bacteroides, Klebs. pneum. классов IgG, IgM, IgA или суммарные антитела. Средний возраст больных составил $48,8 \pm 1,7$ лет. Все пациенты были обследованы согласно Национальным Стандартам Российской Федерации с созданием базы данных, включающей формализованные результаты физикальных, лабораторных, инструментальных методов исследования, а также результаты определения вирусно-бактериальных маркеров. Всего в базе данных у каждого больного учитывалось 135 параметров [3]. В качестве базовой методики анализа выбрана модифицированная процедура проверки гипотезы об однородности двух многомерных выборок, построенная на основе непараметрического критерия рангов U Манна – Уитни, с определением U сначала по каждому признаку, затем с нахождением взвешенной суммы U_w . Полученная величина сравнивалась с табличным квантилем уровня значимости $U_{кр}$, равным 1,645 для 95 % достоверности. Если $U_w \geq U_{кр}$, то имелось существенное различие между уровнем признака в рассматриваемых выборках [1, 4].

Таким образом, поставленная задача была сведена к проверке гипотезы на однородность двух многомерных выборок. Для конкретного изучаемого вирусно-бактериального маркера формируются две выборки. Первая выборка – наборы результатов обследований по всем показателям больных, у которых зафиксировано либо наличие данного маркера (бинарный признак – «есть или нет маркер»), либо, при количественном определении, концентрация его превышала норму; вторая – аналогичные наборы по больным, у которых маркера либо не было, либо его концентрация была меньше таковой у здоровых добровольцев. В ряде случаев возможным было выделение дополнительной, третьей группы со сниженной относительно «нормальной» концентрацией антител, определяемых количественно. Рассматривается гипотеза H_0 : можно ли считать, что обе выборки подчиняются одному тому же вероятностному распределению. Справедливость данной гипотезы будет означать, что наличие изучаемого вирусно-бактериального маркера не влияет на особенности заболевания печени. Альтернативной является гипотеза H_1 о несовпадении распределений, которым подчиняются выборочные значения для пациентов, имеющих высокое содержание данного маркера, с распределением аналогичных значений для больных, у которых данный маркер отсутствует или имеет те же показатели, что и здоровые лица. Одним из возможных подходов мог бы быть проверка одномерных гипотез об однородности по каждому показателю, и если хотя бы по одному из показателей гипотеза об однородности не выполняется, то считается, что и исходные наборы не являются однородными. Имеется много методов проверки гипотезы об однородности для одномерных выборок [5, 6, 7]. Однако при таком подходе признаки рассматриваются как независимые, что некорректно. Кроме того, поскольку в задаче, рассматриваемой в данной работе, число признаков составляет 135, то практическая реализация описанного подхода в рамках данной работы трудно реализуема ввиду ее трудоемкости. Поэтому целесообразно рассмотреть многомерные методы анализа однородности. Наиболее эффективными методами проверки гипотезы об однородности многомерных данных являются методы, разработанные для анализа нормально распределенных данных. Однако предположение о нормальности неубедительно для данных, принимающих два значения: есть симптом (или признак) или нет его. Для решения рассматриваемой задачи может быть использован метод Манна – Уитни, модифицированный с учётом особенностей имеющихся данных. Ниже предлагается модификация метода, учитывающая эти особенности [2,3].

ИСХОДНЫЕ ДАННЫЕ. Имеется набор показателей X_1, X_2, \dots, X_L , и два набора наблюдений $\{x_{ij}^{(1)}\} (i = \overline{1; N_1}; j = \overline{1; L})$ и $\{x_{ij}^{(2)}\} (i = \overline{1; N_2}; j = \overline{1; L})$ по этим показателям, где N_1 и N_2 – объемы наблюдений (X_1, X_2, \dots, X_L) в первом и втором наборах. При этом отдельные наблюдения в каждом из наборов могут отсутствовать. В общем случае признаки имеют определенные «веса», характеризующие важность каждого признака: признак X_j имеет вес w_j , причем $\sum_{j=1}^L w_j = 1$. Если веса отсутствуют или малы, то $w_j = 1$ (они равнозначны) для всех $j = \overline{1; L}$. В качестве весов могут быть использованы пронормированные значения среднего арифметического коэффициентов корреляции между рассматриваемым признаком и другими признаками объединенной совокупности. Уровень значимости достоверности вывода по гипотезе $H_0: \alpha = 0,95$ (наиболее распространенное значение уровня значимости).

1. В тех позициях индексов i, j , по которым наблюдения отсутствуют, записываем недостижимое число, например, полагаем $x_{ij}^{(1)} = x_{\max} + 1$, либо $x_{ij}^{(2)} = x_{\max} + 1$ (либо другое «очень большое» значение), если соответствующее значение отсутствует, где $x_{\max} = \max_{i,j,k} x_{ij}^k$. Все имеющиеся наблюдения неотрицательны, и поэтому записанные значения будут максимальными в таблице данных. В результате наборы наблюдений не будут иметь пропусков.

2. Для каждого признака X_j составляем единый ранжированный ряд из обеих наборов наблюдений $\{x_{ij}^{(1)}\} (i = \overline{1; N_1})$ и $\{x_{ij}^{(2)}\} (i = \overline{1; N_2})$ сопоставляемых выборок, расставив их элементы по степени нарастания признака и приписав меньшему значению меньший ранг $r_{ij} (i = \overline{1; N})$, где $N = N_1 + N_2$ – общее количество всех наблюдений (и их рангов) в объединенной совокупности.

3. Находим сумму рангов T_{1j} , пришедшихся на долю элементов первой выборки, и отдельно сумму T_{2j} – на долю элементов второй выборки; то есть $T_{1j} = \sum^1 r_{ij}$ и $T_{2j} = \sum^2 r_{ij}$, где индексы «1» и «2» у знаков сумм означают, что суммирование берется только по элементам первой или второй выборок соответственно. Затем определяем большую из двух ранговых сумм $T_{j,\max} = \max(T_{1j}, T_{2j})$, и N_{\max} – объем соответствующей выборки, то есть $N_{\max} = N_1$, если $T_{1j} \geq T_{2j}$, и $N_{\max} = N_2$, если $T_{1j} < T_{2j}$.

4. Определяем значение U-критерия Манна-Уитни по каждому признаку на основе соотношения ($1 \leq j \leq L$):

$$U_j = N_1 \cdot N_2 + \frac{N_{\max} \cdot (N_{\max} + 1)}{2} - T_{j,\max}$$
 Затем находится взвешенная сумма $U = \sum_{j=1}^L w_j U_j$ (1)

5. Если объем хотя бы из одной выборок мал – меньше 20, то для нахождения критического уровня $U_{кр}$ при уровне значимости α распределения величины U необходимо составить таблицу значений величины U для данных объемов выборок N_1 и N_2 . Для этого воспользуемся таблицей значений критерия Манна-Уитни для одномерной выборки [2]. Именно, в определенном (например, лексиграфическом) порядке перебираются всевозможные значения величин $\{U_1, U_2, \dots, U_L\}$ и из таблицы одномерных распределений КМИ при каждом наборе значений $\{U_1, U_2, \dots, U_L\}$ вычисляется значение U по формуле (1), а также произведение соответствующих им вероятностей. Полученное произведение добавляется в качестве слагаемого к вероятности соответствующего значения U . В начальный момент значения всех вероятностей равны 0. Отметим, что в рассматриваемой в работе задаче значения N_1 и N_2 больше 20, и поэтому в рассматриваемой конкретной задаче выбор критического уровня $U_{кр}$ осуществляется на основе процедуры, описываемой ниже в пункте 6.

6. Если оба значения N_1 и N_2 больше 20, то для нахождения $U_{кр}$ можно воспользоваться асимптотическим приближением для U . Именно, при справедливости гипотезы H_0 статистика U и при достаточно большом объеме выборочных данных $N_1 \geq 20$ и $N_2 \geq 20$ имеет практически нормальное распределение со средним

$$M(U) = \sum_{j=1}^L w_j M(U_j) = \sum_{j=1}^L w_j \frac{N_1 \cdot N_2}{2} = L \cdot \frac{N_1 \cdot N_2}{2}$$

и дисперсией
$$D(U) = \sum_{j=1}^L w_j^2 D(U_j) = \sum_{j=1}^L w_j^2 \frac{N_1 \cdot N_2 \cdot (N_1 + N_2 + 1)}{12} = \left(\sum_{j=1}^L w_j^2 \right) \cdot \frac{N_1 \cdot N_2 \cdot (N_1 + N_2 + 1)}{12}$$

В частности, если все $w_j=1$, то
$$D(U) = L \cdot \frac{N_1 \cdot N_2 \cdot (N_1 + N_2 + 1)}{12}.$$

Отсюда выводим, что для нахождения $U_{кр}$ можно воспользоваться таблицами стандартного нормального распределения:

$$U_{кр} = L \cdot \frac{N_1 \cdot N_2}{2} + U_{кр}^{ст} \cdot \sqrt{\left(\sum_{j=1}^L w_j^2 \right) \cdot \frac{N_1 \cdot N_2 \cdot (N_1 + N_2 + 1)}{12}}, \text{ где } U_{кр}^{ст}$$

находится из таблиц стандартного нормального распределения как квантиль уровня значимости α .

7. Значение критерия U , полученное в результате обработки имеющихся наблюдений, сравнивается $U_{кр}$. Если полученное значение U меньше или равно $U_{кр}$, то признается наличие существенного различия между уровнем признака в рассматриваемых выборках и принимается гипотеза H_1 . Если же полученное значение U больше $U_{кр}$, то принимается нулевая гипотеза H_0 . Достоверность различий тем выше, чем меньше значение U .

Многомерная модификации метода Манна – Уитни для сравнения двух многомерных выборок (групп наблюдений)

Результаты исследования и их обсуждение

В качестве базовой методики анализа выбрана процедура проверки гипотезы оценки значимости модифицирующего действия вирусно-бактериальных маркеров об однородности двух многомерных выборок, построенная на основе непараметрического критерия рангов U Манна – Уитни, с определением U сначала по каждому признаку, затем с нахождением взвешенной суммы $U_{в}$, или коэффициент связи набора признаков с маркером (k), или

«коэффициент модификации». Полученная величина сравнивалась с квантилем уровня значимости $U_{кр}$, равным 1,645 (табличное данное) для 95 % статистической достоверности, а число 1,282 – для 90 % достоверности. Если $U_{в} \geq U_{кр}$, то имеется существенное различие между уровнем признака в рассматриваемых выборках. Предложено считать признак с достоверностью больше 95 % сильным модификатором, при 90–95 % – модификатором средней силы, а менее 90 % – слабым.

Таблица 1

Оценки значимости модифицирующего действия вирусно-бактериальных маркеров при ХДЗП

Маркер(антитела)	Коэффициент связи набора признаков с маркером (k)	Оценка модифицирующего действия маркера
ChlamidophylapneumoniaeIgM	0,912	Слабый модификатор
ChlamidophylapneumoniaeIgA	1,77*	Сильный модификатор
ChlamidophylapneumoniaeIgG	1,610	Средней силы модификатор
MycoplasmapneumoniaeIgM	1,014	Слабый модификатор
MycoplasmapneumoniaeIgA	1,309	Средней силы модификатор
MycoplasmapneumoniaeIgG	**	
H. pylori (суммарные антитела)	1,203	Слабый модификатор
Вирус простого герпеса (HSV-1) IgM	1,953*	Сильный модификатор
Вирус простого герпеса (HSV-1) IgG	NS	NS
Вирус Эпштейна – Барр (EBV) IgM	1,494	Средней силы модификатор
Вирус Эпштейна – Барр (EBV) IgG	1,596	Средней силы модификатор
Цитомегаловирус (CMV) IgM	2,059*	Сильный модификатор
Цитомегаловирус (CMV) IgG	0	NS
Вирус гепатита А (HAV)	**	
E. coli	1,691*	Сильный модификатор
Candidaalbicans	1,717*	Сильный модификатор
Streptococcuspp.	0,862	Слабый модификатор
Klebsiellapneumoniae	1,670*	Сильный модификатор
Staphylococcus spp.	0,505	Слабый модификатор
Ps. aeruginosa	1,933*	Сильный модификатор
Антиэндотоксиновые антитела	1,353	Средней силы модификатор
Proteusspp.	1,826*	Сильный модификатор
Bacteroides	1,028	Слабый модификатор

Примечание: при $k > 1,645$, $p^* < 0,05$; NS – не имеет статистических отличий; ** – p не вычислено.

Оценка значимости маркеров показала, что достоверно модифицировали клиническую картину ХДЗП наличие Chlamydia pneumoniae IgA (1,77), HSV IgM (1,95), Cytomegalovirus IgM (2,06), а также повышенные по сравнению с нормой уровни суммарных антител к E.coli (1,69), Proteus mirabilis (1,83), Candida albicans (1,72), Klebsiella pneumoniae (1,67), Pseudomonas aeruginosa (1,93). Установлена высокая степень соответствия между вычисленными «коэффициентами модификации» $U_{в}$ и особенностями течения ХДЗП у пациентов, имеющих различия по данным маркерам. Значимость маркера целесообразнее

оценивать как с учётом клинического анализа, так и путём вычисления коэффициента модификации, когда один метод дополняет другой.

На основе полученных данных можно наметить определенный спектр воздействий, потенциально могущих изменить, модифицировать течение заболеваний, через влияние на вирусно-бактериальные инфекционные маркеры. И хотя целью работы не была разработка новых методов лечения, но некоторые перспективные направления могут быть эмпирически предложены на основании проведенных клинических исследований и разработанного математического аппарата. Это может быть использование бактериофагов, в частности – хламидиофагов против парсистирующей хламидийной инфекции, лактулозы для коррекции кишечной микрофлоры, профилактическая вакцинация против HAV, лечение реактивации инфекций CMV и HSV при ХДЗП. Степень и направление этой индуцированной лечебным воздействием модификации болезни можно оценивать с помощью объективных методов математического моделирования. Методологические подходы, в частности – математический аппарат, предложенный в рамках настоящей работы, может дать новую информацию в рамках изучения роли не только вирусно-бактериальных маркеров, но и самих представителей микрофлоры человека при различных заболеваниях. Дальнейшее изучение выявленных особенностей, с применением новых технологий, может позволить уточнить детали патогенеза и процессы модификации течения острых и хронических неинфекционных заболеваний, тем самым внося вклад в теоретическую и практическую медицину. Предлагаемые методические подходы могут быть использованы как при решении других медико-биологических проблем, так и в иных, не связанных с медициной и биологией областях, при которых обрабатываются количественные данные, организованные в многомерные массивы.

Выводы. Установлены вирусно-бактериальные маркеры, достоверно модифицирующие клиническую картину ХДЗП. Предложен алгоритм выявления подобных маркеров, «модифицирующих» течение заболеваний печени. Получен объективный «коэффициент модификации» клинической картины болезни при наличии того или иного вирусно-бактериального маркера, согласующийся с клиническими данными. Выявлены сильные, слабые и средней силы модификаторы ХДЗП. Создана программа для ЭВМ «Автоматизированная система оценки различий между группами больных терапевтического профиля» (Свидетельство о регистрации № 2016611315 от 29.01.16), для количественной оценки и достоверности влияния различных факторов, позволяющая количественно определять вклад того или иного маркера в изменение (модификацию) клинической картины хронических диффузных заболеваниях печени. Наличие возможности вычисления коэффициентов модификации дает возможность создания так называемой «теории

модификации» применительно к ХДЗП.

Список литературы

1. Больнев Л.Н. Таблицы математической статистики / Л.Н. Больнев, Н.В. Смирнов. – 3-е изд. – Москва: Наука, 1983. – 474 с.
2. Дедов А.В. Автоматизированная система оценки различий между группами больных терапевтического профиля / А.В. Дедов, А.А. Дамакин (правообладатели) // Программы для ЭВМ. Базы данных. Топологии интегральных микросхем. Бюллетень. – 2016. – № 2; программа для ЭВМ № 2016611315 от 29.01.16 г.
3. Дедов А.В. Использование при принятии диагностических решений результатов статистического анализа влияния вирусно-бактериальных маркеров на клиническую картину заболевания / А.В. Дедов, Г.А. Попов // Прикаспийский журнал: управление и высокие технологии. – 2014. – № 4. – С.119-133.
4. Медик В.А. Руководство по статистике в медицине и биологии: в 2 т. / В.А. Медик, Б. Б. Фишман, М.С. Токмачев; под ред. Ю.М. Комарова. – Москва: Медицина, 2001. – Т. 2. Прикладная статистика здоровья. – 352 с.
5. Сердюкова Т.В. Мониторинг общей заболеваемости органов пищеварения в Астраханской области по отдельным нозологическим формам с 2006–2010 гг. / Т.В. Сердюкова, Н.Н. Курьянова. М.А. Сердюков // Астраханский медицинский журнал. – 2012. – Т. 7, № 2. – С. 116-119.
6. Bolukbas F.F. High rate of seropositivity of Chlamydia pneumoniae IgA in male patients with nonalcoholic steatohepatitis / F.F. Bolukbas, C. Bolukbas, F. Zeyrek, M. Aslan, H.I. Bahcecioglu, Ozar- dali, I. // Dig Dis Sci. – 2005. – Vol. 50, № 6. – P. 1141-1145.
7. Capron, L. What's in a risk factor? “He who strikes the ball” / L. Capron // Diabetes & metabolism. – 2003. – Vol. 29, № 1. – P. 6–12.