

УДК 61:311.3(075.8)

**ЕЩЁ РАЗ О МИНИМАЛЬНОМ ЧИСЛЕ БОЛЬНЫХ В КАЖДОЙ ГРУППЕ
ДЛЯ ВКЛЮЧЕНИЯ В ИСПЫТАНИЕ****Муслов С.А., Маев И.В., Салманов П.Л., Лапшихина Е.А.***ГБОУ ВПО «Московский государственный медико-стоматологический университет
им. А.И. Евдокимова (МГМСУ)» Минздрава РФ, Москва, e-mail: muslov@mail.ru*

Данная статья посвящена вопросам анализа эффективности различных лечебных вмешательств. Рассмотрены медицинские данные, представленные таблицами сопряженности с размерностью 2x2. При анализе статистической значимости различий использован точный метод Фишера. Показано, что если всех больных разделить на две группы, а клинический исход в каждой группе представить только двумя вариантами, то для того, чтобы иметь веские основания утверждать о преимуществе одного метода лечения (лекарственного средства, препарата и др.) над другим, необходимо как минимум 6 больных. При общем числе пациентов 5 и менее нет веских оснований (на снижаемом уровне значимости 0,05) для отклонения нулевой гипотезы. Перечислены некоторые возможные способы вычислительных процедур для решения поставленных задач.

Ключевые слова: медицинская статистика, значимость отличий, объем выборки, точный метод Фишера**AGAIN ON THE MINIMUM NUMBER OF PATIENTS IN EACH GROUP
FOR INCLUSION IN THE TRIAL****Muslov S.A., Maev I.V., Salmanov P.L., Lapshikhina E.A.***Moscow State Medical Stomatological University n.a. A.I. Evdokimov (MSMSU), Moscow,
e-mail: muslov@mail.ru*

This article is devoted to the correct recognition and analysis of pathological processes in the groups of patients. The case of the two groups of patients and two clinical contrary outcomes are examined. The theoretical analysis found on nonparametric Fisher's exact test. It is shown that if all the patients were divided into two groups, and clinical outcome has two mutually exclusive values, in order to have a reason to assert the superiority of one treatment or the drug over the other, need to be treated at least 6 patients (at a significance level of 0,05). When the total amount of patients in the 5 or less can not be a valid reason for rejecting the null hypothesis H_0 . The possible methods of computational procedures for solving the problem are provided.

Keywords: biostatistics, significance of differences, amount of sampling, Fisher exact method

В клинической практике достаточно частыми задачами, решаемыми статистическим исследованием, являются: установление идентичности сравниваемых групп больных по показателям, которые могут оказать влияние на результат лечения (к примеру, возраст, пол, тяжесть травмы или стадия заболевания и т.д.), диагностики или профилактики заболеваний; выявление различий в результатах лечения (к примеру, летальность, процент выхода на инвалидность, длительность койко-дня и т.п.); анализ статистической достоверности различий. Приведем такие примеры.

При оценке эффективности цитопротекторной терапии (двух-трех препаратов этой группы) при ишемическом инсульте необходимо определить, сопоставимы ли больные в этих группах по полу, возрасту, срокам развития ишемического процесса в мозге, по месту проживания. Вторым этапом анализа будет решение вопроса об эффективности препаратов, используемых для защиты клеток от ишемии. Такими показателями могут являться длительность утраты сознания, длительность периода нестабильности витальных функций, сроки регресса невроло-

гической симптоматики и т.д. Помимо этого врачу-исследователю крайне важно оценить статистическую значимость установленных различий в исходах лечения в группах. Не являются ли они случайными? Необходимо выяснить – являются ли различия в группах следствием естественных различий, существующих в выборках, или разница возникла из-за того, что один из методов лечения или лекарственных препаратов оказался значимо эффективнее другого.

При изучении эффективности нового препарата (скажем, противовирусных препаратов – «старого» ацикловира и «нового» рибавирина) показателями эффективности лечения могут быть сроки купирования катаральных явлений, длительность температурного периода, выраженность и длительность интоксикационного периода, процент госпитализаций, наличие осложнений противовирусной терапии (диспепсия, нейротоксичность). В экспериментальных медицинских исследованиях статистически могут анализироваться частота появления обычных реакций организма экспериментального животного при каком-либо воздействии (скажем, тахикардии при вве-

дении атропина) и необычных реакций (брадикардии при введении атропина), частота возникновения какого-либо заболевания при некоем воздействии (опухоль при облучении ионизирующей радиацией) или, наоборот, исчезновения этого заболевания при определенном лечении (к примеру, уменьшение числа колоний опухоли Раушенбаха в печени мышей после внутривенного введения перфторана). В опытной группе погибли все 3 животных, в группе контроля из 9 животных ни одно не погибло. Можно показать, что различия между группами статистически не значимы ($p > 0,05$). Различия становятся достоверными, если в контроле из 10 животных ни одно не погибло. То есть если в опытной группе погибли все 3 животных, то минимальный объем выборки, чтобы различия между опытом и контролем были достоверны, равен как минимум 13 животных.

В медицинских исследованиях итоговые данные о количестве наблюдений этих градаций признаков в группах больных – частоты или численности n_{ij} сводятся в многопольные таблицы сопряженности. В строках таблиц, как правило, описаны группы, сформированные в зависимости от использованных методов лечения, лекарственных препаратов и т.д. В столбцах представлены возможные уровни признаков, обусловленные различными результатами лечения – исходы лечения. Каждая ячейка таблицы стоит на пересечении столбца и строки. В результате каждая ячейка содержит комбинацию значений двух переменных (очевидно, поэтому эти таблицы называются таблицами сопряженности).

Самыми простыми перекрестными таблицами являются четырехклеточные таблицы с 2 строками и 2 столбцами – так называемые таблицы сопряженности 2×2 (табл. 1). Эти таблицы являются частым способом представления статистических данных при клинических исследованиях по качественным показателям (значительное число результатов клинических наблюдений являются не количественными, а качественными). Нетрудно видеть, что они соответствуют частному случаю номинальных бинарных шкал, то есть состоящих из двух градаций признаков по типу «да – нет» (например, наличие – отсутствие эффекта от лечения) в двух группах больных общим числом $N = n_{11} + n_{12} + n_{21} + n_{22}$.

Таблица 1
Общий вид таблиц сопряженности 2×2

	В	не В
А	n_{11}	n_{12}
не А	n_{21}	n_{22}

Таблицы сопряженности позволяют решить следующие задачи:

1) проанализировать эмпирические частоты наблюдений градаций исследуемых признаков в группах пациентов, подвергнутых различным вмешательствам, и определить наличие или отсутствие «терапевтической пользы»;

2) исследовать степень доверия к установленным различиям (статистическую значимость различий), а именно принять или опровергнуть нулевую гипотезу H_0 об отсутствии влияния данного хирургического метода или проводимой терапии на состояние больного $H_0: p = p_0; H_1: p \neq p_0$.

При этом сначала традиционно формулируется нулевая гипотеза, то есть изначально предполагается, что исследуемые методы, препараты и пробы не оказывают никакого влияния на результаты лечения и диагностики, а полученные различия между группами больных кажущиеся (случайны). Потом делается попытка опровергнуть нулевую гипотезу с определенной вероятностью и с помощью специальных критериев.

Поскольку речь идет о случайных событиях (на самочувствие больных всегда оказывает влияние множество факторов, как улучшающих, так и ухудшающих его состояние, и учесть все врачу никогда не представляется возможным), заключения проведенного исследования могут носить лишь вероятностный характер. Никакой статистический критерий не дает исследователю абсолютной уверенности в различии результатов лечения между группами или их идентичности. Другими словами, утверждения об эффективности лечения при использовании того или иного метода лечения или лекарственного препарата могут высказываться только с определенной вероятностью (0,95; 0,99 или 0,999 как это обычно принято в статистических исследованиях или на определенном уровне значимости 0,05; 0,01 или 0,001 соответственно).

Точный метод Фишера (ТМФ) достоверности наблюдаемых различий, как следует из его названия, является точным и может использоваться независимо от особенностей выборки. Он принадлежит к непараметрическим статистическим методам и основан на переборе всех возможных вариантов заполнения таблицы при фиксированных значениях маргинальных частот, что позволяет получать точные значения вероятности событий, столь же или еще менее вероятных, чем те, которые наблюдались в действительности. После этого делается вывод о том, достаточны ли различия в сравниваемых группах по величине, чтобы считать

их статистически значимыми (другими словами, есть ли связь между результатами и примененными способами лечения).

Цель исследования

Итак, зададимся вопросом, какие различия должны быть в значениях частот n_{ij} , чтобы можно было достоверно заявлять, что различия в результатах лечения достоверны, связаны с применением новых методов лечения и новых лекарственных средств и не связаны со случайными причинами? Какова должна быть минимальная численность выборки пациентов (наблюдений) $N = n_{11} + n_{12} + n_{21} + n_{22}$, чтобы иметь основания утверждать о преимуществе одного способа лечения (хирургического, терапевтического) над другим?

Методы исследования

На первый вопрос дает ответ, например, критерий хи-квадрат, а для малочисленных выборок – точный метод Фишера [1]. Именно ТМФ применен в данном исследовании.

Результаты исследования и их обсуждение

Чтобы ответить на второй вопрос, нужно рассмотреть наиболее «выгодные» варианты, например таблицы сопряженности 2×2 , у которых $n_{12} = n_{21} = 0$. Будем называть такие таблицы наиболее перекошенными. Для наиболее перекошенных таблиц объем выборки $N = n_{11} + n_{12} + n_{21} + n_{22} = n_{11} + n_{22}$ будет минимальным.

Также для определенности предположим, что метод или лекарство, примененные в одной группе (пусть это будет Группа 1), приносят успех, а метод или лекарство, использованные в другой группе (Группа 2), – неуспех (табл. 2). Таким образом, будем рассматривать благоприятные (😊) исходы в Группе 1 и неблагоприятные (😞) исходы в Группе 2. В качестве Успеха предположим, как это принято, успешный исход лечения, отсутствие боли при вмешательстве или осложнений после проведенного лечения, повышение индекса качества жизни и т.д. В качестве Неудачи – соответствующие противоположные исходы.

Таблица 2

Пример наиболее перекошенной таблицы сопряженности 2×2

	Успех	Неудача
Группа 1	n_{11}	0
Группа 2	0	n_{22}

Упростим обозначения. Пусть $n_{11} = n_1$, а $n_{22} = n_2$, как в табл. 3. Тогда $N = n_1 + n_2$.

Таблица 3

	Успех	Неудача
Группа 1	n_1	0
Группа 2	0	n_2

Для определения статистической значимости полученных в результатах лечения различий можем применить прямой расчет гипергеометрических вероятностей, лежащих в основе точного метода Фишера. Так вероятность реализации табл. 3 будет равна

$$p = \frac{C_{n_1}^0 \cdot C_{n_2}^{n_2}}{C_N^{n_2}} = \frac{1 \cdot 1}{C_N^{n_2}} = \frac{1}{C_N^{n_2}} = \frac{(N - n_2)! \cdot n_1!}{N!} = \frac{(n_1!)^2}{(n_1 + n_2)!},$$

где $C_n^m = \frac{n!}{(n-m)!m!}$ – число сочетаний,

$n! = 1 \cdot 2 \cdot \dots \cdot n$ – n -факториал, $0! = 1$.

Поскольку рассматриваемые таблицы являются максимально перекошенными, то соответствующие им гипергеометрические вероятности можно просто сравнить с принятым для анализа уровнем значимости (в противном случае нужно еще учитывать таблицы с теми же маргинальными частотами, но меньшими вероятностями их реализации). Можно также применить ряд статистических мини-программ [2] или универсальные пакеты типа «Statistica», «умеющие» работать с таблицами сопряженности [3]. Решение этих несложных задач максимально облегчается, если есть возможность воспользоваться заранее вычисленными таблицами В.С. Генеса [4] или В.Д. Брицкого [5].

Последовательно по мере возрастания n_1 рассмотрим частные случаи.

1. В результате проведенного в группах лечения нет благоприятных исходов в Группе 1 ($n_1 = 0$). Учтем, что для метода Фишера нет ограничений даже в том случае, если число больных с неким значением признака равно нулю (одно из значений n в ячейке равно 0). В этом случае применение ТМФ дает результат, согласно которому нет оснований говорить об эффективности лечения в Группе 1, даже если число неблагоприятных исходов в Группе 2 очень велико: $n_2 = N$ и $p = \frac{1}{C_N^{n_2}} = \frac{1}{C_N^N} = 1 > 0,05$ независимо от численного значения n_2 .

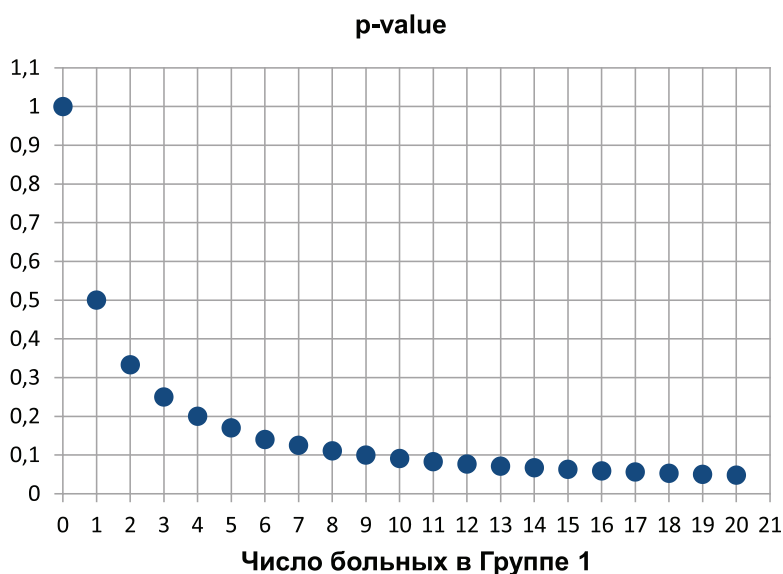


Рис. 2. Величина p-value в зависимости от числа больных в Группе 1

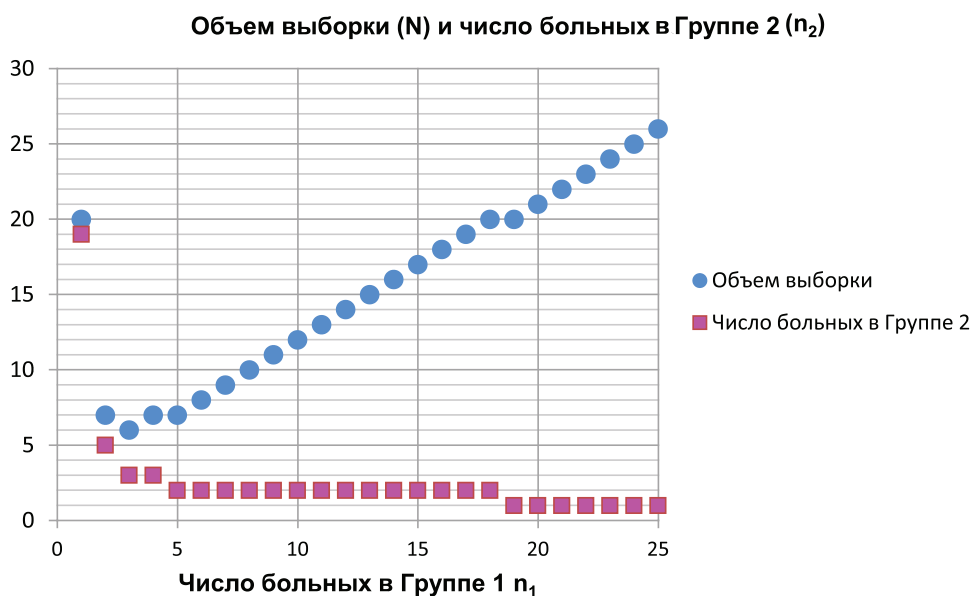


Рис. 3. Зависимость минимального числа больных в двух группах (объем выборки N) в зависимости от числа больных в Группе 1 (с благоприятными исходами), когда, в принципе, возможно установить статистически значимые различия в результатах лечения на уровне значимости 0,05

На самом деле, подсчитаем гипергеометрическую вероятность при $n_1 = n_2 = 3$ (вероятность наблюдения табл. 6).

Вероятность наблюдения этой таблицы равна $p = \frac{C_3^0 \cdot C_3^3}{C_6^3} = \frac{1 \cdot 1}{20} = 0,05$. Поскольку

таблица является наиболее перекошенной из всех возможных таблиц (с данным набором маргинальных частот), то получен-

ное значение вероятности 0,05 и есть то р-значение, которое фигурирует в точном методе Фишера и которое сравнивается с 0,05. Как видно, оно не больше 0,05, поэтому есть достаточные основания отвергнуть нулевую гипотезу (об отсутствии различий в группах) и принять альтернативную гипотезу о связи результатов лечения с применением того или иного метода лечения или лекарственного средства.

Таблица 7

Число больных в Группе 1 n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Объем выборки $N = n_1 + n_2$	20	7	6	7	7	8	9	10	11	12	13	14	15	16	17	18	19	20	20	21	22	23	24	25
Число больных в Группе 2 n_2	19	5	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1

Таблица 8

Общее число больных в группах, N	Варианты таблиц сопряженности 2x2, когда различия в группах достоверны ($p \leq 0,05$)
1	—
2	—
3	—
4	—
5	—
6	$\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$
7	$\begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 5 \end{pmatrix}$
8	$\begin{pmatrix} 6 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}$
9	$\begin{pmatrix} 7 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 5 & 1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 7 \end{pmatrix}$
10	$\begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 7 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 5 & 1 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} 4 & 1 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 1 & 5 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 0 & 6 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 7 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 8 \end{pmatrix}$

5. Дальнейший анализ при $n_1 = 4$ и $n_1 = 5$ показывает, что минимальный объем выборки больных равен 7, причем начиная с $n_1 = 5$ становится равным $n_1 + 2$, а при $n_1 > 18$ – сумме $n_1 + 1$ (табл. 7, рис. 3).

В заключение отметим, что при любых других вариантах таблиц сопряженности, менее перекошенных, чем рассмотренные (табл. 2), но с теми же маргинальными частотами, минимальное число больных

будет заведомо большим. Всего для малочисленных выборок ($N \leq 10$) возможны следующие варианты, когда результаты лечения в первой группе «лучше» и различия достоверны (табл. 8).

Как видно, не все варианты в таблице представлены максимально перекошенными таблицами. Число вариантов, когда различия достоверны, может быть увеличено путем изменения порядка следования строк

или столбцов, а также заменой строк на столбцы (транспонированием).

Выводы

Таким образом, при $n_1 = 0$ не может быть никаких оснований утверждать, что метод лечения, примененный в Группе 1, более эффективен, чем метод в Группе 2, даже при очень большом числе неблагоприятных исходов в Группе 2 (рис. 3). Случай $n_1 = 1$, когда в Группе 1 только один больной (с благоприятным исходом), также отбрасываем. В этом случае необходимо иметь 19 больных в Группе 2 и все (!) – с неблагоприятным исходом, то есть комбинацию: $\begin{pmatrix} 1 & 0 \\ 0 & 19 \end{pmatrix}$.

Только тогда на уровне значимости 0,05 можно утверждать, что лечение больных в Группе 1 эффективнее, чем в другой группе. Объем выборки здесь достаточно большой – 20 больных.

Дальнейший анализ наименьшего числа наблюдений, когда возможно статистически значимо утверждать, что между группами больных есть существенные различия и что один метод лечения или лекарственное

средство эффективнее другого, показал, что наименьшее число больных в обеих группах должно быть не менее 6.

При этом распределение больных по группам и исходам может быть таким как

в табл. 6, то есть $\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$ или $\begin{pmatrix} 0 & 3 \\ 3 & 0 \end{pmatrix}$.

Список литературы

1. Халафян А.А. STATISTICA 6. Статистический анализ данных. Второе издание. – М.: Бином, 2009. – 508 с.
2. Муслов С.А., Салманов П.Л., Беляев П.Я., Василенко А.В. Статистические программы для проверки нулевой гипотезы и вычисления p-value в медицинских исследованиях // Международный журнал прикладных и фундаментальных исследований. – 2011. – № 9. – С. 98–99.
3. Петри А., Сэбин К. Наглядная медицинская статистика. – М.: Геотар-Медиа, 2015. – 216 с.
4. Генес В.С. Таблицы достоверных различий между группами наблюдений по качественным показателям: пособие по статистической обработке результатов наблюдений и опытов в медицине и биологии. – М.: Медицина, 1964. – 80 с.
5. Гублер Е.В. Вычислительные методы анализа и распознавания патологических процессов. – Л.: Медицина, 1978. – 296 с.
6. Клюшин Д., Петунин Ю. Доказательная медицина. Применение статистических методов. – М.: Вильямс, 2008. – 320 с.