

Я. А. Туровский<sup>1,2</sup>, д-р техн. наук, канд. мед. наук, доц., Yaroslav\_turovsk@mail.ru,  
С. В. Борзунов<sup>2</sup>, канд. физ.-мат. наук, доц., А. А. Вахтин<sup>2</sup>, канд. физ.-мат. наук, доц.

<sup>1</sup> Институт проблем управления РАН им. В. А. Трапезникова, Москва

<sup>2</sup> ФГБОУ ВО Воронежский государственный университет

# Алгоритм коррекции статистического оценивания с учетом эффекта множественных сравнений на основе группировки результатов тестов<sup>1</sup>

*Многократное применение критериев статистических сравнений приводит к значительному возрастанию вероятности ошибочного установления различий в тех случаях, когда они статистически незначимы. В настоящей статье предложен алгоритм оценки значимости различий в задаче управления дизайном медико-биологического исследования с учетом вариаций значений попарного статистического критерия. Представлен пример применения алгоритма для окулографических данных.*

**Ключевые слова:** множественные сравнения, оценка значимости, дизайн исследования, окулографический интерфейс

## Введение

В подавляющем большинстве исследований явлений медицинской и биологической природы в настоящее время используют многофакторные зависимости. Соответственно, при этом возникает необходимость осуществить в рамках статистической обработки более чем одно применение статистического критерия. Как следствие, возникает широко известный эффект множественных сравнений, заключающийся в повышении риска необоснованно принять альтернативную гипотезу ( $H_1$ ), что, естественно, скажется на выводах исследования. Тем не менее различные методы противодействия данному эффекту основаны на пересчете уровня значимости, что в конечном итоге может, подобно поправке Бонферрони, привести к ошибочному отклонению гипотезы  $H_0$  [1–4]. В недавно опубликованной работе [5] описан разработанный авторами алгоритм управления дизайном медико-биологического исследования, основанный на оценке доли отклонений основной гипотезы  $H_0$  от общего числа сравнений в пространстве сравнения переменных. Использование данных не только о результатах статистического сравнения, но и об их распределении в рамках экспериментального плана исследования позволяет выделить группы реакций, в которых при сохранении мощности критерия снижается вероятность принятия ложноположительных гипотез. Действительно, если ложноположительные решения должны быть распределены равномерно

по всему пространству сравнения переменных, то группировка положительных решений в каких-то отдельных областях, отличных от случайного распределения, позволяет предположить наличие истинно положительных различий между сравниваемыми группами.

Однако при таком подходе учитываются все значения сравнений при стандартном значении  $p < 0,05$ , которые и формируют группы различий на факторном плане эксперимента (здесь и далее  $p$  означает уровень значимости результата попарного сравнения согласно применяемому статистическому критерию). В то же время подобный подход не учитывает, что полученные в ходе расчета статистических критериев значения  $p$  могут быть существенно меньше 0,05. Таким образом, возникает определенная сложность в интерпретации результатов, полученных согласно работе [5]. Действительно, предположим, что значения, полученные при применении статистического критерия, оказались много меньше 0,0001, но расположены эти различия не в группе, а по отдельности. Тогда, согласно указанному алгоритму, эти различия должны быть признаны как ложноположительные, что увеличивает вероятность ошибки второго рода. Отсюда возникает очевидная необходимость учесть разные значения вероятности ошибки, полученные при применении статистических критериев.

Таким образом, целью работы была модификация алгоритма снижения вероятности ошибки второго рода в условиях эффекта множественных сравнений, основанного на оценке группировки результатов применения тестов в медико-биологических исследованиях.

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-07-01037 А.

## Оценка значимости событий в условиях эффекта множественных сравнений

Рассмотрим задачу выделения из множества всевозможных результатов эксперимента подмножеств, выделяемых на основе дизайна эксперимента. Такие подмножества  $M_i$  (где  $i = 1, 2, \dots, J$ ) универсального множества  $M$  (содержащего результаты полного набора опытов проведенного эксперимента, иными словами  $M$  — пространство исходов анализируемого эксперимента) будем называть "блоками" или "кластерами" испытаний. Как элементы множеств  $M_i$ , так и число указанных множеств  $J$  подлежат определению. Для управления дизайном исследования наибольший интерес представляет ситуация, при которой число выявленных различий (при некотором фиксированном значении  $p$ ) без учета эффекта множественных сравнений сравнимо с числом, к которому стремится их число при только ложноположительных значениях попарного статистического критерия.

### Описание алгоритма

**Шаг 1.** В качестве начального шага предлагаемого алгоритма вычислим вероятность положительного исхода одиночного опыта  $p_+$  в предположении равномерного распределения положительных исходов в множестве согласно формуле  $p_+ = s/|M|$ , где  $s$  — число выявленных в эксперименте положительных исходов.

**Шаг 2.** Далее определим порог  $B$  числа ложноположительных решений, воспользовавшись неравенством

$$\sum_{k=1}^B C_{|M|}^k p^k (1-p)^{|M|-k} > 1 - p_0, \quad (1)$$

где  $p_0$  — уровень значимости, соответствующий применяемому попарному критерию;  $C_{|M|}^k$  — стандартное обозначение для биномиальных коэффициентов. Значение числа положительных исходов, превышающее  $B$ , трактуется как наличие в результатах исследования положительных результатов, которые не являются ложными.

С учетом информации о значении попарного критерия в  $i$ -м эксперименте  $p(i)$  выполним группировки элементарных опытов в блоки  $M_i$ ,  $i = 1, 2, \dots, J$ . Элементы объединяются в блоки исходя из постановки и дизайна эксперимента. В задаче об обнаружении синхронной электрической активности областей головного мозга элементарные исходы — сравнения пар сигналов — объединяются согласно мере близости параметров той или иной серии опытов как соответствие электродов тем или иным областям головного мозга. В целях формирования очередного блока  $M_i$ , вычислим число положительных результатов  $s_i$  и вероятность в предположении равномерного распределения  $p_i = s_i/|M_i|$ , где в знаменателе — число элементов в новом блоке.

**Шаг 3.** Сравним вероятности реализации  $s_i$  исходов в серии  $|M_i|$  испытаний с уровнем значимости в диапазоне  $[\min p(t_i), \max p(t_i)]$ :

$$\min_{1 \leq i \leq |M_i|} p(t_i) \leq C_{|M_i|}^{s_i} p_i^{s_i} (1-p_i)^{|M_i|-s_i} \leq \max_{1 \leq i \leq |M_i|} p(t_i). \quad (2)$$

Условие (2) играет центральную роль в рассматриваемом алгоритме и сравнивает вероятности реализации ровно  $s_i$  исходов в серии  $|M_i|$  независимых испытаний с диапазоном значений критерия, учитывающим его вариацию в различных индивидуальных опытах. Критерий (2) позволяет сформировать блок  $M_i$  за счет значительно меньшего значения попарного критерия в одном или нескольких исходах, уменьшая вероятность ошибки второго рода.

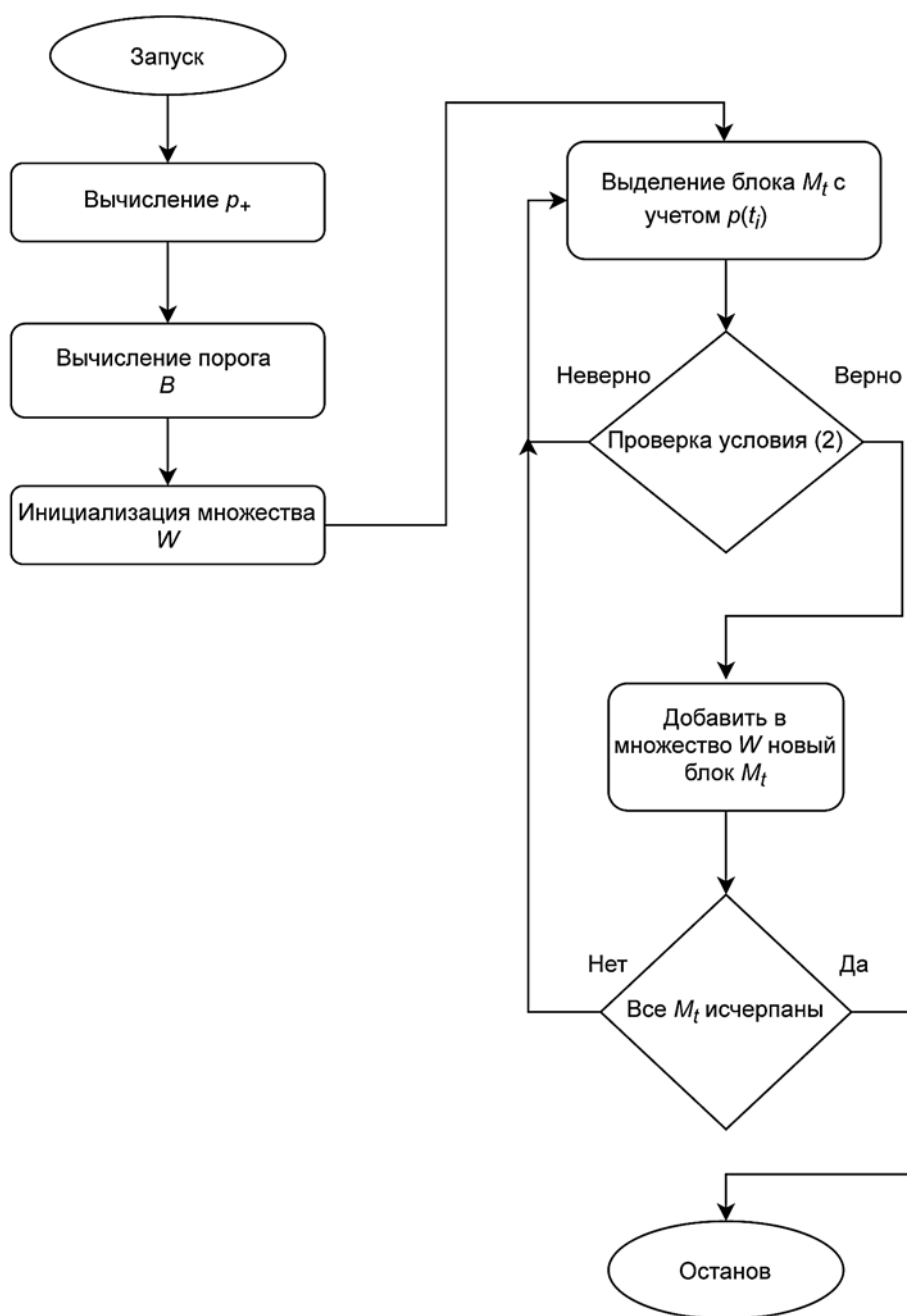
В результате получим множество блоков  $W = \{M_1, M_2, \dots, M_J\}$ . В него сгруппированы положительные исходы, обладающие вероятностями сформировать группу, значимо большими по сравнению с равномерным распределением. Далее происходит останов алгоритма.

Подчеркнем, что критерий (2) дает возможность учитывать и сравнивать между собой как маловероятные, так и практически достоверные события. Это обстоятельство имеет большое значение, в частности, для исследований медико-биологической природы. Такой эмпирический способ учета различных, в том числе очень малых значений критерия позволяет сконструировать простой алгоритм оценки значимости событий в условиях эффекта множественных сравнений. Блок-схема алгоритма приведена на рисунке.

Рассмотрим пример практического применения алгоритма. Пусть имеется экспериментальный план, в котором по одному фактору есть восемь возможных вариантов сравнений, по другому фактору — два, по третьему фактору — четыре, т. е.  $|M| = 8 \cdot 2 \cdot 4$ . В этом случае следует различать возможные сравнения и возможные переменные, полученные при использовании данного факторного плана. В простейшем случае, например, одна переменная и две группы, дадут нам одно сравнение, а одна переменная в трех группах в зависимости от подходов может дать как одно, так и три сравнения. Таким образом, при наличии в трех группах одного показателя в виде, например, окулографической активности, т. е. параметров движения глаз [6], можно воспользоваться критерием Краскейла—Уоллеса [7] и в ходе одного сравнения оценить, извлечены ли все три выборки из одной генеральной совокупности. Если группа одна и та же, но показатели получены в ходе лечения, возможно использование критерия Фридмана для связанных выборок. Однако, если необходимо определить конкретные межгрупповые различия, то нужно с соответствующими поправками применять попарные сравнения. В этом случае число таких сравнений увеличиться до трех.

Пространство событий состоит из пар элементарных опытов. Его образуют сравнения вида "фактор  $i$  — фактор  $j$  — фактор  $k$ ",  $1 \leq i \leq 8, 1 \leq j \leq 2, 1 \leq k \leq 4$ . Удобно рассматривать двумерный "срез" результатов, соответствующий фиксированной координате по одной из осей пространства событий, например, пусть  $k = \text{const}$ .

Предположим, что результаты серии попарных сравнений сформировали результаты, например, как это представлено в табл. 1. Прочерки в таблицах с



Блок-схема алгоритма

Таблица 1

Пример результатов применения попарного критерия — уровни значимости сравнений.  
Наблюдается формирование кластера согласно предлагаемому алгоритму

Варианты постановки серии опытов	Фактор 1-1 — фактор 2-1	Фактор 1-2 — фактор 2-2	Фактор 1-3 — фактор 2-1	Фактор 1-4 — фактор 2-2	Фактор 1-1 — фактор 2-1	Фактор 1-2 — фактор 2-2	Фактор 1-3 — фактор 2-1	Фактор 1-4 — фактор 2-2
1	—	—	—	—	0,0001	—	—	—
2	—	—	—	—	—	—	0,02	—
3	0,03	0,04	—	0,04	—	—	—	—
4	0,04	—	0,03	0,04	—	—	—	—

Пример результатов применения попарного критерия — уровни значимости сравнений.  
Значимые различия не формируют какие-либо блоки

Варианты постановки серии опытов	Фактор 1-1 — фактор 2-1	Фактор 1-2 — фактор 2-2	Фактор 1-3 — фактор 2-1	Фактор 1-4 — фактор 2-2	Фактор 1-1 — фактор 2-1	Фактор 1-2 — фактор 2-2	Фактор 1-3 — фактор 2-1	Фактор 1-4 — фактор 2-2
1	—	—	0,04	—	—	—	0,04	0,03
2	0,02	—	—	—	—	—	—	—
3	—	—	—	0,04	—	—	—	—
4	—	—	0,01	—	—	0,03	0,01	—

примерами означают отсутствие значимого результата применения попарного критерия.

Предположим, что общее число результатов сравнений, при которых выполняется неравенство  $p < 0,05$ , равно 13. Из следствия формулы Бернулли (1) согласно алгоритму следует, что вероятность выпадения от 0 до 6 ложноположительных результатов превышает 95,9 %. Значит, не менее семи из полученных результатов являются истинно положительными. Однако как и в работе [5] задачей является установление того, какие именно различия являются истинно положительными. В данном случае оценим, согласно (2), вероятность того, что кластер из восьми сравнений даст шесть истинно положительных решений, она составит  $3,948 \cdot 10^{-7}$  при 95 %-ном уровне вероятности положительных исходов. В предположении равномерного распределения исходов математическое ожидание числа попаданий равно 1,40. Поскольку число положительных исходов в нашем случае  $6 > 1,40$ , переходим к следующему шагу алгоритма. Вероятность получить кластер при  $p = 0,04$  не выше  $\pi' = 1,057 \cdot 10^{-7}$ , а при  $p = 0,03$  — не ниже  $\pi = 1,921 \cdot 10^{-8}$ . Внутри диапазона  $[\pi, \pi']$  получаем формирование кластера из опытов с положительными исходами.

Обратим внимание, что в табл. 1 значение  $p = 0,0001$  образует кластер из одного элемента. Таким образом, при применении стандартного подхода на основе поправки Бонферрони, из всех сравнений, представленных в табл. 1, должно было бы остаться

только единственное значение  $p = 0,0001 < \frac{p_0}{64}$ , что

привело бы к отклонению всех остальных результатов. Поскольку вероятность того, что шесть результатов сформируют кластер составляет менее  $3,948 \cdot 10^{-7}$ , то даже с учетом поправки Бонферрони можно признать его существование статистически значимым.

Рассматривая весь срез данных, следует признать, что сравнение для  $p = 0,02$  не может быть определено как ложно- или истинно положительное, и следовательно, должно быть исключено из рассмотрения.

В качестве примера, когда кластеры согласно алгоритму не формируются, можно привести результаты исходов в табл. 2. Из данных табл. 2 видно, что критерии значимости не превышают значений, отвечающих равномерному распределению ложноположительных

результатов, и оснований для выделения кластеров нет, поскольку области факторного плана, где наблюдаются статистически значимые различия, не позволяют принять гипотезу, что они формируют какую-либо группу.

## Заключение

Предложено обобщение алгоритма управления дизайном медико-биологического исследования, основанного на оценке доли отклонений основной гипотезы  $H_0$  от общего числа сравнений в пространстве сравнения переменных. Принят во внимание тот факт, что результаты применения статистического критерия позволяют учитывать не только пороговые значения критерия, но и показатели его значений в критической области. Продемонстрирована работа алгоритма в двух различных случаях, когда допускается как формирование кластеров положительных решений (принятия гипотезы  $H_1$ ), так и принятия единичной гипотезы  $H_1$  при достаточно малых значениях вероятности ошибки первого рода. На основе данных о разных значениях вероятности ошибки, алгоритм поможет подготовить дизайн медико-биологического исследования, что снизит вероятности ошибок второго рода в случае большого числа множественных сравнений. Полученный эффект достигается в первую очередь за счет введения дополнительной информации в виде распределения положительных решений в пространстве исходов применения статистических критериев.

## Список литературы

1. Гланц С. Медико-биологическая статистика. М.: Практика, 1998. 459 с.
2. Hochberg Y., Tamhane A. C. Multiple Comparison Procedures. New York: Wiley, 1987. 482 p.
3. Benjamini Y., Cohen R. Weighted false discovery rate controlling procedures for clinical trials // Biostatistics. 2017. Vol. 18, No. 1. P. 91–104. DOI: 10.1093/biostatistics/kxw030.
4. Lawrence J. Familywise and per-family error rates of multiple comparison procedures // Statistics in Medicine. 2019. Vol. 38, No. 19. P. 3586–3598. DOI: 10.1002/sim.8190.
5. Туровский Я. А., Борзунов С. В., Вахтин А. А. Алгоритм оценки результатов статистического анализа данных биомедицинской природы в условиях эффекта множественных сравнений // Программная инженерия. 2021. Т. 12, № 9. С. 470–474. DOI: 10.17587/prin.12.470-474.
6. Suefusa K., Tanaka T. A comparison study of visually stimulated brain-computer and eye-tracking interfaces // Journal of Neural Engineering. 2017. Vol. 14. P. 036009-1–036009-16.
7. Рунион Р. Справочник по непараметрической статистике: Современный подход. М.: Финансы и статистика, 1982. 198 с.

# An Algorithm for Correction of Statistical Estimations Taking into Account the Effect of Multiple Comparisons based on Test Results Grouping

**Ya. A. Turovsky**, Yaroslav\_turovsk@mail.ru, V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, 117997, Russian Federation; Voronezh State University, Voronezh, 394018, Russian Federation, **S. V. Borzunov**, sborzunov@gmail.com, Associate Professor, Voronezh State University, Voronezh, 394018, Russian Federation, **A. A. Vahtin**, alvahtin@gmail.com, Associate Professor, Voronezh State University, Voronezh, 394018, Russian Federation

*Corresponding author:*

**Turovsky Yaroslav A.**, Principal Researcher, V. A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences Moscow, 117997, Russian Federation; Head of Lab, Voronezh State University, Voronezh, 394018, Russian Federation  
E-mail: Yaroslav\_turovsk@mail.ru

*Received on December 05, 2021*

*Accepted on January 28, 2022*

*Repeated application of statistical comparison criteria leads to a significant increase of the probability of erroneous identification of differences in cases where they are statistically insignificant. An algorithm is proposed for assessing the significance of differences in the problem of managing the design of a biomedical study, taking into account variations in the values of a pairwise statistical criterion. An example of using the algorithm for oculographic data is given.*

**Keywords:** multiple comparisons, significance assessment, study design, eye tracking interface

**Acknowledgements:** The reported study was funded by RFBR according to the research project no. 19-07-01037 A.

*For citation:*

**Turovsky Ya. A., Borzunov S. V., Vahtin A. A.** An Algorithm for Correction of Statistical Estimations Taking into Account the Effect of Multiple Comparisons based on Test Results Grouping, *Programmnaya Ingeneria*, 2022, vol. 13, no. 3, pp. 148—152.

DOI: 10.17587/prin.13.148-152

## References

1. **Glantz S. A.** *Primer of Biostatistics*, McGraw-Hill, 2011. 320 p.
2. **Hochberg Y., Tamhane A. C.** *Multiple Comparison Procedures*, New York, Wiley, 1987, 482 p.
3. **Benjamini Y., Cohen R.** Weighted false discovery rate controlling procedures for clinical trials, *Biostatistics*, 2017, vol. 18, no. 1, pp. 91—104, DOI: 10.1093/biostatistics/kxw030.
4. **Lawrence J.** Familywise and per-family error rates of multiple comparison procedures, *Statistics in Medicine*, 2019, vol. 38, no. 19, pp. 3586—3598. DOI: 10.1002/sim.8190.
5. **Turovsky Ya. A., Borzunov S. V., Vahtin A. A.** An Algorithm for Evaluating the Results of Statistical Analysis of Biomedical Data under the Conditions of the Effect of Multiple Comparisons, *Programmnaya Ingeneria*, 2021, vol. 12, no. 9, pp. 470—474, DOI: 10.17587/prin.12.470-474 (in Russian).
6. **Suefusa K., Tanaka T.** A comparison study of visually stimulated brain-computer and eye-tracking interfaces, *Journal of Neural Engineering*, 2017, vol. 14, pp. 036009-1—036009-16, DOI: 10.1088/1741-2552/aa6086.
7. **Runyon R. P.** *Nonparametric Statistics: A Contemporary Approach*, Addison-Wesley Publishing Company, 1977, 218 p.

ООО "Издательство "Новые технологии". 107076, Москва, ул. Матросская Тишина, д. 23, стр. 2  
Технический редактор *Е. М. Патрушева*. Корректор *А. В. Чугунова*.

Сдано в набор 31.01.2022 г. Подписано в печать 28.02.2022 г. Формат 60×88 1/8. Заказ П1321  
Цена свободная.

Оригинал-макет ООО "Авансд солишнз". Отпечатано в ООО "Авансд солишнз".  
119071, г. Москва, Ленинский пр-т, д. 19, стр. 1. Сайт: [www.aov.ru](http://www.aov.ru)