

DOI: <https://doi.org/10.17816/phbn677119>

EDN: ALBZGF



Комментарии к использованию статистических критериев в экспериментальных исследованиях в фармакологии

П.П. Хохлов, А.А. Нужнова, М.И. Костина, А.А. Блаженко, Е.Р. Бычков

Институт экспериментальной медицины, Санкт-Петербург, Россия

АННОТАЦИЯ

Содержательность результатов исследовательской работы во многом обусловлена глубиной статистической обработки и качеством их представления. Однако в связи с недостаточным качеством статистического анализа и представления данных, в настоящее время актуальна проблема состоятельности результатов опубликованных исследований. Некорректная обработка первичных данных может привести к искаженным выводам, а также значительно затрудняет обобщающий анализ данных конкретной исследовательской темы. Авторы обсуждают наиболее частые и некорректные приемы, допускаемые в обработке исходных данных и планировании эксперимента: проблемы расчета объема выборки, некорректного использования статистических критериев, указания численных статистических параметров и описательной статистики, использования критериев согласия и представления данных, и приводят рекомендации о желательности или неуместности применения конкретных методов. При графической визуализации данных необходимо придерживаться некоторых правил. В случаях, когда распределение выборки нормальное, самой удобной мерой положения является выборочное среднее. Общепринятым считается указывать на графиках также стандартное отклонение как меру изменчивости. Наглядным является и представление доверительного интервала. Представление ошибки среднего также встречается, однако следует помнить, что стандартная ошибка среднего ничего не говорит о разбросе данных — это оценка того, насколько выборочное среднее отличается от такового по генеральной совокупности. Для описания данных, не согласующихся с нормальным распределением, применяют медиану вследствие ее нечувствительности к «выбивающимся» данным. В качестве меры разброса в этом случае уместно представить межквартильный размах ($Q1-Q3$). Для визуального представления результатов при больших объемах данных рекомендуется использовать столбчатые гистограммы, «ящик с усами», индивидуальные значения. Авторы настоятельно советуют не использовать таблицы для визуализации объемных результатов ввиду сложности прочтения при их представлении.

Ключевые слова: объем выборки; критерии различия; параметрические оценки; непараметрические оценки; применение критериев.

Как цитировать

Хохлов П.П., Нужнова А.А., Костина М.И., Блаженко А.А., Бычков Е.Р. Комментарии к использованию статистических критериев в экспериментальных исследованиях в фармакологии // Психофармакология и биологическая наркология. 2025. Т. 16, № 2. С. 125–130. DOI: 10.17816/phbn677119 EDN: ALBZGF

DOI: <https://doi.org/10.17816/phbn677119>

EDN: ALBZGF

Commentary on the Use of Statistical Tests in Experimental Pharmacological Studies

Platon P. Khokhlov, Alina A. Nuzhnova, Marina I. Kostina,
Alexandra A. Blazhenko, Evgeny R. Bychkov

Institute of Experimental Medicine, Saint Petersburg, Russia

ABSTRACT

The meaningfulness of the research findings is largely determined by the depth of statistical processing and the quality of their presentation. However, the paucity of high-quality statistical analysis and data presentation poses a significant challenge for maintaining the consistency of published research results. Incorrect processing of primary data can result in distorted conclusions and significantly complicate the generalization of data on a specific research topic. The authors discuss the most frequent and incorrect techniques used in working with initial data and planning an experiment, including issues related to the calculation of sample size, the incorrect use of statistical tests, the identification of numerical statistical parameters and descriptive statistics, and the use of goodness-of-fit and data presentation tests. They also provide recommendations on the appropriateness or inappropriateness of specific methods. Graphic visualization requires that certain principles be followed. For normal distributions, the sample mean is the most practical measure for position. It is a generally accepted practice to show the standard deviation on the graphs as a measure of variability. Confidence intervals are also presented to offer a visual reference. Standard errors may also be represented, however, they are not indicative of data variability, but rather, they are an estimation of the difference between the sample mean and the mean population value. Medians are used to describe non-normal distributions because they are insensitive to outliers. In those specific cases, interquartile ranges ($Q1-Q3$) can be used as a measure of variability. In the context of large data sets, bar graphs, histograms, box-and-whiskers, and individual values may be considered to enhance the comprehension of visual presentations. The authors strongly advise against the use of tables for displaying large amounts of data, citing the challenges associated with reading such presentations.

Keywords: sample size; test of differences; parametric estimates; non-parametric estimates; use of tests.

To cite this article

Khokhlov PP, Nuzhnova AA, Kostina MI, Blazhenko AA, Bychkov ER. Commentary on the Use of Statistical Tests in Experimental Pharmacological Studies. *Psychopharmacology and biological narcology*. 2025;16(2):125–130. DOI: 10.17816/phbn677119 EDN: ALBZGF

Received: 28.03.2025

Accepted: 24.06.2025

Published: 04.07.2025

ВВЕДЕНИЕ

В настоящее время наблюдается довольно упрощенный уровень представления результатов, что связано с некорректным использованием статистических методов. Это затрудняет сравнение результатов, полученных разными авторами, а также результатов, полученных разными методами, что приводит к искаженным и ошибочным выводам исследований.

Таким образом, можно констатировать наличие насущной проблемы в корректности применения современных статистических методов и стандартизации представления результатов, дающей возможность проведения сравнительных обзоров в области экспериментальной фармакологии.

Наиболее распространенными неточностями при использовании статистической обработки данных можно считать следующие.

1. Малое количество особей в группе.
2. Некорректное использование статистических критериев.
3. Формальное указание статистических критериев без количественных показателей. Например, в случае использования однофакторного дисперсионного анализа и последующего апостериорного теста следует указывать не только значение уровня значимости p , полученного в результате *post-hoc*-критерия, но и само значение F -критерия (например $F_{(2,35)}=9,1, p=0,01, n=12$).
4. Редкое использование критериев согласия. Перед непосредственным проведением сравнительного анализа необходимо установить принадлежность данных определенному распределению.
5. Пренебрежение описательной статистикой: мерой центральной тенденции (среднее), мерой рассеяния (стандартное отклонение).
6. Нерепрезентативность представления данных.

Из-за подобных неточностей или некорректного использования статистических методов изначально актуальные результаты могут выглядеть необъективными. Также иногда это приводит к невозможности сравнения результатов. Таким образом, актуальным представляется создание кратких рекомендаций по применению разных методов использования статистики.

РАСЧЕТ РАЗМЕРА ВЫБОРКИ

В подавляющем большинстве случаев стандартные исследования в сфере фармакологии направлены на нахождение различий между контрольными и экспериментальными группами, выявление токсичной дозы и взаимосвязи между двумя признаками. В зависимости от типа данных используют те или иные статистические критерии.

Отметим, что при планировании эксперимента важно заранее иметь представление о том, какие именно статистические методы будут использованы. Здесь важную роль играют чувствительность (мощность критерия),

которая зависит от уровня значимости α (в стандартном случае равен 0,05), размер эффекта, деленный на стандартное отклонение (параметр нецентральности), и объем выборки [1].

Для достижения нужного объема выборки, достаточного для оптимальной чувствительности критерия, используют ряд формул. Например, формула для расчета размера выборки для сравнения двух групп имеет вид:

$$n = \frac{2SD^2(Z_{\alpha/2} + Z_{\beta})^2}{d^2},$$

где SD — стандартное отклонение, $Z_{\alpha/2}$ — Z -оценка для значения, соответствующего $\alpha/2$, Z_{β} — Z -оценка для значения, соответствующего β (вероятность ошибки II рода, стандартно равна 0,2), d — размер эффекта. Величины SD и d являются эмпирическими, их получают из пилотного эксперимента либо из предыдущих исследований, однако такая возможность имеется не всегда. В случае невозможности нахождения эмпирических переменных для нахождения количества животных в группе также используют так называемый «метод уравнения ресурсов» (resource equation method) [2]. Предполагается, что данный расчет используется для последующего применения дисперсионного анализа. Этот метод опирается на то, что число степеней свободы, достаточное для достижения предела погрешности статистической модели, должно находиться в диапазоне от 10 до 20. Значение из этого диапазона должно получаться путем вычисления разности между общим количеством животных и общим количеством групп. Например, при наличии 4 групп по 5 животных в каждом получится $5 \times 4 - 4 = 16$, то есть наличие 5 особей в группе достаточно для дальнейшего анализа. Данный метод является достаточно грубым и не рекомендуется для регулярного пользования. Использовать его надлежит только в случае невозможности расчета по стандартным формулам, подобным указанной выше. Следует помнить, что данный расчет предполагает большие размеры эффекта [3].

Автоматический расчет выборок доступен в некоторых пакетах программ, как, например, G Power.

ПРОВЕРКА МАССИВА ДАННЫХ НА СООТВЕТСТВИЕ НОРМАЛЬНОМУ РАСПРЕДЕЛЕНИЮ

Начальным этапом анализа данных следует считать проверку массива данных на принадлежность определенному виду распределения. Обычно осуществляется проверка на нормальность, что впоследствии сужает круг используемых критериев.

Оценить, принадлежат ли данные нормальному распределению, возможно рядом способов.

Визуально такую задачу можно решить графическим методом, построив частотное распределение, график вероятности-вероятности (P-P график), график квантиль-квантиль (Q-Q график) и рядом других способов. Такие методы не подразумевают использование статистического теста [4].

Самым мощным критерием проверки на нормальность считается *W*-критерий Шапиро–Уилка [5]. В случае если уровень критерия значимости более 0,05, то эти распределения совпадают, что дает право использовать параметрические методы статистики, которые являются более строгими относительно аналогичных непараметрических. Одной из его особенностей является возможность использования для малых выборок (формально — 3–4 точки, но настоятельно рекомендуется брать больше), но при этом он эффективен и для больших.

Также во многих пакетах программ предлагается использование критерия Колмогорова–Смирнова — менее мощного, чем критерий Шапиро–Уилка [6]. Он рекомендуется для анализа исключительно больших выборок. Нет точного описания, насколько большой должна выборка для этого, но в некоторых источниках можно найти информацию, что требуется выборка более 15 точек.

Менее популярными считаются тесты Андерсона–Дарлинга (Anderson–Darling test), Д’Агостино–Пирсона (D’Agostino–Pearson test) и др. Первый тест основан на тесте Колмогорова–Смирнова, более чувствительно подчеркивая отклонения к концам распределений. Тест Д’Агостино–Пирсона проверяет нулевые предположения о том, что распределение данных симметрично и что эксцесс равен нулю. Однако перечисленные тесты менее мощные и также не могут быть использованы при маленьких выборках, что зачастую важно для экспериментальной фармакологии.

АНАЛИЗ ПАРНЫХ НАБЛЮДЕНИЙ

Часто в экспериментальной фармакологии используются сравнения двух групп, например, группы без воздействия (контрольная группа), и группа с введением препарата (экспериментальная группа). В таких случаях статистический анализ можно проводить по двум сценариям: анализ парных наблюдений и анализ независимых наблюдений.

В первом случае, например, при введении подопытным животным препарата в течение длительного времени и измерения неких показателей в начале и в конце эксперимента у одних и тех же особей, следует применять парный *t*-тест (если данные соответствуют нормальному распределению) или его непараметрический аналог — критерий Вилкоксона, либо критерий знаков (если выборка очень мала или данные не соответствуют нормальному распределению).

При анализе непарных наблюдений можно использовать непарный *t*-критерий: при небольших выборках

он стремится к *t*-распределению, однако при больших выборках (от 30 точек) это распределение стремится к нормальному. Стоит отметить, что параметрические критерии сильно чувствительны к выбросам: значительно выбивающиеся из общей массы значения могут сильно повлиять на результат анализа. При отклонении гипотезы о нормальном распределении для сравнения двух независимых выборок следует использовать непараметрический аналог *t*-теста — *U*-критерий Манна–Уитни.

В случае использования параметрических критериев также не стоит забывать об условии равенства дисперсий. Данную проверку можно выполнить с помощью *F*-теста Фишера. В некоторых пакетах программ он выполняется автоматически при использовании параметрических тестов.

МНОЖЕСТВЕННОЕ СРАВНЕНИЕ

В случае, когда требуется сравнить более чем 2 группы, следует применять дисперсионный анализ. Если данные соответствуют нормальному распределению, то обращаются к однофакторному либо двухфакторному/многофакторному дисперсионному анализу ANOVA. Напомним, что дисперсионный анализ не дает знаний о конкретных различиях между группами — его результаты лишь дают сведения о том, принадлежат ли выборки к одной генеральной совокупности. Для определения наличия статистически значимых изменений используют апостериорные тесты (post hoc-критерии).

Поправка Бонферрони — достаточно консервативный метод, к нему в настоящее время относятся с осторожностью. Это строгий критерий, который минимизирует вероятность допустить ошибку I рода, однако при большом количестве групп может допустить ошибку II рода.

Метод Холма–Бонферрони также имеет большую мощность, чем поправка Бонферрони. Он также решает проблему падения мощности поправки Бонферрони при возрастании количества групп.

Метод Шидака (Сидака), подобно перечисленным выше методикам, способствует предотвращению ошибки I рода путем коррекции значения *p* [7] и более предпочтителен, чем поправка Бонферрони.

Критерий Тьюки, так же как и поправки Бонферрони, представляет собой модификацию критерия Стьюдента, однако считается более либеральным методом, выполняющим сравнения средних. Активно используется в настоящее время. Работать с этим критерием необходимо, когда размеры выборок совпадают или очень близки.

Критерий Ньюмена–Кейлса — более мощный метод, сравнивающий средние значения аналогично критерию Тьюки, однако он в некоторой мере пренебрегает предотвращением ошибки I рода [8]. В отличие от критерия Тьюки он не предусматривает расчет доверительных интервалов. Авторы статьи не рекомендуют этот метод, необходимо тщательное обоснование причины его использования в статистической обработке экспериментальных данных.

Критерий Даннета используется для сравнения среднего значения экспериментальных групп со средним значением контрольной группы [9].

Напомним, что все вышеупомянутые критерии требуют равенства дисперсий в анализируемых группах.

В случае отрицательного теста на нормальность вместо дисперсионного анализа следует пользоваться его непараметрическим аналогом, а именно критерием Краскела–Уоллиса. Для множественных сравнений в данном случае также существует перечень критериев.

Для непараметрического дисперсионного анализа с повторными измерениями используют критерий Фридмана.

Основополагающий метод работы с ненормально распределенными данными — критерий Данна — считается основным непараметрическим тестом для множественных сравнений. Его используют в том числе и для неравных выборок. Так как непараметрические тесты оперируют сравнением рангов, некоторые параметрические анализы легко преобразуют в непараметрические: например, используют варианты критериев Ньюмена–Кейлса и Даннета для выборок одинакового объема.

ОПИСАТЕЛЬНАЯ СТАТИСТИКА И ПРЕДСТАВЛЕНИЕ ДАННЫХ

В настоящее время можно констатировать пренебрежение описательной статистикой: при описании результатов нередко не приводят сведения о мерах центральной тенденции и изменчивости либо представляют их некорректно. Хотя графическое представление данных является исчерпывающим, во многом полезно указывать и численные данные. К тому же для наглядности результатов также следует указывать доверительный интервал или размер эффекта.

Авторы напоминают, что при графической визуализации данных необходимо придерживаться некоторых правил.

В случаях, когда распределение выборки нормальное, самая удобная мера положения — выборочное среднее. Общепринятым считается указывать на графиках также стандартное отклонение как меру изменчивости. Наглядным является и представление доверительного интервала. Представление ошибки среднего также встречается, однако следует помнить, что стандартная ошибка среднего ничего не говорит о разбросе данных — это оценка того, насколько выборочное среднее отличается от такового по генеральной совокупности.

Для описания данных, не согласующихся с нормальным распределением, применяют медиану вследствие ее нечувствительности к «выбывающим» данным. В качестве меры разброса в этом случае уместно представить межквартильный размах ($Q1-Q3$).

Для визуального представления результатов при больших объемах данных рекомендуется использовать столб-

чатые гистограммы, «ящик с усами», индивидуальные значения. Авторы настоятельно советуют не использовать таблицы для визуализации объемных результатов ввиду сложности прочтения при их представлении.

КАТЕГОРИАЛЬНЫЕ ДАННЫЕ

Категориальные данные оцениваются с помощью построения матрицы сопряженности и дальнейшего применения метода Пирсона или точного теста Фишера (если в таблице сопряженности есть нули).

ПРОГРАММЫ ДЛЯ РАСЧЕТА СТАТИСТИЧЕСКИХ КРИТЕРИЕВ

Расчет статистических критериев удобно проводить в программах GraphPad Prism, Statistica, SPSS, R, MedCalc, Python.

ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

Вклад авторов. П.П. Хохлов, А.А. Нужнова, М.И. Костина, А.А. Блаженко, Е.Р. Бычков — анализ данных, написание статьи; Хохлов П.П. — разработка общей концепции. Все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Источник финансирования. Финансирование настоящего исследования осуществлено в рамках государственного задания Министерства науки и высшего образования Российской Федерации, тема FGWG-2025-0020 «Поиск молекулярных мишеней для фармакологического воздействия при аддиктивных и нейроэндокринных нарушениях с целью создания новых фармакологически активных веществ, действующих на рецепторы ЦНС».

Оригинальность. При создании настоящей работы авторы не использовали ранее опубликованные сведения (текст, данные).

Доступ к данным. Все данные, полученные в настоящем исследовании, доступны в статье.

Генеративный искусственный интеллект. При создании настоящей статьи технологии генеративного искусственного интеллекта не использовались.

Рассмотрение и рецензирование. Настоящая работа подана в журнал в инициативном порядке и рассмотрена по обычной процедуре. В рецензировании участвовали два внешних рецензента и член редакционной коллегии.

ADDITIONAL INFORMATION

Authors contribution. P.P. Khokhlov, A.A. Nuzhnova, M.I. Kostina, A.A. Blazhenko, E.R. Bychkov: data analysis, writing—original draft; P.P. Khokhlov: conceptualization. All authors made substantial contributions to the conceptualization, investigation, and manuscript preparation, and reviewed and approved the final version prior to publication.

Competing interests. The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

The source of financing. This work was supported by the state funding FGWG-2025-0020 “Search for molecular targets for pharmacological action

in addictive and neuroendocrine disorders with the aim of creating new pharmacologically active substances acting on CNS receptors.”

Statement of originality. The authors did not use previously published information (text, data) to create this paper.

Data availability statement. DATA generated in this study are available in the article.

Generative AI. Generative AI technologies were not used for this article creation.

Provenance and peer-review. This work was submitted to the journal on its own initiative and reviewed according to the standard procedure. Two external reviewers, and a member of the editorial board participated in the review.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

1. Glantz S. *Medico-biological statistics*. Transl. from English by Danilov YuA, Buzikashvili NE, Samoilov DV. ed. *Praktika*; 1999. 459 p. ISBN 5-89816-009-4
2. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? *Indian J Psychol Med*. 2013;35(2):121–126. doi: 10.4103/0253-7176.116232
3. Festing MF, Altman DG. Guidelines for the design and statistical analysis of experiments using laboratory animals. *ILAR J*. 2002;43(4):244–258. doi: 10.1093/ilar.43.4.244
4. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab*. 2012;10(2):486–489. doi: 10.5812/ijem.3505
5. Mohd R, Nornadiah M, Yap B. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J Stat Model Anal*. 2011;2:21–33.
6. Stephens MA. EDF statistics for goodness of fit and some comparisons. *J Am Stat Assoc*. 1974;69(347):730–737. doi: 10.2307/2286009
7. Blakesley RE, Mazumdar S, Dew MA, et al. Comparisons of methods for multiple hypothesis testing in neuropsychological research. *Neuropsychology*. 2009;23(2):255–264. doi: 10.1037/a0012850
8. Seaman MA, Levin JR, Serlin RC. New developments in pairwise multiple comparisons: some powerful and practicable procedures. *Psychol Bull*. 1991;110:577–586. doi: 10.1037/0033-2909.110.3.577 EDN: HKDMYN
9. Lee S, Lee DK. What is the proper way to apply the multiple comparison test? *Korean J Anesthesiol*. 2018;71(5):353–360. doi: 10.4097/kja.d.18.00242

ОБ АВТОРАХ

***Платон Платонович Хохлов**, канд. биол. наук;
Институт экспериментальной медицины; адрес: Россия, 197022, Санкт-Петербург, ул. Академика Павлова, д. 12;
ORCID: 0000-0001-6553-9267; eLibrary SPIN: 8673-7417;
e-mail: platonkh@list.ru

Алина Алексеевна Нужнова, студент;
ORCID: 0009-0002-1607-1471; eLibrary SPIN: 5521-5096;
e-mail: nuzhnova.aa@edu.spbstu.ru

Марина Игоревна Костина, студент;
ORCID: 0009-0005-1060-5489; eLibrary SPIN: 3909-2574;
e-mail: koctena@list.ru

Александра Александровна Блаженко, канд. мед. наук;
ORCID: 0000-0002-8079-0991; eLibrary SPIN: 8762-3604;
e-mail: alexandrablazhenko@gmail.com

Евгений Рудольфович Бычков, д-р. мед. наук;
ORCID: 0000-0002-8911-6805; eLibrary SPIN: 9408-0799;
e-mail: bychkov@mail.ru

AUTHORS INFO

***Platon P. Khokhlov**, Cand. Sci. (Biology);
Institute of Experimental Medicine; address: 12, Akademika Pavlova st., Saint Petersburg, 197022, Russia;
ORCID: 0000-0001-6553-9267; eLibrary SPIN: 8673-7417;
e-mail: platonkh@list.ru

Alina A. Nuzhnova, student;
ORCID: 0009-0002-1607-1471; eLibrary SPIN: 5521-5096;
e-mail: nuzhnova.aa@edu.spbstu.ru

Marina I. Kostina, student;
ORCID: 0009-0005-1060-5489; eLibrary SPIN: 3909-2574;
e-mail: koctena@list.ru

Alexandra A. Blazhenko, MD, Cand. Sci. (Medicine);
ORCID: 0000-0002-8079-0991; eLibrary SPIN: 8762-3604;
e-mail: alexandrablazhenko@gmail.com

Evgeny R. Bychkov, MD, Dr. Sci. (Medicine);
ORCID: 0000-0002-8911-6805; eLibrary SPIN: 9408-0799;
e-mail: bychkov@mail.ru

* Автор, ответственный за переписку / Corresponding author