



Статистический анализ в биомедицинских исследованиях: от достоверности к «тенденциям»?

А.Н. Хохлов 

*Сектор эволюционной цитогеронтологии, биологический факультет, Московский государственный университет имени М.В. Ломоносова, Россия, 119234, г. Москва, Ленинские горы, д. 1, стр. 12
e-mail: khokhlov@mail.bio.msu.ru*

На основании своего многолетнего редакторского опыта работы в биомедицинских научных журналах, а также опыта чтения лекций по оформлению научных статей для аспирантов Московского государственного университета имени М.В. Ломоносова автор анализирует сложившуюся ситуацию со статистическим анализом результатов биологических, биомедицинских и медицинских исследований — в частности, данных, полученных в работах, направленных на поиск лекарственных препаратов, предположительно замедляющих старение и подавляющих развитие возрастных болезней. Подчеркивается, что во многих случаях такой анализ либо вообще не проводится, либо проводится некорректно. Довольно часто авторы анализируют данные, не подчиняющиеся нормальному распределению, с помощью параметрического теста Стьюдента. При этом все чаще многие исследователи, не получив достоверных результатов ($p > 0,05$), заменяют в своих статьях понятие «достоверность» на понятие «тенденция», что представляется абсолютно недопустимым. Эта проблема затрагивает не только биомедицинские, но и другие естественнонаучные исследования, однако именно для биомедицины она представляется наиболее важной, ибо в данном случае речь идет о жизни и здоровье людей. Предлагается больше внимания уделять этому вопросу как рецензентам соответствующих статей, так и редакторам журналов, в которых они публикуются. Кроме того, подчеркивается важность преподавания основ статистической обработки результатов исследований студентам и аспирантам биологических, биомедицинских и медицинских институтов, а также соответствующих факультетов университетов.

Ключевые слова: биомедицинские исследования, статистический анализ, параметрические и непараметрические критерии, размеры выборок, кривые выживания, уровень значимости, геронпротекторы, экспериментальная геронтология

DOI: 10.55959/MSU0137-0952-16-80-4-7

Если ввести в поисковую строку базы данных Google Scholar слова «statistics biomedical research», можно получить огромное количество (более 8 миллионов!) ссылок, ведущих на статьи и книги, посвященные вопросам корректного использования статистических методов для анализа данных биомедицинских исследований. Это с очевидностью указывает на огромную важность данного вопроса для специалистов соответствующего профиля во всем мире. Очень много лет назад, еще когда я был молодым аспирантом, один из моих коллег рассказал мне о том, что в Великобритании проверили значительное количество научных публикаций, содержащих данные, требующие статистического анализа. В результате обнаружили, что приблизительно в 2/3 этих работ такой анализ был выполнен с грубыми нарушениями, т.е. сделанные авторами выводы были совершенно не обос-

нованными. Ясно, что формирование адекватной и репрезентативной выборки проанализированных в данном исследовании публикаций в свою очередь подразумевало корректный с позиций статистического анализа подход, в чем, конечно, нельзя было быть уверенным на 100%, однако сделанные выводы все же наводили на серьезные сомнения в «правильности» результатов, опубликованных в подвергнутых анализу работах.

Боюсь, что на сегодняшний день ситуация не только не улучшилась, но даже стала гораздо более настораживающей, о чем свидетельствует большое количество появляющихся вновь и вновь публикаций на эту тему ([1, 2] и др.). Как отметил С. Гланц в своей монографии 1992 г., «Как мы видим, статистические ошибки встречаются примерно в половине статей. Однако дальнейшие исследования показали, что журналам, в которых

взяли за правило обращать внимание не только на медицинскую, но и на статистическую сторону дела, удалось существенно снизить долю ошибочных статей. И эта доля нисколько не изменилась в тех журналах, которые так и не ввели статистического рецензирования» [1].

Мы обычно оцениваем важность рассматриваемых в научных публикациях проблем и полученных результатов по количеству скачиваний и цитирований этих работ, хотя, надо заметить, во многих случаях это не вполне корректно [3]. Я нашел в Google Scholar ссылку на профиль (https://scholar.google.com/citations?user=_QnLm3kAAAAJ&hl=en&oi=ao) коллеги из Оксфордского университета по имени Дуглас Альтман (Douglas G. Altman), статьи которого как раз в основном посвящены рассматриваемой проблеме статистического анализа в биомедицине. При этом всего у Альтмана в его профиле в настоящее время более 1 070 000 ссылок (!) на его работы (индекс Хирша – более 290!), а количество цитирований некоторых его отдельных статей по статистическому анализу в биомедицине достигает нескольких десятков тысяч [4–6]. Хотя, объективности ради, надо заметить, что на некоторые его не менее интересные работы в этой области ссылок на порядок меньше [7, 8]. Надо полагать, мои цитирования его публикаций в настоящей работе добавят еще кое-что в его копилку. Впрочем, если я правильно понял, он и так является самым цитируемым ученым Оксфордского университета (между прочим, одного из лучших в мире!) по данным Google Scholar Citations, обходя даже обычно лидирующих по количеству ссылок специалистов по искусственному интеллекту и визуализации головного мозга (brain imaging). Так что «классическим» исследователям, работающим в области биомедицинских исследований и не занимающимся профессионально методами статистического анализа данных, остается только завидовать Альтману и его коллегам.

Конечно, можно возразить, что работы Альтмана были опубликованы довольно много лет назад и именно поэтому набрали такое количество цитирований. Однако и в последнее десятилетие появилось большое количество публикаций по статистическому анализу в биомедицинских исследованиях [10–13], что явно свидетельствует о неослабевающем интересе к данной проблеме у специалистов соответствующих областей исследования.

Во всех упомянутых выше работах речь, как правило, идет о корректных подходах к обработке результатов самых разных биологических, биомедицинских и медицинских исследований. Совсем недавно я проанализировал [14] ситуацию со статистическим анализом данных в экспериментально-геронтологических статьях, поступающих в журналы, редактором которых я являюсь – «Ad-

vances in Gerontology» [15, 16] и «Moscow University Biological Sciences Bulletin» [17, 18]. В оба этих издания поступает много работ от ученых из разных стран. И, к сожалению, в большинстве этих рукописей мы обнаруживаем серьезные проблемы, касающиеся статистического анализа полученных результатов. Частично они были перечислены в одной нашей недавней публикации [19], инициированной курсом лекций «Основы наукометрии и подготовки научных публикаций» для аспирантов биологического факультета Московского государственного университета имени М.В. Ломоносова.

В этих лекциях мы довольно большое внимание уделяли и уделяем необходимости корректной статистической обработки экспериментальных данных, публикуемых в научных статьях. Собственно, в биомедицинских экспериментах – например, на крысах – одно животное всегда тяжелее другого, даже если мы берем вполне однородную группу особей. Вообще-то это касается и любых других физиологических, биохимических и молекулярно-биологических показателей. Однако эти различия показателей должны быть случайными, а распределение животных по изучаемому показателю – нормальным (гауссовским). И если мы сравниваем контрольную и опытную группы (например, изучая влияние на людей, животных или клеточные культуры какого-то потенциального геропротектора) с помощью параметрического критерия Стьюдента, то такой подход подразумевает нормальность распределения животных по изучаемому параметру в обеих группах. В «правильных» компьютерных программах, предназначенных для статистического анализа данных – например, SigmaPlot (Grafitti LLC, US) – при попытке запустить тест Стьюдента система анализирует эти данные на нормальность и после этого оценивает достоверность различий между опытной и контрольной группами только в том случае, если распределение изучаемых показателей достоверно не отличается от нормального. Если отличается, то программа просто предложит вам использовать непараметрический тест, отказавшись использовать параметрический. Но, судя по всему, большинство исследователей ограничивается тестом Стьюдента, встроенным в MS Excel, который никакой проверки на нормальность не подразумевает. А о существовании непараметрических критериев статистического анализа они, как ни печально, часто просто не имеют представления. При этом объективности ради надо, конечно, заметить, что нормальность распределения данных – хоть и необходимое, но не единственное требование к возможности использования параметрического критерия сравнения [20].

К сожалению, как правило, и рецензенты, и редакторы биомедицинских научных журналов не очень много внимания уделяют статистической обработке данных, полученных авторами статей.

И это понятно: довольно трудно корректно проанализировать громоздкие математические выкладки, которые обосновывают достоверность полученных результатов. Кроме того, во многих случаях подробности статистического анализа данных авторы оставляют «за кадром».

Конечно, общепринятый уровень достоверности $p = 0,05$ (для наиболее придирчивых — $p = 0,01$) является во многом «искусственным» и отражает лишь некоторые условные договоренности, достигнутые исследователями на протяжении многих десятилетий. Однако договоренности нужно соблюдать, и это касается также упомянутого «барьера достоверности». Впрочем, не будучи «крутым» специалистом в статистике, я не могу правильно объяснить своим студентам и аспирантам, почему $p = 0,049$ — это хорошо, а $p = 0,051$ — плохо. Хотя я слышал от одного из коллег, что поговорка «семь раз отмерь, один раз отрежь» возникла не на пустом месте: в случае использования критерия знаков для анализа парных сравнений объектов двух выборок 2^6 (имеется в виду 6 одинаковых знаков — например, 6 раз подряд показатель в контроле превышает таковой в опыте) — это 64%-ная вероятность неслучайности обнаруженных различий, а 2^7 — уже 128%-ная (т.е. эта вероятность уже достигла и даже, условно, «превысила» 100%). Нельзя также не заметить, что делаются неоднократные попытки обосновать увеличение строгости требований к значимому уровню различий, например — уменьшение его до $p < 0,005$ [21].

Однако давайте представим себе, что мы сделали все совершенно правильно и корректно использовали одну из многочисленных существующих программ анализа полученных данных. И... оказалось, что достоверных различий, свидетельствующих, например, об эффективности исследуемого нами биомедицинского препарата, в нашей работе обнаружено не было. И вот тут мы подходим к той самой теме, которая вынесена в заголовок настоящей статьи.

Можно просто построить гистограммы со столбиками разной высоты для контрольной и опытной группы, но без «усов», показывающих ошибку описываемых данных. А потом в обсуждении заметить: вот, мол, в контроле столбики ниже, а в опыте выше, так что препарат со всей очевидностью действует. То есть, как обычно пишут наши авторы, «обнаружена очевидная тенденция». И, как ни печально, многих рецензентов, редакторов, а также читателей уже опубликованной работы это вполне может устроить. Автор микробиологической работы, поступившей в один из моих журналов, написал мне в ответ на замечания: «У нас в микробиологии не принято статистически анализировать полученные данные». И это при том, что в статье были приведены результаты сравнения влияния различных культуральных

сред на рост бактерий! Как можно в такой работе сделать какие-либо выводы без использования статистических методов, мне совсем не понятно. При этом, и это особенно огорчает, в ответ на мои возражения автор прислал несколько своих ранее опубликованных статей во вполне приличных журналах (и даже в изданиях, входящих в Q1), где статистическая обработка результатов действительно отсутствовала!

Известно, что для получения достоверных результатов количество животных, используемых в геронтологических исследованиях, должно быть достаточно большим. В случае построения кривых выживания — не менее 50 в каждой группе. В противном случае на «хвостах» таких кривых нам просто не с чем будет работать. При тестировании геропротекторов с использованием показателей биологического возраста количество объектов как в контрольной, так и в опытной группе может быть меньшим, но все равно достаточным для корректных выводов о влиянии исследуемого препарата на старение и продолжительность жизни. Это же касается количества клеток или культуральных флаконов и гемоцитометров (камер Горяева) в цитогеронтологических экспериментах [22, 23]. И мне не кажутся оправданными объяснения некоторых коллег, которые говорят мне что-нибудь вроде: «Эти животные очень дорогие и редкие, так что в каждой группе у нас было не более 2–3 особей». Или: «Где же мы возьмем больше одной королевы голых землекопов?». Впрочем, полученные результаты они, тем не менее, легко публикуют в высокорейтинговых журналах, особенно если эти журналы получают с авторов большую плату за публикацию (APC — article processing charge) и, как ни печально, заинтересованы в том, чтобы закрывать глаза на проблемы со статистической обработкой результатов [3, 24]. Однако даже если животные не «очень дорогие», авторы иногда могут вообще не указать их количество в своих экспериментах, но, тем не менее, без проблем опубликоваться в одном из самых престижных научных журналов [3]. И это при том, что существуют проверенные способы оценки минимального объема выборки для получения значимых результатов — например, с помощью программы «G*Power»: (<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>) [25].

И самое главное: просто ужасает огромное количество научных статей, авторы которых, не получив значимых результатов ($p > 0,05$), спокойно отмечают выявленную ими «очевидную тенденцию» — например, влияния изучаемого препарата на те или иные живые организмы или культивируемые клетки. Особенно мне нравится утверждение: «Да, у нас $p > 0,05$, но совсем ненамного, так что мы думаем, что эффект точно обнаружен». Я всегда считал, что отсутствие достоверных отли-

чий между опытом и контролем свидетельствует (не забываем об упомянутой выше «договоренности») лишь о том, что таких различий просто нет. Если же авторы действительно имеют основания полагать, что различия все-таки должны быть, то им нужно значительно увеличить размеры контрольной и опытной группы объектов и провести эксперименты заново. При этом, как уже отмечалось выше, удивляет реакция коллег на такие замечания: «А мы уже публиковали такие результаты в очень солидных журналах, и у нас все принимали без проблем». Проверяем — действительно публиковали, причем в очень «крутых» журналах из Q1. Непонятно, куда в таких случаях смотрели рецензенты и редакторы.

И еще несколько «геронтологических» замечаний (по своей основной специальности я являюсь геронтологом). Как известно, на основании очень дорогих и длительных лонгитудинальных исследований могут быть получены гораздо более интересные и важные результаты, чем в поперечных исследованиях. Именно таким образом выявляются корреляции между различными физиологическими показателями либо привычками людей и их заболеваемостью, смертностью и продолжительностью жизни (долголетием). Однако факторный анализ и статистический анализ получаемых в таких исследованиях результатов должны проводиться очень хорошими специалистами в области математической обработки биомедицинских данных. Себя я точно к ним не отношу, так что мне так и не понятно до конца, почему достаточно небольшие коэффициенты корреляции, получаемые в таких работах, ученые-геронтологи считают заслуживающими внимания. Скажем, мне кажется странным, что по данным «The Duke First Longitudinal Study» удовлетворение работой и карьерой является «строгим предиктором» долголетия для мужчин ($r = 0,28$, что, на мой взгляд, маловато) [26]. Считается, что чем больше объем выборок, тем меньше может быть коэффициент корреляции для выявления представляющих интерес факторов. Однако, к сожалению, в различных работах приводятся совсем разные «барьеры» для размеров выборок, хотя специалисты уже давно достаточно точно определяют их «по Чеддоку» [27].

Впрочем, мне также не очень понятны подходы, которые позволяют многим геронтологам счи-

тать представляющими интерес минимальные достоверные различия кривых выживания контрольной и опытной группы животных в экспериментально-геронтологических исследованиях: «Да, как средняя, так и максимальная продолжительность жизни выросли только на 0,8%, но это увеличение, по данным статистического анализа, достоверно и свидетельствует об эффективности изученного нами геропротектора». Не исключаю, что мне просто не хватает квалификации для понимания такого рода результатов.

Заключение

Итак, ситуация со статистическим анализом данных в биомедицинских (и в частности, в экспериментально-геронтологических) работах, к сожалению, представляется мне неудовлетворительной. В большом количестве исследований полученные результаты либо вообще не подвергаются статистическому анализу, либо анализируются некорректно, а во многих случаях авторы легко заменяют понятие «достоверность» на понятие «тенденция». И это делается, в том числе, в работах, посвященных поиску критически важных препаратов против болезней, уносящих ежегодно миллионы жизней! Ошибочные выводы в таких исследованиях могут нам очень дорого обойтись. И примеров этому можно в Интернете легко найти огромное количество. Чего стоят бесконечные накладные с поисками препаратов против болезни Альцгеймера [28]!

Мне кажется, что нужно больше уделять внимания этому вопросу как рецензентам соответствующих статей, так и редакторам журналов, в которых они публикуются. Кроме того, хотелось бы подчеркнуть необходимость преподавания основ статистической обработки результатов исследований студентам и аспирантам биомедицинских институтов и соответствующих факультетов университетов.

Выражаю свою искреннюю благодарность Г.В. Моргуновой и Н.А. Хохлову за их неоценимую помощь в работе над рукописью статьи. Работа выполнена в рамках государственного задания МГУ имени М.В. Ломоносова без использования животных и без привлечения людей в качестве испытуемых. Автор заявляет об отсутствии у него конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Glantz S.A. *Primer of biostatistics*. 3rd ed. N.Y.: McGraw-Hill; 1992. xvi, 440 pp.
2. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533(7604):452–454. <https://doi.org/10.1038/533452a>
3. Khokhlov A.N. How scientometrics became the most important science for researchers of all specialties.

Moscow Univ. Biol. Sci. Bull. 2020;75(4):159–163. <https://doi.org/10.3103/S0096392520040057>

4. Bland J.M., Altman D.G. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*. 1986;1(8476):307–310.

5. Altman D.G. *Practical statistics for medical research*. NY: Chapman and Hall/CRC, 1990. 624 pp. <https://doi.org/10.1201/9780429258589>

6. Bland J.M., Altman D.G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* 1999;8(2):135–160. <https://doi.org/10.1177/096228029900800204>.
7. Altman D.G. Statistics and ethics in medical research: III How large a sample? *Br. Med. J.* 1980;281(6251):1336–1338. <https://doi.org/10.1136/bmj.281.6251.1336>.
8. Altman D.G. Statistics in medical journals: some recent trends. *Statist. Med.* 2000; 19(23):3275–3289. [https://doi.org/10.1002/1097-0258\(20001215\)19:23<3275::aid-sim626>3.0.co;2-m](https://doi.org/10.1002/1097-0258(20001215)19:23<3275::aid-sim626>3.0.co;2-m).
9. Yan F., Robert M., Li Y. Statistical methods and common problems in medical or biomedical science research. *Int. J. Physiol. Pathophysiol. Pharmacol.* 2017;9(5):157–163.
10. Marino M.J. Statistical analysis in preclinical biomedical research. In: *Research in the biomedical sciences*. Academic Press, 2018, pp. 107–144. <https://doi.org/10.1016/B978-0-12-804725-5.00003-3>
11. Fang J.-Q. (Ed.) *Statistical methods for biomedical research*. Singapore: World Scientific Publishing, 2021. 1160 pp.
12. Padhee S., Ahmed S. Application of statistics in biomedical research. *J. Integr. Med. Res.* 2024;2(2):66–71. https://doi.org/10.4103/jimr.jimr_1_24
13. Ordak M., Paoletti G., Di Bona D., Sousa-Pinto B., Martini M., Bognanni A., Bousquet J., Canonica G.W. Statistical analysis in allergy and immunology: A review with practical examples. *Allergy.* 2025;80(10):2755–2766. <https://doi.org/10.1111/all.70013>
14. Khokhlov A.N. Some brief remarks on statistical significance and “trends” in gerontological research. *Adv. Gerontol.* 2025;15(4):298–301. <https://doi.org/10.1134/S2079057025600788>
15. Khokhlov A.N. Gerontology in the 21st century: From failures to advances. Hopefully. *Adv. Gerontol.* 2023;13(1):1–3. <https://doi.org/10.1134/S2079057024600289>
16. Khokhlov A.N. The renewed ‘Advances in gerontology’ 2025: Results, prospects, changing strategy. *Adv. Gerontol.* 2025;15(2):43–45. <https://doi.org/10.1134/S2079057025600429>
17. Kirpichnikov M.P., Morgunova G.V., Khokhlov A.N. Our journal—2020: what and how we publish. *Moscow Univ. Biol. Sci. Bull.* 2020;75(1):1–6. <https://doi.org/10.3103/S0096392520010034>
18. Morgunova G.V., Khokhlov A.N. 75 years of the journal Vestnik moskovskogo universiteta and 45 years of its biological series. *Moscow Univ. Biol. Sci. Bull.* 2022;77(1):1–5. <https://doi.org/10.3103/S0096392522010023>
19. Khokhlov A.N., Morgunova G.V. Is it worth teaching biology students the basics of scientometrics and the instructions for the design of scientific articles, and if so, why? *Moscow Univ. Biol. Sci. Bull.* 2021;76(3):77–82. <https://doi.org/10.3103/S0096392521030081>
20. Ермолаев-Томин О.Ю. *Математические методы в психологии: учебник для бакалавров*. 4-е изд., пер. и доп., М.: Издательство Юрайт, 2012. 511 с.
21. Benjamin D.J., Berger J.O., Johannesson M. et al. Redefine statistical significance, *Nat. Hum. Behav.* 2018;2(1):6–10. <https://doi.org/10.1038/s41562-017-0189-z>
22. Khokhlov A.N. Cell kinetic approaches to the search for anti-aging drugs: Thirty years after. *Moscow Univ. Biol. Sci. Bull.* 2018;73(4):185–190. <https://doi.org/10.3103/S0096392518040041>
23. Khokhlov A.N., Morgunova G.V., Klebanov, A.A. Demographic approaches to the study of aging on cell cultures. *Moscow Univ. Biol. Sci. Bull.* 2019;74(4):262–267. <https://doi.org/10.3103/S0096392519040060>
24. Khokhlov A.N., Morgunova G.V., Scientific publications – the bad, the good, for a fistful of dollars. *Sci. Ed. Publ.* 2021;6(1):59–67. <https://doi.org/10.24069/2542-0267-2021-1-59-67>
25. Faul F., Erdfelder E., Buchner A., Lang A.-G. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods.* 2009;41(4):1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
26. Palmore E.B. Predictors of the longevity difference: A 25-year follow-up, *The Gerontologist.* 1982;22(6):513–518. <https://doi.org/10.1093/geront/22.6.513>
27. Chaddock R.E. *Principles and methods of statistics*. Houghton Mifflin; 1925. xvi, 471 pp.
28. Khokhlov A.N. Why freshwater hydra does not get Alzheimer’s disease. *Moscow Univ. Biol. Sci. Bull.* 2023; 78(3):198–204. <https://doi.org/10.3103/S0096392523700104>

Поступила в редакцию 02.09.2025

После доработки 20.11.2025

Принята в печать 05.12.2025

OPINION ARTICLE

Statistical analysis in biomedical research: From significance to “trends”?

A.N. Khokhlov 

*Evolutionary Cytoogerontology Sector, School of Biology, Lomonosov Moscow State University,
1–12 Leninskie gory, Moscow, 119234, Russia*

**e-mail: khokhlov@mail.bio.msu.ru*

Drawing on his many years of experience editing for biomedical journals and lecturing on scientific article formatting for PhD students at Lomonosov Moscow State University, the author analyzes the current state of statistical analysis of biological, biomedical, and medical research results — in particular, data obtained in studies aimed at discovering drugs that supposedly slow aging and suppress the development of age-related diseases. He emphasizes that in many cases, such analysis is either not performed at all or is performed incorrectly. Authors often analyze

data that does not follow a normal distribution using the parametric Student's t-test. Moreover, many researchers, having failed to obtain reliable results ($p > 0.05$), are increasingly replacing the term "significance" in their articles with "trend," which seems completely unacceptable. This problem affects not only biomedical but also other natural science research, but it is especially important for biomedicine, as it is at stake in human life and health. It is proposed that greater attention should be paid to this issue by both reviewers of relevant articles and editors of the journals in which they are published. Furthermore, the importance of teaching the basics of statistical processing of research results to undergraduate and graduate students in biological, biomedical, and medical institutes, as well as relevant university schools, is emphasized.

Keywords: *biomedical research, statistical analysis, parametric and nonparametric tests, sample sizes, survival curves, significance level, geroprotectors, experimental gerontology*

Funding: The study was conducted under the state assignment of Lomonosov Moscow State University.

Сведения об авторе

Хохлов Александр Николаевич — докт. биол. наук, зав. сектором эволюционной цитогеронтологии биологического факультета МГУ. Тел.: 8-495-939-15-90; e-mail: khokhlov@mail.bio.msu.ru; ORCID: <https://orcid.org/0000-0001-7454-7023>