

◆ РЕДАКЦИОННЫЕ МАТЕРИАЛЫ

© В. В. ФАДЕЕВ. 2002

В. В. Фадеев

ПРЕДСТАВЛЕНИЕ ДАННЫХ В ОРИГИНАЛЬНЫХ РАБОТАХ И ИХ
СТАТИСТИЧЕСКАЯ ОБРАБОТКА

Журнал "Проблемы эндокринологии" был основан в 1955 г. и, как это указывается в периодически публикуемых в журнале памятках "К сведению авторов", в нем издаются статьи по клинической и экспериментальной эндокринологии. На сегодняшний день журнал "Проблемы эндокринологии" фактически является основным печатным изданием российских эндокринологов. Такое положение журнала диктует необходимость поддержания высоких требований к выходящим в нем статьям.

До настоящего времени требования к оформлению статей, которые публиковались в разделе "К сведению авторов", излагались достаточно кратко. Так, требования к статистической обработке представляемых в редакцию статей фактически ограничивались фразой: "Нужно указать, являются ли приводимые числовые значения первичными или производными, результатом одного наблюдения или целой серии, привести пределы точности, надежности, интервалы достоверности, оценки, рекомендации, принятые или отвергнутые гипотезы, обсуждаемые в статье".

Как показывает практика научного редактирования поступающих в редакцию журнала статей, требований, приведенных в столь телеграфном варианте, недостаточно. Часть статей приходится возвращать авторам в связи с достаточно стандартными погрешностями в оформлении или представлении результатов.

В дальнейшем редакция журнала "Проблемы эндокринологии" планирует опубликовать более подробные требования к представляемым в журнал статьям. Они будут базироваться на положениях, принятых в международной медицинской научной литературе. В настоящее время более 500 научных журналов приняло 5-ю редакцию "Единых требований к рукописям, представляемым в биомедицинские журналы", которые были разработаны "Международным комитетом редакторов медицинских журналов" [6].

Эта небольшая статья призвана обсудить наиболее типичные погрешности представления данных, с которыми постоянно приходится сталкиваться при научном редактировании. Мы не ставили перед собой цель разбирать принципы планирования эксперимента и статистической обработки результатов медико-биологических исследований. Наша цель — повышение методологического уровня статей и, таким образом, повышение престижа журнала. В этом плане наибольшее внимание необходимо обратить на статистическую обработку данных, которая, как об этом свидетельствует анализ многих работ, опубликованных в крупных отечественных журналах, представляет существенные проблемы [2–4].

В первую очередь хотелось бы остановиться на разделе "Материалы и методы", который должен присутствовать в любой оригинальной работе. Непременным компонентом этого раздела является описание конкретных методов статистической обработки (если речь не идет об описании серии клинических случаев). Широко используемые фразы типа: "Обработка производилась общепринятыми статистическими методами" — неприемлемы. Статистическая обработка является не вспомогательным, а основным инструментом, с помощью которого в медико-биологических исследованиях делаются те или иные предположения и выводы.

Следует учесть, что похожие названия могут иметь принципиально разные методы статистического анализа. Например, точный критерий Фишера может быть одно- и двусторонним, при анализе повторных изменений в одной и той же группе используется не привычный всем, а парный *t*-тест Стьюдента. Не говоря уже о том, что фраза типа "проводился корреляционный анализ" мало что говорит читателю, поскольку последний может подразумевать принципиально различные тесты: Пирсона, Спирмена и др.

Поскольку в большинстве случаев статистическая обработка не ограничивается использованием какого-либо одного метода, в разделе "Материалы и методы" рекомендуется поместить общее описание использованных тестов. Конкретные же методы статистического анализа целесообразно указывать по ходу изложения материала в разделе "Результаты". В большинстве же представляемых рукописей в тексте статьи присутствуют указания лишь на тот факт, что " $p < 0,05$ " или " $p < 0,01$ " и т. д.

Возвращаясь к разделу "Материалы и методы", отметим, что в нем целесообразно указать задачи, которые ставились при анализе данных, например [4] описание изучаемых показателей в группах (использована только описательная статистика); парные групповые сравнения; исследование связей двух или нескольких показателей; анализ дожития.

Если речь идет о проспективном клиническом исследовании, необходимо указать принципы разделения пациентов на группы, в частности, проводилась ли рандомизация, а если да, то каким методом. Далее, в случае контролируемого исследования, необходимо указать, проводился ли статистический анализ данных только тех больных, которые закончили лечение, или всех включенных в исследование, поскольку при значитель-

ной доле выбывших может существенно измениться сопоставимость групп [5].

Следует уделить особое внимание проблеме множественных сравнений [1]. Дело в том, что чем больше статистических сопоставлений проводится на одних и тех же данных, тем больше вероятность того, что значение " p " (вероятность справедливости нулевой гипотезы) может быть ошибочно принято за отражающее истинную биологическую закономерность. Наиболее типичной ошибкой являются попарные сравнения в рамках 3 групп и более с помощью *t*-теста Стьюдента (см. ниже). Следовательно, в разделе "Материалы и методы" целесообразно отметить, приводятся ли в работе все сопоставления, которые выполнены в ходе анализа данных. В случае попарных сравнений необходимо указать на соблюдение соответствующих поправок (например, поправка Бонферрони, критерий Ньюмена–Кейлса и др.).

В настоящее время для статистической обработки рекомендуется использовать общеизвестные статистические компьютерные программы, поскольку ручной подсчет и использование самостоятельного разработанного программ не обеспечивают должного уровня точности и воспроизводимости. Таким образом, в разделе "Материалы и методы" необходимо указать конкретный программный пакет и его производителя, например: Statistica ("StatSoft", США), SPSS ("SPSS Inc", США), STATGRAPHICS ("Manugistic Inc.", США) и т. д. Нередко удивляет упоминание авторами факта использования компьютерных программ наряду с ошибками в написании их названий ("Exell" или "STATGAF") и типичными некоррекциями в использовании статистических методов. Очевидно, что сам факт использования компьютера, во-первых, уже не повышает престижа работы, во-вторых, не страхует от ошибок и не может их нивелировать.

При описании лабораторных методов необходимо четко указать название коммерческих наборов для определения гормонов и привести нормативы гормональных показателей. Для инструментальных методов необходимо указать название и основные характеристики приборов (марку ультразвукового аппарата и т. д.).

Другой проблемой, на которой хотелось бы остановиться, является представление данных в разделе "Результаты".

Описание выборок медико-биологических исследований, как известно, зависит от характера распределения показателей. В большинстве направляемых в журнал работ данные представлены в виде " $X \pm x$ ". Что авторы понимают под этим, в большинстве случаев не указывается. Оказывается, что X — почти всегда — среднее арифметическое (M), а x — его стандартная ошибка (m).

При таком варианте изложения (если авторы все-таки указали на это) возникает ряд проблем. Во-первых, среднее арифметическое адекватно характеризует выборку только в случае распределения близкого к нормальному. Оценка же нормальности распределения производится и описывается в единичных случаях. Еще более проблематично использование для описания выборки стандартной ошибки средней, поскольку эта ошибка характеризует само среднее арифметическое, но никак не разброс показателей в выборке (!), который наиболее интересен читателям. Другими словами, описание " $M \pm m$ " не позволяет читателям оценить однородность выборки. В случае нормального распределения однородность выборки характеризует стандартное отклонение SD .

Соблазн привести m вместо SD очевиден. Во-первых, m по определению (!) меньше SD , т. е. приводя m , автор "демонстрирует" большую однородность выборки. С другой стороны, автор может столкнуться с ситуацией, когда SD окажется больше M (данные, например, будут выглядеть как 45 ± 100 мл). Такая ситуация характерна для распределения, отличающегося от нормального. Не зная, как описывать такую выборку, авторы очень часто приводят " $M \pm m$ " и избегают заведомо некорректной ситуации.

По современным представлениям, большинство распределений медико-биологических показателей не являются нормальными. Особенно низка вероятность того, что нормальному распределению будут подчиняться показатели в малых выборках. Для описания распределений, не являющихся нормальными, используют медиану (Me) и интерквартильный размах, как правило, в виде 25-го и 75-го процентилей. Например: 45 (22,84) — Me , 22 — 25-й, а 84 — 75-й перцентиль.

Следует иметь в виду, что при округлении числовых данных в средних значениях (или медиане) принято приводить не более чем на 1 десятичный знак больше, чем в исходных данных, а в SD (или Mo) — не более чем на 2 знака больше, чем в исходных данных. Например, среднее артериальное давление должно описываться не как $80,8796 \pm 10,3321533$ мм рт. ст., а как $80,9 \pm 10,33$ мм рт. ст., а лучше даже $81 \pm 10,3$ мм рт. ст. (кроме того, очевидно, что сотые доли мм рт. ст. при обсуждении артериального давления абсолютно лишены клинического значения).

Если число наблюдений в исследовании мало (менее 20—15), т. е. речь идет о малой выборке, использования описательной статистики следует избегать. Как это ни кажется абсурдным, но приходится сталкиваться со средним и стандартным отклонением по выборке из 3 значений. Конечно, теоретически среднее арифметическое и стандартное отклонение может быть рассчитано и для 2 значений, одно из которых 1, а другое 0, однако в любой из описанных ситуаций полученные показатели могут исказить представление о выборке. Таким образом, в малых выборках, которые анализируются достаточно часто, по возможности целесообразно приводить первичные данные.

Похожая ситуация возникает при использовании процентов. Проценты используют для наглядного сопоставления различных по объему выборок. Если объем выборки больше 100 наблюдений, рекомендуется указывать процент не более чем с одним десятичным знаком. Если объем выборки меньше 100, то проценты указываются только целыми числами.

Если выборка меньше 20, то проценты использовать нецелесообразно. В этой ситуации процент будет значительно больше, чем само число (получается описание типа: "50% больных выздоровели, а еще у 1 процесс приобрел хроническое течение"). При малых выборках приводятся только абсолютные числа наблюдений. В целом при использовании процентов необходимо указывать числители и знаменатели, с применением которых проценты вычислены: % (n/N) [4].

Задачей этой статьи является краткий разбор наиболее типичных неточностей, которые содержат статьи, направляемые в редакцию журнала. В большей степени эта статья ориентирована на то, чтобы заострить внимание авторов на таком важном аспекте, как представление данных и их статистическая обработка. Мы не будем здесь описывать плюсы и минусы тех или иных методов статистической обработки, но использование одного и наиболее излюбленного метода мы проанализируем. Речь идет о *t*-тесте Стьюдента.

Как показывает анализ отечественных публикаций [2], этот метод для проверки гипотез использует подавляющее большинство авторов. Фактически *t*-тест затмил собой все многообразие методов, которое может предложить статистика. Наряду с этим складывается ситуация, которую автор недавно переведенного на русский язык руководства по медицинской статистике С. Гланц [1] описывает следующим образом: "...Результаты многих биологических и медицинских работ основаны на неправильном использовании статистических методов и способны только ввести в заблуждение. Большинство ошибок связано с неправильным использованием критерия Стьюдента". Этим мы хотим подчеркнуть, что проблема носит в какой-то мере интернациональный характер. В одном из опросов врачей-исследователей выяснилось, что большая часть из них не знает, каковы условия использования этого столь популярного теста [2].

Итак, во-первых, *t*-тест используется только для сравнения количественных признаков, но не качественных (стандартная ситуация: в разделе "Материалы и методы" указано, что "для статистической обработки использовался *t*-тест". После этого в тексте статьи идет описание napодобие: "распространенность явления "А" в основной группе 80%, а контрольной — 10%; $p < 0,05$ ". Получается, что с помощью *t*-теста сравнивали качественные данные). Для сопоставления качественных признаков используют совершенно другие методы ("Хи-квадрат", точный критерий Фишера и др.).

Во-вторых, *t*-тест (как и все другие параметрические методы) может быть использован только для сравнения групп с нормальным распределением, т. е. перед использованием *t*-теста необходимо с помощью ряда тестов (Колмогорова—Смирнова, Шапиро—Уилка и др.) удостовериться в нормальности распределения в группах.

В-третьих, *t*-тест правомерно использовать при равенстве дисперсий в выборках, т. е. когда разброс данных в выборках примерно одинаков. Наконец, в-четвертых, *t*-тест предназначен для сравнения двух групп и не может (!) использоваться для попарного сравнения 3 групп и более (см. выше). В последней ситуации из параметрических тестов может быть использован дисперсионный анализ (ANOVA), непараметрическим аналогом которого является критерий Крускала—Уоллиса.

Эти достаточно жесткие ограничения в последние десятилетия делают все более популярным использование непараметрических тестов (критерии Манна—Уитни, Уилкоксона, Крускала—Уоллиса и др.). Эти тесты лишь немногим уступают по мощности параметрическим, но не требуют какого-либо типа распределения. Как показал анализ данных литературы, непараметрические методы в отечественных оригинальных работах используются очень редко [2].

В заключение хотелось бы остановиться на использовании в статьях статистических терминов. В упоминавшихся в начале этой статьи "Единых требованиях..." [6] в разделе "Статистика" указано: "Избегайте употребления статистических терминов, таких как "рандомизированный" (что

означает случайный способ отбора), "значимый", "корреляции" и "выборка", для обозначения нестатистических понятий".

В статьях, поступающих в редакцию, часто встречается либо использование статистических терминов (чаще "корреляция") вне какой-либо связи с проведенным статистическим анализом, либо подмена правильных статистических терминов теми, которые глобализуют простые статистические выкладки. Наиболее типичным примером последнего феномена является полное вытеснение термина "статистически значимый" различными вариантами слова "достоверный" или даже "высокодостоверный".

Совершенно очевидно, что если исследователю с помощью того или иного статистического метода удалось отвергнуть нулевую гипотезу об отсутствии различий между выборками с тем или иным уровнем значимости (например, всеми любимым $p < 0,05$), это свидетельствует лишь об одном: есть некое косвенное подтверждение того, что исследованные выборки различаются (т. е. не принадлежат к одной большой выборке) по исследованному показателю. Поскольку эти различия были подтверждены статистическими методами, они являются "статистически значимыми". Статистическая значимость различий не может свидетельствовать о неких достоверных клинических различиях и не несет никакой смысловой, в данном случае клинической нагрузки, т. е. наличие статистически значимых различий или статистически значимой корреляции не свидетельствует о прямой причинно-следственной, а особенно патогенетической связи явлений.

Причина, по которой русским эквивалентом английского термина "significance" стал термин не "статистически значимый", а "достоверный", во многом ясна. Часто авторы работ подосознательно или намеренно стремятся отождествить неслучайность выявленных различий (только об этом и свидетельствует "статистическая значимость") с некоей достоверностью, т. е. по сути с биологической или патогенетической закономерностью.

Некорректных вариантов использования авторами статей слова "достоверность" применительно к статистическому анализу очень много: "...достоверность определялась по критерию Стьюдента", "...высокая достоверная корреляционная связь...", "...высокодостоверное увеличение...", "...статистически достоверно большими были...", "...коэффициенты достоверно отличались друг от друга...". Иногда можно встретить откровенный набор слов napодобие: "Вероятность случайности различий соответствует достоверности".

Наиболее ярким примером попытки глобализации полученных данных является попытка полной подмены статистического контекста клиническим. Так, вместо термина "статистически значимые различия" можно встретить фразу "обнаружены статистически и клинически достоверные различия".

Замена термина "статистическая значимость" (2 слова обязательно вместе) на "достоверность" применительно к статистическому анализу неприемлема.

Примером того, как статистическое понятие некорректно применяется вне контекста статистического анализа, является использование терминов "корреляция" и "коррелирует". Очевидно, что в быту и в разговорной речи эти термины часто используют, чтобы просто обозначить взаимосвязь между явлениями. Такой вариант абсолютно неприемлем в научной литературе, поскольку термин "корреляция" имеет совершенно определенное значение, а сам корреляционный анализ является базовым методом проверки гипотез о взаимосвязи признаков в медико-биологических исследованиях. Например, авторы обнаружили, что некий феномен "А" встречается чаще в группе с более тяжелым течением какого-то заболевания. Далее, среди прочих рассуждений, используется фраза типа: "А" достоверно коррелирует с тяжестью заболевания". Корреляция может быть статистически значимая, при этом упоминать ее можно лишь на основании проведенного корреляционного анализа.

ЛИТЕРАТУРА

1. Гланц С. Медико-биологическая статистика: Пер. с англ. — М., 1999.
2. Леонов В. П. // *Международ. журн. мед. практики.* — 1998. — № 4. — С. 7—12.
3. Леонов В. П. // Там же. — 1999. — № 4. — С. 7—19.
4. Реброва О. Ю. // Там же. — 2000. — № 4. — С. 43—45.
5. Флетчер Р., Флетчер С., Вагнер Э. Клиническая эпидемиология. Основы доказательной медицины: Пер. с англ. — М., 1998.
6. International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals // *Ann. Intern. Med.* — 1997. — Vol. 126. — P. 36—47.

Поступила 25.06.01