

2' 2007

МЕЖДУНАРОДНЫЙ ЖУРНАЛ МЕДИЦИНСКОЙ ПРАКТИКИ

International Journal of Medical Practice

ТЕМА НОМЕРА:

ФАЛЬШИВКИ



Издательство Медиа Сфера

2
2007

МЕЖДУНАРОДНЫЙ ЖУРНАЛ МЕДИЦИНСКОЙ ПРАКТИКИ

Главный редактор В.В. Власов

Зав. редакцией А.В. Полунина

Редактор: А.В. Полунина

Переводчики: Е.А. Султанова, Е.А. Фролова,
И.В. Журавлев, В.В. Забненкова

Адрес для корреспонденции:

127238 Москва, а/я 54, Медиа Сфера

Тел.: (495) 482-4329, Факс: (495) 482-4312.

www.mediasphera.ru; e-mail: mjmp@mediasphera.ru

Оригинал-макет изготовлен издательством МЕДИА СФЕРА

Компьютерная верстка: М.Л. Калужнин, Е.Л. Коган, С.В. Олефир

Корректоры: Е.А. Папоян, В.Ю. Глазунова, И.В. Корягина



Издательство Медиа Сфера

127238 Москва, Дмитровское шоссе, 46, корпус 2, этаж 4.

Отдел подписки и распространения: Тел/Факс: (495) 488-6637

Отдел рекламы: Тел.: (495) 488-6000, Факс: (495) 482-4312

© Издательство Медиа Сфера

Отпечатано:

ФГУП Издательство "Известия"

ОШИБКИ СТАТИСТИЧЕСКОГО АНАЛИЗА БИОМЕДИЦИНСКИХ ДАННЫХ

В.П. Леонов

Томский государственный медицинский университет

В статье рассмотрено много примеров ошибочного использования статистики в российской биологии и медицине. Дана классификация основных видов ошибок, обсуждаются обзоры по этой теме за предыдущие 50 лет. Дана классификация причин возникновения таких ошибок в биологии и медицине. С позиций меметики рассмотрены более 1500 статей, монографий и диссертаций, содержащих типичные ошибки использования и описания статистики. Приведены цепочки мемов с ошибочными описаниями методов сравнения средних и статистической значимости результатов и доверительной вероятности. Показано существование устойчивых мемов ошибочного описания статистики внутри отдельных научных школ, например в Сибирском государственном медицинском университете. Вводится новый тип мема — мем камуфляжного описания, используемый для маскировки неправильного применения статистики в медицинском исследовании. Установлено, что наиболее часто встречаются случаи ошибочного использования *t*-критерия Стьюдента, без необходимой проверки обязательных для этого условий. Наиболее часто подобные ошибки встречаются в публикациях журнала «Бюллетень экспериментальной биологии и медицины». Достаточно часто в российских публикациях приводятся описания некорректного применения многомерных статистических методов, таких как факторный и кластерный анализ. На основе анализа причин живучести таких ошибок в биомедицине, даны конкретные рекомендации по разрешению этой проблемы. Основным средством для этого является создание в медицинских исследовательских организациях лабораторий биостатистики.

In article many examples of mistakes of use of statistics in the Russian biology and medicine are analysed and classification of such mistakes is given. Such reviews for the previous 50 years are discussed. The reasons of occurrence of such mistakes in biology and medicine are classified. From positions of the Memetics are considered more than 1500 articles, monographies and the dissertations containing typical mistakes of use and the description of statistics. Chains of the Memes with erroneous descriptions of methods of comparison of average and are given the statistical importance of results and confidential probability. Existence of steady the Memes of the erroneous description of statistics inside separate scientific schools, for example at the Siberian state medical university is shown. The new type of the Meme — the Meme of the camouflage description used for masking of wrong application of statistics in medical research is entered. It is established, that most frequently there are cases of erroneous use of *t*-criterion of the Student, without necessary check of obligatory conditions for this purpose. Most frequently similar mistakes meet in publications of magazine «The Bulletin of experimental biology and medicine». There is enough frequently in the Russian publications descriptions of incorrect application of multivariate statistical methods, such as factorial and cluster the analysis are resulted. On the basis of the analysis of the reasons of survivability of such mistakes in biomedicine, concrete recommendations for the sanction of this problem are given. The basic means for this purpose is creation in the medical research organizations of laboratories of biostatistics.

Как известно, не ошибается лишь тот, кто ничего не делает. Впрочем, и это утверждение тоже не абсолютно. Ошибки статистического анализа биомедицинских данных стали объектом изучения с момента появления этого компонента научной деятельности. Неправильное использование статистического аппарата приводит к неверным выводам. Эти ошибки особенно часто происходят в медицине: врачи не блещут математическими знаниями. Жертвами таких неверных выводов впоследствии становятся больные, и таких жертв немало, хотя оценить их число совсем непросто. В статье даны многочисленные примеры таких ошибок, их анализ, классификация, обсуждаются причины их появления. На примере диссертаций, утвержденных Высшей аттестационной комиссией РФ (ВАК), опубликованных монографий и статей в журналах дан меметический анализ описаний методов статистики, показана локализация мемов ошибок статистического анализа внутри отдельных научных школ. Приведены примеры камуфляжных описаний этапа статистического анализа.

За рубежом данная проблема обсуждается уже давно. Статистическая непорядочность медиков считается профессиональной негодностью, поскольку представляет для общества не меньшую опасность, чем любая иная непорядочность. Во многих европейских странах уже введены обязательные стандарты этики исследований. В России во времена лысен-

ковщины одним из первых борьбу со статистической непорядочностью начал академик А.Н. Колмогоров. В современной российской биомедицине данная проблема весьма актуальна, поскольку ни на одном из этапов проведения научных исследований в этой отрасли знания нет механизмов, гарантирующих качество статистического анализа экспериментальных данных. Отсутствие таких механизмов объясняется конфликтом интересов участников этих исследований и общества, которое финансирует эти исследования. Ситуация усугубляется также инертностью руководителей соответствующих вузов, научно-исследовательских институтов (НИИ), ВАК и редакций биомедицинских периодических изданий, усматривающих в решении этой проблемы угрозу собственным корпоративным интересам. На основе анализа причин живучести таких ошибок в биомедицине даны конкретные рекомендации по разрешению этой проблемы.



ОШИБКА — ЭТО.....

Ошибки не есть лженаука. Лженаука — это непризнание ошибок!

П.Л. Каница

«Словарь русского языка» С.И. Ожегова (М: Русс. яз. 1986) определяет это слово так: «ОШИБКА. Не-

правильность в действиях, мыслях». Действительно, очень часто ошибки в выполнении статистического анализа биомедицинских данных являются в первую очередь следствием неправильных мыслей непосредственных исполнителей этого анализа и тех, кто собирал эти данные.

«Часто вспоминают слова Дизраэли: «Существуют три вида лжи: просто ложь, наглая ложь и статистика». При этом под статистикой обычно понимают ложные наборы статистических данных, взятые чаще всего с потолка и ставящие целью подкрепить чьи-то предвзятые суждения. Но бывают и другие случаи, когда используются вполне доброкачественные наборы данных и никаких дурных целей не ставят. Однако неправильное использование статистического аппарата приводит к неверным выводам. Эти ошибки особенно часто происходят в медицине: ведь известно, что врачи не блещут математическими знаниями и поэтому легко впадают в ошибки, когда пользуются статистическими методами. Их жертвами становятся больные, и таких жертв немало, хотя оценить число таких жертв совсем непросто» [1].

М. Рейтман выделяет два вида статистических ошибок: 1) заведомо ложные данные, результаты анализа которых также заведомо ложны; 2) ложные результаты как следствие неправильного использования статистического инструментария. За 20 лет статистического анализа биомедицинских данных автору этих строк достаточно редко приходилось сталкиваться с первым видом статистических ошибок. И не потому, что они действительно редки в реальной жизни. Просто заведомо ложные, специально подтасованные данные гораздо труднее, а подчас и невозможно идентифицировать. Это обусловлено тем, что сами исходные данные, на основе которых авторы получают публикуемые результаты, обычно недоступны рядовому читателю.

ОШИБКА ИЛИ ОБМАН?

Тьмы низких истин мне дороже
нас возвышающий обман.

А.С. Пушкин

«В науке, как в любой другой профессиональной области, жулики были, есть и, увы, будут. Правде надо смотреть в глаза, какой бы горькой она ни казалась. Раньше у нас главными побудительными мотивами для научного мошенничества были карьеристские устремления и страх. (...) Взамен появился новый, более коварный стимул для слабых волей научных сотрудников. Это стремление во что бы то ни стало получить грант или финансирование собственных исследований. По мере интеграции отечественной науки в мировую эта опасность будет усугубляться. И чем раньше мы перейдем к западному опыту борьбы с тамошними жуликами от науки, тем будет лучше» [2].

В статье [2] приводятся примеры жульничества: «...согласно анамнезу, 17-летний юноша, страдающий необычным сердечным недугом, имел четырех детей: 4, 5, 7 и 8 лет. Это была не описка, ибо цифры фигурировали и в таблице, и в тексте». Это, конечно, довольно грубая фальсификация. Но известны и весьма изощренные примеры: «Из 18 статей Дарси в 13 его со-

авторами были либо научный руководитель, либо те, от кого зависело получение гранта. «Нужные люди», как правило, не вникали в суть работы» [2]. Однажды и в нашей практике был случай, когда аспирант Сибирского государственного медицинского университета (Томск), занимавшийся исследованием биохимических показателей у наркоманов, пытался фальсифицировать исходные данные.

В подавляющем же большинстве случаев наблюдаются ошибки, отнесенные во вторую группу, — неправильное использование статистического инструментария, и, как следствие, ложные выводы.

Широко известен случай, когда автор статьи вместе с результатами ошибочно выполненного анализа опубликовал и исходные данные. Это помогло в дальнейшем опровергнуть его выводы. В 1939 г. в журнале «Яровизация» была опубликована статья Н.И. Ермолаевой — аспирантки Лысенко. Работа называлась «Еще раз о «гороховых законах» [3]. Академик А.Н. Колмогоров проанализировал данные и опубликовал статью «Об одном новом подтверждении законов Менделя» [4]. В ответной публикации «По поводу статьи академика А.Н. Колмогорова» [5] Т.Д. Лысенко выводы А.Н. Колмогорова оценил так: «Я не чувствую себя достаточно компетентным, чтобы разбираться в этой системе математических доводов и доказательств. К тому же меня как биолога сейчас не интересует, хорошим или плохим математиком был Мендель. Свою же оценку статистических работ Менделя я уже неоднократно освещал в печати, заявляя, что эти работы никакого отношения к биологии не имеют... В данной заметке мне хочется лишь указать, что названная выше статья известного математика А.Н. Колмогорова тоже никакого отношения к биологической науке не имеет. ... Поэтому-то нас, биологов, и не интересуют математические выкладки, подтверждающие практически бесполезные статистические формулы менделистов».

После августовской сессии ВАСХНИЛ 1948 г. гонению подверглась не только генетика, но и непосредственно статистика [6]. Нападки на статистику, один из основных инструментов генетики, сразу дали ожидаемый результат. Биологи и медики тех лет поняли, что использовать статистику опасно, т.к. могут причислить к менделистам-морганистам, обвинить в преклонении перед иностранщиной и в космополитизме. Это была вполне реальная опасность: за использование статистики в биомедицинских диссертациях стали даже отказывать в присуждении ученых степеней. Один из таких примеров мы можем найти в статье секретаря Фрунзенского райкома ВКП(б) г. Москвы Е.А. Фурцевой (будущего министра культуры СССР) [6]. Последствия этих гонений российская биомедицина пожинает до сих пор.

Неизбежны ли ошибки статистического анализа биомедицинских данных? Как известно, не ошибается только тот, кто ничего не делает. Ошибки встречаются везде, будь то наука, техника или политика. Задача исследователя — свести к минимуму возможность появления ошибок. Очевидно, что одним из наиболее разумных путей уменьшения числа ошибок в медицинской науке должно быть выполнение этапа статистического анализа профессионалом в данной

области, т.е. биостатистиком. За рубежом для этих целей существуют многочисленные лаборатории биостатистики [7]. Как и в любом ином деле, привлечение профессионалов само по себе является определенной гарантией качества конечного результата.



ОБЗОРЫ ОШИБОК ПРИМЕНЕНИЯ СТАТИСТИКИ В МЕДИЦИНЕ

Сомнительный результат не может признаваться в качестве научного.

Н. Бор

В отечественной медицине одно из первых исследований, содержащее большой и нелицеприятный обзор таких ошибок — книга А.Я.Боярского «Статистические методы в экспериментальных медицинских исследованиях», изданная более 50 лет назад [8]. Автор детально проанализировал большое количество статей, опубликованных в течение нескольких лет в известных советских медицинских журналах. По сути дела это было **первое отечественное исследование, в котором автор на значительном материале обнаружил крайне неблагоприятную ситуацию с применением статистики в экспериментальной медицине и биологии.**

«Уже беглое ознакомление с состоянием дела показывает, что статистическая обработка экспериментальных данных является наиболее слабым местом во многих исследованиях. (...) Трудно требовать от медика, чтобы он, наряду со знаниями в своей собственной области, был в то же время достаточно вооруженным, скажем, в радиотехнике, для конструирования аппаратуры, улавливающей биотоки, или в статистике, для нахождения наиболее правильных методов статистической обработки своих экспериментальных данных. И подобно тому, как медику, несомненно, приходится обращаться за содействием к радиотехнику, для правильной статистической обработки экспериментальных данных нередко приходится обращаться к специалисту-статистику. (...) **Так или иначе, но бесспорным фактом являются и недостаточная вооруженность медиков статистическими знаниями, и недостаточно высокий научный уровень статистической методики в большинстве их экспериментальных работ**» [8].

Одна из первых зарубежных публикаций, содержащая анализ статистических ошибок в медицинских статьях, относится к 1929 г. [9]. В этой статье сообщается, что примерно половина статей, публикуемых в журнале *Physiological Reviews*, содержит примеры ошибочного использования статистики. Результаты более общего исследования были опубликованы в 1932 г. [10]. Начиная с 60-х годов XX века публикуются такие обзоры [11–22]. Большую работу в этом направлении проводит статистик Дуглас Альтман (Douglas G. Altman) [22–31]. В этих обзорах отмечается рост доли публикаций, в которых используется статистика, а также уменьшение доли ошибочного применения этих методов. Такая тенденция характерна не только для европейских, но и для китайских журналов [19].

Следует отметить, что в российской медицинской периодике подобные публикации крайне редки. Не

потому что такие исследования не ведутся, а потому что не каждый журнал отважится опубликовать статью, в которой будет дан критический анализ его содержания. Правда, имеются примеры и иного отношения к критике. Так в начале 1997 г. автор этих строк направил в журнал «Кардиология» рукопись «Применение методов статистики в кардиологии (по материалам журнала «Кардиология» за 1993–1995 гг.)». Поскольку редакция журнала в течение года ответа не дала, то я решил, что рукопись отвергнута. Однако в апреле 2004 г. на семинаре по биометрике в Самаре одна из слушательниц передала мне копию опубликованной статьи [32].



ПРИЧИНЫ ВОЗНИКНОВЕНИЯ ОШИБОК И МЕРЫ БОРЬБЫ С НИМИ

Как он дышит, так и пишет...

Б. Окунджава

Большая распространенность ошибочного использования статистики в медицинских исследованиях имеет несколько причин. Сформулируем кратко лишь основные из них. То что такие ошибки проникают на страницы российских журналов и диссертаций, свидетельствует, что на этапах рецензирования этих работ отсутствует контроль качества статистического анализа профессиональными биостатистиками. Гораздо более высокое качество статистических выводов в статьях зарубежных медицинских журналов обеспечивается именно наличием такого статистического рецензирования [15, 30, 33, 34]. Например, во всемирно известном журнале *BMJ* такое рецензирование проводит группа специалистов, в состав которой входит и упоминавшийся выше Д. Альтман.

Точно так же статистики не участвуют и в рецензировании диссертаций. Однако причины возникновения таких ошибок действуют и на более раннем этапе исследований. Большинство авторов — не профессиональные исследователи. Они прежде всего клиницисты, и поэтому не должны иметь исчерпывающей статистической подготовки. Этот же вывод можно сделать и на основе содержания квалификационной характеристики биолога [35]. Во всех аналогичных квалификационных характеристиках медицинских специальностей в разделе «Знание смежных дисциплин» вообще отсутствует упоминание математики и методов исследований. И это вполне разумно, поскольку медицинский вуз не может, да и не должен обеспечивать своим выпускникам профессиональный уровень владения биостатистикой. Его задача — подготовить хороших врачей, лекарей, а не исследователей.

Д. Альтман обращает внимание на этический аспект этой проблемы [29]. Если врач использует неправильную технологию лечения или дает неправильную дозу лекарства, то большинство его коллег сочтут такое поведение непрофессиональным и неэтичным. А что думать о враче-исследователе, который допускает статистическую неряшливость? Весьма обстоятельно данный этический аспект рассмотрен в работе [21], в которой утверждается, что **статистическая непорядочность медиков является профессиональной некомпетентностью**.

петентностью и представляет для общества не меньшую опасность, чем любая иная непорядочность. В связи с этим хочу обратить внимание читателей на публикацию [36]. «Спустя 5 лет после крупного скандала, потрясшего научное сообщество, германские университеты готовы вот-вот принять свои первые обязательные **стандарты этики исследований**. Университеты должны обеспечить выполнение этих новых правил к концу этого месяца, в ином случае они рискуют, что их не будут считать имеющими право на получение грантов из главного исследовательского фонда страны, Deutsche Forschungsgemeinschaft (DFG). Эти правила следуют международным нормам при определении научного должностного преступления или нарушения научной этики как умышленной или грубо пренебрежительной фальсификации или фабрикации данных. Другими перечисленными серьезными должностными преступлениями являются введение в заблуждение, плагиат и ущерб исследованиям других. Возможные санкции включают лишение контрактов на исследования и **аннулирование академических званий**. (...) Правила о научных должностных преступлениях, о неэтичном поведении явились плодом многих разбирательств после того, как основанная DFG-группа нашла фальсификацию в дюжине статей пары исследователей рака (Фридгельм Герман и Марион Брак, [37]) (...) **Выявление умышленной или грубой небрежности при обработке или получении данных является трудным делом, как это печально известно.** (...) По новым правилам институты должны назначать независимого секретаря-омбудсмана, который будет инициировать проверки и в то же время защищать тех, кто поднял тревогу. Вдобавок, чтобы ускорить будущие расследования, новые правила устанавливают, что во всех случаях, когда это только возможно, **первичные данные исследований должны сохраняться в течение десяти лет.** Это, «вероятно, одна из тех областей, где исследователи наиболее беспечны», говорит Иоханнес Дикганс (Johannes Dichgans), нейролог и омбудсмен университета в Тюбингене. **Пренебрежение архивированием записей данных исследований или их умышленное разрушение должно было бы осуждаться как грубая небрежность и потому быть наказуемым**» [36].

Чем же можно объяснить такую терпимость к многолетнему воспроизводству статистических ошибок в российских публикациях биомедицинской тематики? Основной барьер на пути уменьшения многочисленных ошибок в использовании статистических методов — не критическое отношение специалистов как к собственной исследовательской практике, так и к практике своих коллег, а также противодействие руководителей медицинских вузов и НИИ попыткам улучшения ситуации. Результатом этого является неявное следование следующим установкам:

□ — использовать тот статистический метод, который предложил руководитель (ослушаться нельзя, а если кто-то обвинит в некорректности, то виноват шеф);

— применять данный метод потому, что он есть в доступном статистическом пакете (надо же написать в статье, что использовал статистический пакет);

— использовать те методы, которые ранее уже были самим автором использованы (за них никто не ругал, значит можно и впредь их применять);

— использовать данный метод потому, что его уже использовал другой автор статьи (диссертации) по этой же тематике: раз статью опубликовали (диссертацию утвердили), значит можно использовать, ведь в редакциях (диссертационных советах) не дураки сидят.

Нередко следование этим установкам приводит к генерации цепочек так называемых «мемов», когда несколько авторов последовательно заимствуют друг у друга одну и ту же ошибку [38].

Можно выделить следующие причины появления и живучести ошибок:

1. Отсутствие знаний по статистике у авторов публикаций, членов журнальных редколлегий и диссертационных советов, что не позволяет им осознать всю степень ошибочности выводов в работе.

2. Деформация цели публикации. Очень часто истинная цель публикации — не стремление обнаружить надежные, достоверные результаты исследования, а лишь сам факт быстрой публикации пусть даже ненадежных выводов как доказательство научной состоятельности и учености автора.

3. Отдаленность во времени и пространстве возможных негативных последствий опубликования сомнительных результатов от автора, времени и места публикации. Автор такой публикации, зная о том, что его выводы не бесспорны, понимает, что именно эта неочевидность и не будет способствовать стремительному их внедрению в реальную медицинскую практику. А если такое и случится, то доказать, что основной причиной негативных последствий были именно недоброкачественные выводы автора этой публикации, тоже весьма проблематично.

4. Ограниченность доступа широкой читательской аудитории к медицинским публикациям, содержащим ошибочные выводы. Это связано как с ограниченными тиражами медицинской периодики, так и с ограничением доступа к медицинским диссертациям. В ряде медицинских вузов для доступа к уже защищенным диссертациям необходимо разрешение ректора.

5. Отсутствие доступа для читателей к исходным данным автора публикации. Автор, зная, что любой читатель, сомневающийся в истинности авторских выводов, имеет возможность немедленно (или же спустя некоторое время) получить эти данные и самостоятельно перепроверить авторские выводы, будет совершенно иначе относиться к процедуре статистического анализа. В настоящее время среди английских медиков уже обсуждается необходимость публикации «сырых данных», используемых авторами публикаций [31]. В частности, предлагается размещение таких данных в электронном виде [39], что позволит в реальном масштабе времени дополнять их и каждому участнику проводить повторный анализ таких дополненных данных. «Сырые данные» некоторых клинических испытаний уже доступны для анализа в электронном виде, и это позволило получить качественно новые выводы.

6. Незаинтересованность редакций периодических изданий, ректоров медицинских вузов, диссрта-

ционных советов, ВАК РФ и дирекций медицинских НИИ в повышении статистического качества журнальных публикаций и диссертаций. Предельно четко сформулировал эту позицию в беседе с автором этих строк ректор Сибирского государственного медицинского университета, академик РАМН, профессор Новицкий В.В. Когда я обсуждал с ним предложение об организации независимой статистической экспертизы диссертаций, представляемых в диссертационные советы его вуза, он сказал буквально следующее: «Как ученый я за, а как ректор — против».

Каждая из описанных причин может быть объектом воздействия, однако ввиду сложных взаимодействий перечисленных причин и их преимущественно социальной природы в основном остается уповать на добросовестность отдельного ученого, т.е. на то, что в истории науки составляло ее эмос, и что остается основой ее эффективности.



ОШИБКИ ОПИСАНИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ

Вырази ложную мысль ясно,
И она сама себя опровергнет.
Л. Вовенарг

Кроме ошибок, являющихся результатом неправомерного использования тех или иных статистических методов, часто встречаются некорректные описания того, что и как было сделано исследователем с использованием этих методов. Нередко читателю по такому описанию невозможно установить, какие конкретно статистические методы были использованы, насколько корректно выполнен статистический анализ, заслуживают ли доверия его результаты и авторские выводы, построенные на их основе. В большинстве российских медицинских журналов требования к описанию этой части исследования вообще отсутствуют. В других эти требования являются калькой не вполне удачных рекомендаций, заимствованных из иных журналов [40]. Между тем, имеется немало методических материалов, регламентирующих корректное описание данного этапа исследований: книги [41] и разнообразные рекомендации на сайтах медицинских журналов. Наиболее полно и подробно такое описание дано на сайте BMJ (<http://resources.bmj.com/bmj/authors>).

Анализ более 1500 опубликованных статей и диссертаций биомедицинской тематики, проведенный нами в течение последних 15 лет, показал, что примерно в 25% работ, в которых авторы применяли статистические методы, вообще не описываются подробно использованные методы и критерии. Большинство таких работ изобилуют выражениями типа « $p < 0,05$ », « $p < 0,01$ » или « $p > 0,05$ ». Однако нигде в тексте не сообщается, к каким конкретно статистическим критериям относятся данные вероятности.

В 63 статьях журнала «Бюллетень экспериментальной биологии и медицины» (БЭБМ) за 1997 г., в которых использовалось выражение « $p < 0,05$ », вообще не упоминается использованный авторами статистический критерий. При чтении таких работ возникает ощущение того, что выражения « $p < 0,05$ » или

« $p < 0,01$ » «самозарождаются». Между тем, в контексте описываемых задач выражения « $p < 0,05$ » и « $p < 0,01$ » могут относиться только к конкретным статистическим критериям, например критериям Пирсона, Фишера, Вилкоксона, Стьюдента и т.д. Отсутствие упоминания критериев в тексте позволяет предположить, что авторы либо не считают уточнение использованных статистических критериев достаточно важным для того, чтобы приводить их в диссертации, монографии или статье, либо же специально не приводят ее исходя из иных мотивов. Подобный стиль описания наиболее характерен для журналов БЭБМ, «Кардиология» и «Вестник РАМН».

Почему важно подробно описать в тексте статьи или диссертации всю схему статистического анализа экспериментальных данных, в том числе указать использованные при этом статистические критерии? Как для каждого лекарственного средства имеются свои специфические противопоказания к применению, так и для конкретных статистических критериев имеются ограничения. Поэтому детальное описание схемы анализа экспериментальных данных — важнейший аргумент в пользу адекватности выбранного критерия анализируемым данным и проверяемой статистической гипотезе, а стало быть, и обоснование надежности выводов. В противном случае выводы сомнительны и не могут считаться научно обоснованными.

Оставшиеся 75% работ содержат в той или иной форме описание использованных методов статистики, обычно в разделе «Материал и методы». В большинстве работ описание статистических методов и критериев дается в форме клише с весьма расплывчатым содержанием.

Можно выделить три группы подобных клише. В первой группе с разными вариациями просто констатируется сам факт использования статистических методов. Доля таких работ по отношению ко всем работам, содержащим описания, составляет примерно 45%. Для этой группы характерны следующие описания:

1. Результаты **обработаны статистически**.
2. Обработку данных проводили **статистическими методами**.
3. Результаты обработаны **методом вариационной статистики** (варианты: стандартными программами, общепринятыми методами и т.п.).
4. Результаты **подвергнуты статистической обработке**.

Пожалуй, наиболее необычное описание этого важного этапа исследования дано в статье «Варьирование активностей цистеиновых катепсинов почек и печени» (БЭБМ 1995;12:586—9): «Математическую обработку проводили **обычным способом** с использованием компьютера». Далее в статье авторы нигде не упоминают ни одного конкретного метода статистического анализа и статистического критерия, хотя и используют выражения « $p < 0,05$ » и « $p < 0,01$ ». Попробуйте представить себе аналогичное описание лечения больного: «Лечение производилось **обычным способом** с использованием лекарственного средства»!

Для второй группы работ (примерно 30%) характерно указание на более конкретные детали (назва-

ние критерия, компьютера и т.п.), не позволяющие определить, каким образом в действительности проводился анализ:

1. Материал обрабатывали статистически по методу Лакина (варианты: по В.Ю. Урбаху; по таблицам Стрелкова).

2. Достоверность значений определяли по t-критерию Стьюдента (варианты: по среднеквадратичному отклонению усредненных значений параметров; методом определения среднего квадратического отклонения от средней величины).

3. Статистическая обработка материала произведена с использованием компьютера IBM PC, (Pentium II...) по стандартным программам с помощью статистического пакета.

Очевидно, что для читателя ссылка на фамилию (Лакин, Урбах, Стрелков) ни о чем не говорит, поскольку в книгах данных авторов описано много разнообразных методов. Что сможет узнать врач о методике лечения больных, если автор статьи или диссертации напишет: «Лечение больным назначалось по Машковскому»? Делать подобную ссылку все равно, что утверждать наличие методики Машковского, только по той причине, что есть лекарственные справочники с такими авторами.

Третья группа проанализированных работ (примерно 25%) содержит достаточно безграмотные, а подчас и просто абсурдные либо противоречивые сочетания перечисляемых терминов или критериев. В таких описаниях авторы часто вводят свою собственную терминологию, не раскрывая в тексте работы смысла используемых терминов. Вот типичный пример такого описания: «Достоверность различия между средними арифметическими сравнениями вариационных рядов устанавливалась по степени вероятности положительной гипотезы в соответствии с описанием, приведенным в монографии Г.Ф. Лакина»[42]. Поясним это бессмысленное утверждение: правильным образом сравнивают две средние величины и это то же самое, что оценивать различие между ними; «степень вероятности положительной гипотезы» — бессмысленная фраза!

МЕМЕТИЧЕСКИЙ АНАЛИЗ ОПИСАНИЙ МЕТОДОВ СТАТИСТИКИ

Дьявол скрывается в деталях.

Концепции меметики [38, 44] позволяют описывать в терминах генетики размножение, распространение, отбор, мутации и рекомбинации, а также смерть мемов — элементарных единиц, квантов культуры. Такими информационными квантами — мемами — могут выступать литературные клише и обороты, используемые авторами печатных работ, сформулированные идеи, лозунги, религиозные догмы, мода на тот или иной стереотип поведения, музыкальные мотивы и т.д.

Траекторию жизни мема можно представить по аналогии с распространением вируса, который может существовать только в клетке инфицированного переносчика. Так и мем может существовать только в определенных носителях. Ими могут быть речь,

пресса, книги, звуковые записи, видеозаписи, компьютерные файлы и т.д. Переносчиком же мема может быть лишь человек. При этом в процессе переноса мем легко может подвергаться перестройке, рекомбинации с другими мемами, существующими на данном носителе. Естественный отбор сохраняет в данном индивиде в основном те мемы, которые легко им запоминаются, понимаются другими индивидами, приносят его носителю определенные дивиденды и по этим причинам являются наиболее инфекционными. При этом меметика не занимается вопросами истинности или ложности переносимой в мемах информации.

Психологам хорошо известно, что большинству специалистов свойственно преувеличивать глубину познаний другого профессионала в некой сфере деятельности, если она отлична от его собственной сферы профессиональных интересов. Причем чем более удалены друг от друга эти сферы деятельности, тем сильнее это заблуждение. Чем выше уровень квалификации специалиста и чем критичнее его отношение в своей области знания, тем доверчивее и не критичнее его отношение к данным иных сфер знания. Именно по этой причине исследователь более всего подвержен «инфицированию» мемами из этих пограничных либо удаленных сфер деятельности, и, соответственно, познание о этих сферах представлено в его сознании, поведении, речи и научных произведениях в форме подобных мемов. В частности, исследователи в области медицины предрасположены к «инфицированию» мемами из терминологии точных наук, и в первую очередь — статистики.

Попытаемся взглянуть на ошибки описания статистических методов с позиций меметики, выделяя при необходимости те или иные обороты, выражения, а нередко и отдельные предложения или группы предложений как ментальные репликаторы — мемы.

Приведенные выше клише являются типичными мемами. Часть таких мемов вполне адекватно описывает несложный статистический инструментарий, применяемый авторами. Однако подавляющее их большинство несет на себе отпечаток смысловой противоречивости и вербальной эклектичности. В такой смысловой инженерии медик уподобляется ребенку, которому в руки попал конструктор «Сделай сам» с набором отдельных элементов, но без руководства по сборке нужного изделия. По этой причине часто наблюдается смысловая удаленность содержания выполненного анализа от его описания. Нередко можно наблюдать «меметические цепочки», состоящие из дословно повторяющихся предложений, следующих в ряду статей или диссертаций. Ниже даны четыре авторских описания, относящиеся к определению достоверности различия (табл. 1). Три из них приведены в диссертациях [46—48], четвертое — в монографии [45].

Различающиеся части мема выделены *жирным курсивом*.

Как видим, эти четыре описания весьма похожи друг на друга. Менее всего отличаются между собой описания из диссертаций [47, 48] и монографии [45].

Обратим внимание на наличие во всех 4 случаях субмема «показатель точности», причем в 3 случаях

Таблица 1. Мемы описания сравнения средних из 4 диссертаций

Диссертация [46] (стр.41)	Диссертация [47] (стр.53)	Монография [45] (стр.48)	Диссертация [48] (стр.40)
Достоверность различия сравниваемых величин определяли с помощью показателя точности Р по таблице Стьюдента, где он располагается в зависимости от значений t и n.	Достоверность определяли с помощью абсолютного показателя точности «р», который находили по специальной таблице, где он расположен в зависимости от (t) Стьюдента и числа степеней свободы «n».	Достоверность <i>различий</i> средних арифметических определяли по абсолютному показателю точности процентных точек распределения Стьюдента в зависимости от коэффициента достоверности (t) и числа степеней свободы (n).	Достоверность различных средних арифметических определяли по абсолютному показателю точности (<i>Р</i>) по таблице процентных точек распределения Стьюдента в зависимости от коэффициента достоверности (t) и числа степеней свободы (n).

Таблица 2. Мемы описания различия 2 средних

Описание из монографии [45]	Описание из диссертации [48]
На основании критерия t по таблице Стьюдента определялась вероятность различия. Различие считалось достоверным при $p < 0,05$, т.е. в тех случаях, когда вероятность различия составляла больше 95% (стр. 48)	На основании t по таблице Стьюдента определялась вероятность различия (p). Различие считалось достоверным при $p < 0,05$, т.е. в тех случаях, когда вероятность различия составляла больше 95% (стр. 40)

он называется «абсолютным показателем точности». Попытаемся выяснить, какой смысл могли вкладывать авторы этих работ в «показатель точности», используя для этой цели следующий мем из монографии [45] и диссертации [48] (табл. 2).

Авторы монографии не приводят в этом меме объяснения величины *p*. Можно предположить, что они подразумевают под ней уровень значимости, и тогда в последнем предложении использованного мема нет противоречия. Действительно, «при $p < 0,05$ » мы имеем достоверную вероятность более 95%. Таким образом, вероятнее всего, авторы монографии [45] подразумевали уровень значимости. Совершенно бессмысленную конструкцию имеет этот мем в диссертации [48], где величина *p* определяется как «вероятность различия». В одном и том же предложении диссертант утверждает, что «различие считалось достоверным при $p < 0,05$, т.е. в тех случаях, когда вероятность различия составляла больше 95%». Таким образом, мы имеем случай рекомбинации двух мемов, в результате которой новый мутантный мем приобрел искаженную бессмысленную форму.

Анализ мемов со сложной конструкцией не всегда позволяет достаточно надежно представить понятия, которые авторы пытаются описать с их помощью. По своему языку такие описания имеют стиль известных произведений А. Платонова «Котлован» и «Чевенгур», а также речи экс-премьера Российской Федерации В. Черномырдина. Приведем одно из подобных описаний, содержащихся на с. 30 диссертации [49] и попытаемся его проанализировать: «Для определения значимости выборочных показателей, *оценки сущности двух или нескольких показателей*, а также определения связи между явлениями, полученными (так в тексте, выделено нами — В.Л.) в результате выборочных исследований, были использованы сле-

дующие формулы: ...». Итак, каков же смысл первого утверждения — «определение значимости выборочных показателей»?

Предположим, что в результате 50 наблюдений получен выборочный показатель — среднее арифметическое значение систолического давления 90 мм рт. ст. Что тогда понимать, согласно автору, под «значимостью выборочного показателя»? Как предлагает автор «оценить сущность двух или нескольких показателей», например того же среднего систолического давления, содержания иммуноглобулинов, липопротеидов, холестерина и т.д., и какой смысл вкладывает диссертант в сам термин «сущность»? Не ясно и то, как собирается автор «определять связи между явлениями», поскольку ни в приводимых ниже формулах, ни в тексте диссертации об этом ничего не сказано. Продолжим авторский текст: «Пределы возможных колебаний (доверительные интервалы) средних величин, полученные в выборочном исследовании, принимались равными удвоенной средней ошибке ($+2m$), что дает основание с вероятностью 95% отнести полученную закономерность ко всей генеральной совокупности».

В первой половине этого предложения автор пытается объяснить технологию построения им 95% доверительных интервалов (ДИ). Однако из его объяснения не ясно, какой смысл имеет величина ($+2m$). Является ли это величина всей шириной 95% ДИ, либо же это полуширина (только средняя плюс $2m$), и тогда полная ширина ДИ равна $4m$? Итак, какой же ДИ строит диссертант? Не меньше вопросов возникает и при изучении второй половины данного предложения, в которой речь идет о «полученной закономерности», поскольку ранее нигде автор ничего не говорит ни о видах закономерностей, ни о методах их оценки.



ЛОКАЛИЗАЦИЯ ОШИБОЧНЫХ ОПИСАНИЙ ВНУТРИ НАУЧНЫХ ШКОЛ

Дилетант обнаруживает себя прежде всего незнанием терминологии.

Г. Селье

Описания, приведенные в следующих 2 кандидатских и 3 докторских диссертациях, представляют особый интерес. 4 из них относятся к одной и той же научной специальности (14.00.17 — нормальная физиология) и выполнены в одном и том же вузе — Сибирском государственном медицинском университете, а 5-я диссертация защищена по близкой специальности (14.00.16 — патологическая физиология). Интересно и то, что один и тот же человек выступает в роли научного руководителя в кандидатских и научного консультанта в докторских диссертациях. Автор же докторской диссертации [69] выступает в качестве научного руководителя в кандидатской диссертации [68]. Все эти факты позволяют говорить о принадлежности диссертантов к одной научной школе. Эти работы достаточно близки по:

1. срокам выполнения работ: 1992, 1993, 1994 и 1995 г.;

2. содержанию исследования и месту выполнения;

3. структуре названия работ — в трех диссертациях одно и то же первое слово, а в двух диссертациях полностью совпадают первые четыре слова (табл. 3).

Все это позволяет сделать вывод о том, что в научной школе СГМУ культивируется и прочно закрепилось ошибочное понимание и описание самого смысла доверительной вероятности. Приведем несколько цитат, касающихся смысла этого понятия, и далее сравним их с анализируемым ошибочным методом. Жирным шрифтом выделим ту часть цитат, которая будет наиболее важна для нашего анализа.

1. «Доверительная вероятность — вероятность того, что оцениваемый вектор характеристики (параметров) совокупности генеральной покрывается доверительной областью (доверительным интервалом — оценкой интервальной при одном параметре). **Доверительная вероятность должна быть достаточно большой, т.е. отвечать принципу практической достоверности.** Другие названия: надежность доверительной области, доверительный коэффициент» [52, с. 117].

2. «А теперь рассмотрим интервальные оценки, т.е. возможность, исходя из выборочных значений $x_1, x_2, \dots, x_n$, построения некоторого интервала, содержащего истинное значение параметра Θ с заданной вероятностью. Метод получения таких интервалов сводится к построению нижнего и верхнего доверительных пре-

делов Θ_n и Θ_v (Θ_n и Θ_v есть функции от выборочных значений), и если мы говорим, что истинное значение параметра лежит между этими доверительными пределами, то это утверждение выполняется с вероятностью $(1 - \epsilon)$ и не выполняется с вероятностью ϵ [число $(1 - \epsilon)$ называется доверительным уровнем]. Следовательно, **выбрав достаточно малое число (обычно в биологических задачах 0,05 или 0,01), мы по выборочным значениям $x_1, x_2, \dots, x_n$, построим доверительный интервал (Θ_n, Θ_v) , соответствующий доверительному уровню $1 - \epsilon$.** Обычно строят так называемые центральные доверительные интервалы. Очень часто, когда доверительный уровень измеряется в процентах, применяется термин $100(1 - \epsilon)\%$ -ный интервал (например, 95%-ный доверительный интервал confidence interval, если $1 - \epsilon = 0,95$)» [53, с. 44].

3. «В этом параграфе мы рассмотрим вопрос получения интервальных оценок, т.е. возможность построения некоторого интервала, содержащего истинное значение параметра с заданной точностью. Сам метод получения таких интервалов сводится к построению так называемых доверительных пределов, а полученные интервалы называются доверительными. Процедура, которую мы собираемся описать в этом параграфе, состоит в построении нижнего и верхнего доверительных пределов уровня $(1 - \alpha)$, обладающих следующим свойством: если мы говорим, что истинное значение параметра лежит между этими пределами, то это утверждение верно с вероятностью $1 - \alpha$ (и неверно с вероятностью α). **Очевидно, что « α » следует выбирать достаточно малым (обычно используемые значения равны 0,05 и 0,01).** Число $1 - \alpha$ называют коэффициентом доверия или доверительной вероятностью» [54, с. 114].

4. Закончим этот экскурс в описание понятия доверительной вероятности часто цитируемым в биомедицинских публикациях учебным пособием Г.Ф. Лакина [43, с. 106]. «**Вероятности, признанные достаточными для уверенного суждения о генеральных параметрах на основании известных выборочных показателей, называют доверительными вероятностями.** Понятие о доверительных вероятностях предложено Р.Фишером. Оно вытекает из принципа, который положен в основу применения теории вероятностей к решению практических задач. Согласно этому принципу, маловероятные события считают практически невозможными, а события, вероятность которых близка к единице, принимают за почти достоверные. Обычно в качестве доверительных используют вероятности $P_1 = 0,95$; $P_2 = 0,99$; $P_3 = 0,999$ ».

Итак, основной вывод, который делается во всех процитированных определениях, — величина довери-

Таблица 3. Мемы описания доверительной вероятности

Описание из диссертации [68] (стр. 35)	Описание из диссертации [69] (стр. 49)	Описание из диссертации [70] (стр. 39)	Описание из диссертации [71] (стр. 59)	Описание из диссертации [61] (стр. 68)
Достоверными считались отличия с уровнем доверительной вероятности $p < 0,05$	Достоверными считались отличия с уровнем доверительной вероятности $p < 0,05$	Достоверными считались отличия с уровнем доверительной вероятности $p < 0,05$	Достоверными считались отличия с уровнем доверительной вероятности $p < 0,05$	Достоверными считали различия с уровнем доверительной вероятности $p < 0,05$

тельной вероятности, т.е. степени нашей уверенности в декларируемых утверждениях, должна быть достаточно близка к 1. Близость этой величины к 0 говорит о недоверии к декларируемому утверждению. Таким образом, 5 докторов и кандидатов наук, использовавших приведенный выше ошибочный мем, утверждали, что степень их уверенности в своих выводах не более 5%! Наличие подобного растиражированного ошибочного мема в нескольких защищенных диссертациях означает, что это абсурдное утверждение не смутило ни руководителей и консультантов этих диссертаций, ни членов диссертационных советов, ни представителей ведущих организаций, ни оппонентов и рецензентов, и даже членов экспертных советов ВАК РФ. Таким образом, научная среда, где циркулирует данный ошибочный мем, гораздо шире, нежели коллектив данных авторов. Напомним, что эти утверждения приведены не в дипломных работах выпускников вуза, а в квалификационных работах на ученые степени кандидатов и докторов медицинских и биологических наук, авторы которых уже имели не одну публикацию, а некоторые из них годами читают лекции студентам. Поэтому сомнительно, что приведенные мемы — результат неточного изложения авторами своих мыслей. Скорее наоборот, данные мемы и есть адекватное отображение устоявшегося в данной научной школе и в СГМУ абсолютно искаженного представления о доверительной вероятности. Поскольку содержание данного ошибочного описания одобрено многими официальными уровнями (ведущая организация, диссертационный совет, экспертный совет ВАК, президиум ВАК РФ), то понятно, что и последующие исследователи, читая эти диссертации и слушая лекции этих преподавателей, также будут инфицированы этими мемами, которые вновь будут реплицироваться уже в их статьях и диссертациях.



МЕМЫ КАМУФЛЯЖНЫХ ОПИСАНИЙ

Недостоверна видимость натуры
в сравнение с данными литературы.
И.В. Гете (Фауст)

Внимательное изучение ошибочных диссертационных описаний статистического анализа обнаруживает ряд признаков, которые позволяют в ряде случаев говорить о камуфляжном характере описаний. Подобные описания камуфлируют непонимание авторами смысла использованных ими статистических методов. По своему содержанию они представляют не более чем попытками «онаучить» авторские выводы перечислением разнообразных статистических методов и терминов. Видимо, по мнению авторов, такое перечисление должно способствовать созданию у читателей положительного образа исследования, в котором применялись современные методы статистического анализа, и потому вызвать к авторским выводам большее доверие.

Рассмотрим следующий пример: «Статистический анализ проводился на основе банка данных, доступ к которому реализовывался через пакет программ STATGRAPHICS в которой (так в тексте. В.Л.) использо-

вались вариационный, корреляционный, регрессивный и дискриминантный методы, помимо средних значений рассчитывались величины дисперсий первого и второго порядка, определялась погрешность и достоверность, оценка параметров проводилась по Т-критерию Стьюдента» [55, с. 41]. С другими работами этой научной школы данную диссертацию роднит не только близость научного направления и объекта исследования, но и наличие мема «на основе банка данных, доступ к которому реализовывался через пакет программ». Имеется здесь и часть других мемов, в частности упоминание «вариационного» и «регрессивного» анализа. Если под «регрессивным» анализом автор, видимо, подразумевал регрессионный анализ, что такое «вариационный анализ», можно только гадать. Упоминание об этом таинственном методе можно встретить и в других публикациях [56, с. 42].

Следующее ядро мема — «дисперсии первого и второго порядка». Наиболее вероятный механизм появления этого мема — комбинация определения дисперсии и определений начальных и центральных моментов. Судя по приведенному выше описанию, автором «определялась погрешность и достоверность». Видимо, полагая, что эти определяемые величины не требуют дополнительного комментария, автор не поясняет ни смысл этих величин, ни методику их определения. Не меньше вопросов вызывает и то, как «оценка параметров проводилась по Т-критерию Стьюдента». В частности, из этого отрывка совершенно непонятно, оценка каких именно параметров проводилась автором с помощью t-критерия Стьюдента. Ответов на эти вопросы в тексте диссертации нет, как нет и описания того, с помощью каких статистических критериев проводились проверки многочисленных гипотез, после чего появлялись выражения « $p < 0,05$ ». Нет там, конечно, и никаких результатов «регрессивного» и дискриминантного анализа. Так что ответ на вопрос: «А использовались ли вообще все эти виды статистического анализа, и не является ли упоминание о них не более чем приукрашиванием диссертации?», — очевиден.

Можно выделить два вида камуфляжных описаний. В первом случае автор декларирует некоторые статистические методы или критерии, якобы использованные им в исследовании. Однако далее в тексте не приводит никакой информации, ассоциирующей с результатами применения этих методов. В качестве типичного случая такого камуфляжного мема можно привести следующую цитату: «Фактические данные обработаны методами математической статистики для малых выборок на ЭВМ: метод сравнения средних, метод сращения дисперсий, проверка гипотез (так в тексте. В.Л.) по критерию «хи-квадрат» [57, с. 38].

Однако во всем дальнейшем тексте диссертации нет даже малейших намеков на результаты применения этих методов математической статистики! Нет ни привычных выражений « $M \pm m$ » и « $p < 0,05$ », отсутствуют сами проверяемые «гипотезы» и обсуждение результатов их проверки, как в «Заключении», так и в «Выводах». Так что и здесь ответ на аналогичный вопрос: «А использовались ли вообще все эти виды статистического анализа, и не является ли упоминание

о них не более чем приукрашиванием диссертации?», — также достаточно очевиден.

Такое же камуфляжное описание можно обнаружить в [58]: «Использованы вариационный, корреляционный, регрессионный и дискриминантный методы ... Для определения информативности маркеров использован факторный анализ. Клинико-иммунологическая классификация больных в трехмерном пространстве признаков проведена методом кластерного анализа» (с. 31). Впечатляет, не правда ли? Однако, прочитав всю диссертацию, мы не обнаружим даже малейшего намека на результаты использования всех этих видов статистического анализа! Не удивлюсь, если в отзывах оппонентов и ведущей организации по этой диссертации в качестве положительного аргумента отмечено использование диссертантом разнообразных статистических методов, включая и многомерные.

Еще один пример камуфляжного описания. В редакцию «Сибирского медицинского журнала» поступила на рецензию рукопись «Опыт лечения вазилипом больных с острым коронарным синдромом без подъема сегмента ST», в которой сообщалось, что «...для анализа выживаемости использовали логранговый критерий». Из текста же рукописи следовало, что за период лечения вообще не было смертельных исходов. Возникает вопрос, с какой же целью и как можно проводить «анализ выживаемости», если все пациенты остались живы? Когда этот вопрос был задан мной автору рукописи — сотруднице НИИ кардиологии ТНЦ СО РАМН, она не смогла ответить на данный вопрос, сославшись на то, что статистический анализ выполнял другой человек, который не является автором. Резонно в таком случае задаться вопросом, насколько этично считать себя автором, не будучи в состоянии отвечать за все содержание рукописи?

СМУТНО ПИШУТ О ТОМ, ЧТО СМУТНО ПРЕДСТАВЛЯЮТ

Суди дружок, не слыше сапога!

А.С. Пушкин

Смутное представление о том, что означают те или иные величины, используемые в статистическом анализе, приводит к тому, что авторы публикаций путают их между собой. Вот какое описание приведено в [59]: «Статистическую обработку полученных данных проводили методом Фишера—Стьюдента. Различия считали достоверными при $t < 0,05$ ». Как видим, авторы этой работы перепутали уровень значимости p и критерий Стьюдента t .

Очень часто авторы путают между собой и такие величины и понятия, как уровень значимости и доверительная вероятность. Вот что, к примеру, можно прочесть в статье [60]: «Для всех статистических тестов в качестве критерия статистической достоверности рассматривался **уровень значимости более 0,95**».

Обратимся к толкованию этого термина в статистическом словаре: «**УРОВЕНЬ ЗНАЧИМОСТИ** — одна из характеристик качества критерия статистической проверки гипотез. Пусть выдвинута гипотеза H_0 (основная, или «нулевая»). Всякое стат. ре-

шение, принимаемое на основе ограниченного ряда наблюдений, неизбежно сопровождается *вероятностью* ошибочного заключения. С вероятностью *альфа* гипотеза H_0 может оказаться отвергнутой, в то время как на самом деле она является справедливой (*ошибка первого рода*), или, наоборот, с вероятностью *бета* может быть принята гипотеза H_0 , в то время как на самом деле она является ошибочной (*ошибка второго рода*). ... В частности, при фиксированном объеме выборки обычно задаются величиной *альфа* вероятности ошибочного отвержения проверяемой гипотезы H_0 . Эту вероятность ошибочного отклонения «нулевой» гипотезы принято называть **УРОВНЕМ ЗНАЧИМОСТИ** ... На практике часто пользуются след. стандартными значениями *альфа*: 0,1, 0,05, 0,025, 0,01, 0,005, 0,001. Особенно распространенной является величина **УРОВНЯ ЗНАЧИМОСТИ** *альфа*, равная 0,05. Она означает, что в среднем в пяти случаях из ста ошибочно отвергают высказанную гипотезу при пользовании данным *критерием статистическим*» [52, с. 542].

Дополним это толкование термина «уровень значимости», обратившись к определению понятия «доверительная вероятность», и, применив его, объясним смысл авторского утверждения. Для этого используем уже упомянутую популярную среди биологов и медиков книгу Г.Ф. Лакина [43], на стр. 107 которой читаем: «...С доверительной вероятностью тесно связан *уровень значимости альфа*, под которым понимают разность $\alpha = 1 - P$ ». Согласно авторскому утверждению [60], получается, что при уровне значимости более 0,95 они использовали доверительную вероятность менее 0,05! **Фактически авторы статьи из журнала «Генетика» утверждают, что они верят собственным результатам менее чем на 5%!** Таким образом, используя уровень значимости более 0,95, авторы ошибочно отвергали высказанные гипотезы более чем в 95 случаях из 100. О какой же «достоверности» можно говорить в этом случае? Самое печальное в таких ошибках то, что читателю невозможно установить саму природу ошибки. Произошла ли только ошибка описания, и тогда мы имеем дело всего лишь с терминологической путаницей, но сами выводы верны, или же данная ошибка как авторское заблуждение присутствовала уже на этапе выполнения анализа, и тогда уже ошибочны сами выводы? Именно поэтому в широко используемых в последние годы критериях качества исследований признается невозможность разделить качество отчета и качество исследования. Единственный выход — при наличии отчета с видимыми дефектами предполагать, что и само исследование проведено методически ущербно.

Аналогичную путаницу обнаружим и в других работах:

— «Достоверными считали различия с уровнем доверительной вероятности меньше 0,05» [61];

— «Затем находилась достоверность различий по критерию Стьюдента при уровне доверительного интервала $P = 0,05$ » [62];

— «Достоверность различий оценивали с помощью t -критерия Стьюдента при 95% уровне значимости» [63];

— «Статистическую обработку проводили с использованием *t*-критерия Стьюдента, при уровне доверительной вероятности $p < 0,05$ » [64].



КОВАРНЫЙ *t*-КРИТЕРИЙ СТЬЮДЕНТА

Нужно делать так, как нужно,
а как не нужно — делать не нужно.

Винни-Пух

Наибольшей популярностью при проверке гипотез о равенстве генеральных средних пользуется *t*-критерий Стьюдента. При чтении статей БЭБМ и «Вестника РАМН» складывается впечатление, что большинство авторов этих журналов знают и используют лишь *t*-критерий Стьюдента. Например, в выпусках БЭБМ за 1997 г. *t*-критерий использован в 125 статьях, тогда как корреляционный и дисперсионный анализ применен всего лишь в 15, критерий Колмогорова—Смирнова — в 1, парная линейная регрессия — в 3, точный критерий Фишера — в 3 статьях.

Как известно, использование *t*-критерия Стьюдента имеет два ограничения, а именно, нормальность распределения в обеих сравниваемых группах и равенство генеральных дисперсий. Из 1562 проанализированных нами статей, монографий, диссертаций и авторефератов, авторы которых использовали *t*-критерий Стьюдента, упоминание о проверке нормальности распределения исследуемых признаков было только в 23 работах! О проверке второго ограничения — равенства генеральных дисперсий — упоминалось лишь в 1 работе.

Очевидно, что без наличия исходных данных читатель не может проверить факт нормальности распределения признаков, анализируемых в таких публикациях. Наши исследования нормальности распределения биомедицинских признаков, проведенные в течение 20 лет более чем на 10 тыс. количественных переменных, показали, что примерно 75% используемых переменных не подчиняются нормальному распределению. Проверка второго условия возможности применения *t*-критерия Стьюдента чаще всего для читателя доступна. Предположим, что анализируемые авторами признаки действительно подчиняются нормальному распределению. В этом случае для проверки второго требования — равенства генеральных дисперсий — необходимы лишь по два параметра для каждой из сравниваемых групп: выборочные дисперсии и объемы наблюдений. Используя их, можно вычислить *F*-критерий Фишера и далее оценить достигнутый уровень значимости p . Чаще всего авторы подобных публикаций приводят в таблицах объем наблюдений n и либо стандартное (среднеквадратичное) отклонение s , либо ошибку среднего m для каждой из сравниваемых групп. В первом случае проверку гипотезы о равенстве генеральных дисперсий двух совокупностей можно произвести, вычислив *F*-критерий Фишера по формуле $F = s_1^2/s_2^2$, где s_1^2 — максимальная по величине дисперсия, а s_2^2 — минимальная дисперсия. Во втором же случае, используя значения m и n , можно получить значение выборочной дисперсии s^2 по формуле $s^2 = m^2 \cdot n$ для каждой из групп и далее вычислить *F*-критерий Фишера. Наши

исследования, а также результаты исследований других авторов показывают, что в большинстве случаев для количественных признаков, изучаемых в биомедицине, характерно значительно увеличение дисперсии в опытной группе (группе вмешательства). Причем такое увеличение чаще всего сопровождается и увеличением среднего значения в опытной группе. Иногда же увеличение дисперсии происходит и на фоне практически неизменного среднего. Отметим, что степень искажения истинности выводов при использовании критерия Стьюдента с нарушением необходимых условий определяется в основном тремя параметрами: видом распределения в сравниваемых группах, равенством (неравенством) генеральных дисперсий и соотношением объемов выборок [74].

Детальный пример такой послепубликационной проверки допустимости применения *t*-критерия Стьюдента рассмотрен в электронной публикации (http://www.biometrika.tomsk.ru/kk/index_6.htm#96). В работе [65] авторы изучали 4 группы пациентов: здоровые доноры (контрольная группа); больные с непсихотическими расстройствами; больные с умственной отсталостью; больные с параноидной шизофренией. Оказалось, что в данной работе гипотеза о равенстве дисперсий в большинстве случаев уверенно отвергается. Величины дисперсий различаются между собой иногда в десятки раз! Если даже предположить, что во всех сравниваемых группах наблюдается нормальное распределение, что само по себе весьма маловероятно, тем не менее, критерий Стьюдента не может быть использован в данных условиях вследствие неравенства генеральных дисперсий. Затем авторы использовали *t*-критерий Стьюдента, применяя его поочередно для разных пар групп сравнения, что для данного критерия без использования специальных поправок недопустимо. В результате такого приема возникает так называемая «ошибка множественных сравнений» (см. [34], с. 113), приводящая к тому, что авторы будут чаще обнаруживать желаемое различие, нежели оно существует на самом деле. Данная ошибка является второй по распространенности после использования критерия Стьюдента без проверки двух имеющихся ограничений. Из чего следует, что выводы авторов такой работы не могут быть признаны корректно обоснованными методами статистики, а стало быть, их надежность и ценность сомнительны. В итоге ошибочные выводы этой работы вводят в заблуждение не только самих авторов, но и читателей, на которых к тому же действует магия академических регалий авторов. «Врачам известно множество методов диагностики и лечения, эффективность которых была «доказана» статистическими методами и которые, тем не менее, канули в Лету, не выдержав испытания практикой. ... Вред, причиняемый ошибками такого рода, очевиден. Исследователь заявляет о «статистически достоверном» эффекте лечения, редактор помещает статью в журнал, врач, неспособный критически оценить публикацию, применяет неэффективный метод лечения. В конце этой цепи находится больной, который расплачивается за все, подвергаясь ненужному риску и не получая действительно эффективного лечения. Не

Таблица 4. Результаты проверки гипотез о равенстве групповых средних различными критериями

Исследуемая количественная переменная	Z9	Z15	X21	Z14	Z16	X21	Z9
Группирующая переменная	A5	A7	A5	A12	A7	A10	A6
Значения групповых средних	10,17	141,7	0,14	37,4	0,77	1,24	8,10
	8,21	107,9	0,11	19,9	0,72	1,04	9,15
Величина t-критерия Стьюдента для случая равных дисперсий	2,94	1,98	2,69	1,57	2,23	2,09	1,84
Величина t-критерия для случая неравных дисперсий	1,76	2,54	2,03	2,93	1,85	1,73	1,77
Величина достигнутого уровня значимости p для t-критерия Стьюдента при равных дисперсиях	0,0048	0,0138	0,0107	0,1235	0,0297	0,0401	0,0714
Величина достигнутого уровня значимости p для t-критерия по Саттерзвайту при неравных дисперсиях	0,1049	0,0665	0,0755	0,0307	0,0830	0,1229	0,0851
Величина достигнутого уровня значимости p для t-критерия по Кохрану—Коксу при неравных дисперсиях	0,1059	0,0693	0,0797	0,0482	0,0866	0,1259	0,0879
Значение F-критерия Фишера при проверке гипотезы о равенстве дисперсий	9,29	2,57	2,71	4,55	2,00	1,59	5,23
Величина достигнутого уровня значимости p для F-критерия Фишера при проверке гипотезы о равенстве дисперсий	0,0000	0,023	0,055	0,24	0,10	0,30	0,0000

Примечание. X21 — отношение iR (периода изометрического расслабления) к длительности интервала RR на электрокардиограмме при пробе с дипиридамолом (ПД) перед началом лечения;

Z9 — толщина межжелудочковой перегородки после лечения, мм;

Z14 — конечный систолический объем в покое после лечения, мл;

Z15 — ударный объем в покое после лечения, мл;

Z16 — фракция выброса в покое после лечения, усл. ед.;

A5 — наличие (отсутствие) депрессии сегмента ST при велоэргометрии;

A6 — наличие (отсутствие) депрессии сегмента ST при пробе ПД;

A7 — наличие (отсутствие) обызвествления коронарных артерий;

A10 — наличие (отсутствие) нарушений сердечного ритма;

A12 — наличие (отсутствие) нарушений сердечного ритма при пробе ПД.

следует сбрасывать со счетов и ущерб от самого факта проведения бессмысленных исследований. Деньги и подопытные животные приносятся в жертву науке, больные рискуют ради сбора ошибочно интерпретируемых данных» [34, с. 24]. Полагаю, что авторам подобных некачественных исследований и публикаций стоит почаще вспоминать известный принцип «Не навреди»....

О том, сколь значительно могут отличаться значения t-критерия и достигнутого уровня значимости p в случае ошибочного его использования, можно судить по результатам статистического анализа, проведенного автором этих строк на реальных данных, полученных в отделении ИБС НИИ кардиологии Томского научного центра СО РАМН (табл. 4).

Как видим, для исследованных переменных принятие или отклонение гипотезы о равенстве генеральных средних в группах во многом определяется результатом проверки гипотезы о равенстве генеральных дисперсий.



**«... НЕ ЗНАЯ ЗАКОНОВ ЯЗЫКА
ИРОКЕЗСКОГО, МОЖЕШЬ ЛИ ТЫ
ДЕЛАТЬ ТАКОЕ СУЖДЕНИЕ ПО СЕМУ
ПРЕДМЕТУ...»**

Как блестящие идеи, так и научные нелепости одинаковым образом можно облечь во впечатляющий мундир формул и теорем.

В.В. Налимов

В ряде случаев встречаются достаточно одиозные ошибки, которые демонстрируют полнейшее непонимание их авторами сути используемых методов. Чаще всего они связаны с использованием сложных статистических методов. Первый пример заимствован нами из диссертации [66] (http://www.biometrika.tomsk.ru/kk/index_3.htm#39): «Статистическая обработка данных ... проводилась на основе пакета ... анализа биомедицинской информации BMDP ... Используются вариационный, корреляцион-

Таблица 5. Факторные вклады, определяющие клиническую форму пневмонии [66]

Вклад фактора	Признак и его характеристика	Коэффициент корреляции
I (22,6%)	Аномалии иммунитета	0,724
	Стигмы дисэмбриогенеза	0,625
	Аллергические диатезы	0,574
	Частые респираторные заболевания	0,502
II (16,0%)	Хронические очаги в носоглотке	0,702
	Паразитарные инвазии	0,616
	Преморбидный фон (без иммунопатологии)	-0,599
III (11,2%)	Аллергические реакции	0,768
	Этиология (возбудитель)	0,598
IV (10,3%)	Неблагоприятное микросоциальное окружение	0,900

ный, регрессионный, факторный, кластерный анализы на ЭВМ ЕС 1033. ... Проверка достоверности выявленных различий осуществлялась по критерию Стьюдента...».

Стр. 150: «В результате проведенного анализа 224 случаев пневмонии на древовидной схеме выделилось 5 кластеров (см. рис. 5.10), соответствовавших 5 группам больных».

Проанализируем эти и некоторые другие фрагменты диссертации.

1. Из содержания диссертации непонятно, что подразумевает автор под «**вариационным анализом**».

2. В диссертации имеется ряд таблиц, содержащих результаты факторного анализа. Так, на с. 69 автор сообщает: «При выявлении факторных вкладов, определяющих приоритетное влияние на критерий «клиническая форма», установлено, что преобладают четыре главных фактора, составляющие 60,1% объясняемой дисперсии, т.е. степени воздействия их на описываемый критерий» (табл. 5).

Как видно из содержания второго столбца таблицы, автор использует признаки дискретные, качественные, имеющие несколько альтернативных вариантов ответа. Факторный же анализ базируется на анализе **непрерывных, количественных признаков, связанных между собой линейными корреляционными соотношениями**. Обратимся к описанию процедуры факторного анализа Р4М, приведенной на с. 21—30 второй части 44-го выпуска сборника «Программное обеспечение ЭВМ» (Минск, 1983), содержащего описание упоминаемого диссертантом пакета ВМДР. «Программа Р4М выполняет факторный анализ **корреляционной или ковариационной матрицы**, воспроизводя ее в пространстве меньшей размерности с помощью матрицы факторных нагрузок и корреляций между факторами. Так как процедура выделения факторов имеет неоднозначное решение, то для введения ограничений, связанных с возможностью дать каждой переменной в пространстве общих факторов простое факторное объяснение, производится вращение факторов». Исходя из этого можно утверждать, что полученные диссертантом результаты факторного анализа совершенно искаженно отражают

истинную структуру связей признаков. Более того, поскольку в данной процедуре может быть использовано достаточно большое количество алгоритмов вращения факторов, то и приведенное автором решение является одним из многих. При выборе другого алгоритма вращения мы получим другие вклады факторов, а также и другие коэффициенты связи исходных признаков и фактора, а значит и другие приоритетные признаки. Автор не приводит своих аргументов в пользу выбора данного решения по сравнению с другими аналогичными. Не уточняет автор и выбранный им алгоритм вращения факторов.

3. Аналогичные вопросы выбора приводимого в диссертации решения из многих возможных вариантов возникают и при знакомстве с результатами кластерного анализа.

Как и другие многомерные методы, процедура кластерного анализа может породить множество решений. Каждое из таких решений определяется уникальной комбинацией нескольких параметров, основными из которых являются подмножество используемых признаков, мера расстояния между наблюдениями, мера расстояния между кластерами и алгоритм кластеризации. Кроме того, разномасштабность количественных признаков, а также совместное использование количественных и качественных признаков еще более усложняют выбор этой комбинации. Приведенное диссертантом решение является одним из очень многих возможных, и без его сравнения с другими аналогичными решениями утверждать его предпочтительность весьма опрометчиво.

4. В тексте диссертации не проводятся результаты проверки условий допустимости применения t-критерия Стьюдента.

Мы проверили гипотезы о равенстве генеральных дисперсий, случайно выбрав одну из таблиц [66, табл. 5.15, с. 168], на основании которой диссертант делает заключения о различии средних. Результаты проверки гипотез о дисперсиях показывают, что для многих сравниваемых пар генеральные дисперсии статистически значимо различаются. **В одном из таких случаев критерий Фишера $F=1202$, т.е. различие дисперсий более чем в 1000 раз!** Таким образом, мы вновь

имеем дело с ситуацией, когда критерий Стьюдента неприменим. Из этого следует, что выводы диссертанта по данной таблице сомнительны. Описанное на примере одной диссертации справедливо в отношении большинства случаев применения факторного анализа в медицинских исследованиях.

Второй пример связан с рукописью, поступившей в «Сибирский медицинский журнал». Цель исследования заключалась в том, чтобы «...количественно оценить частоту и динамику симптомов тревоги, депрессии и болевого синдрома при ежедневном наблюдении у больных гипертонической болезнью в условиях стационара». Для этого за группой из 42 пациентов с артериальной гипертензией наблюдали в течение 16—17 дней, измеряя ряд показателей, характеризующих состояние больных. В результате получился многомерный временной ряд, часть показателей которого имела количественную природу (артериальное давление), а другие показатели оценивались в баллах. Окончив эти наблюдения, авторы объединяют 42 временных ряда в один временной ряд, создав таким образом новый ряд длительностью уже в 678 дней! И далее этот новый ряд начинают анализировать, причем не всегда адекватными методами. Основная ошибка здесь заключается в самой процедуре объединения 42 временных рядов в один. Получается, что новый ряд есть результат наблюдения за неким «усредненным пациентом» в течение 678 дней! Некорректность этого приема столь очевидна, что не требует детального рассмотрения. Укажем только на самые явные противоречия. Во-первых, наблюдения за 42 пациентами могли быть проведены в разные периоды течения заболевания. Во-вторых, существует масса вариантов объединения 42 рядов в один ряд. Смена последовательности следования пациентов в суммарном ряду будет изменять результаты анализа такого ряда. В-третьих, искусственное увеличение протяженности временного ряда многократно увеличивает число степеней свободы, столь же искусственно повышая точность и надежность корреляционных коэффициентов и остальных оценок параметров.

Наша критика в адрес авторов, безграмотно применяющих критерий Стьюдента, нередко приводит к курьезным последствиям. В тех случаях, когда этот критерий неприменим, авторы используют другой критерий, незаслуженно приписывая ему свойства критерия Стьюдента. Проанализируем цитату из доклада [67]: «Статистический анализ показал, что по ряду цитогенетических показателей имеется несоответствие распределения эмпирических данных нормальному закону и поэтому **t-критерий Стьюдента не может использоваться для сравнения средних значений выборок. Более адекватным для этой цели является непараметрический критерий однородности Колмогорова—Смирнова. ... приблизительно у 70% обследованных работников Северского химического комбината (СХК) и жителей г. Северска отмечается существенное повышение частоты хромосомных aberrаций не-радиационной природы**

Как видим, используя критерий Колмогорова—Смирнова, авторы делают вывод о существенном повышении средней частоты хромосомных aberrаций у определенной категории лиц. Однако данный вывод весьма сомнителен. В отличие от критерия Стьюдента, критерий Колмогорова—Смирнова чувствителен к любому различию в параметрах распределений: «... двухвыборочный критерий Колмогорова—Смирнова чувствителен не только к различию в положении двух распределений, но также и к форме распределения. **Фактически он чувствителен к любому отклонению от гипотезы однородности, но не указывает, с каким именно отклонением мы имеем дело**» [56].

Рассмотренные выше особенности этого критерия проиллюстрируем с помощью массива данных, в котором имеются две группы наблюдений, в каждой из которых по 100 наблюдений. Количественный признак представлен переменной VAR1, группирующий признак — VAR2. Вторая группа была сконструирована из первой группы наблюдений следующим образом. После сортировки признака VAR1 в первой выборке по возрастанию к первым 50 наблюдениям прибавили одно и то же число, тогда как от следую-

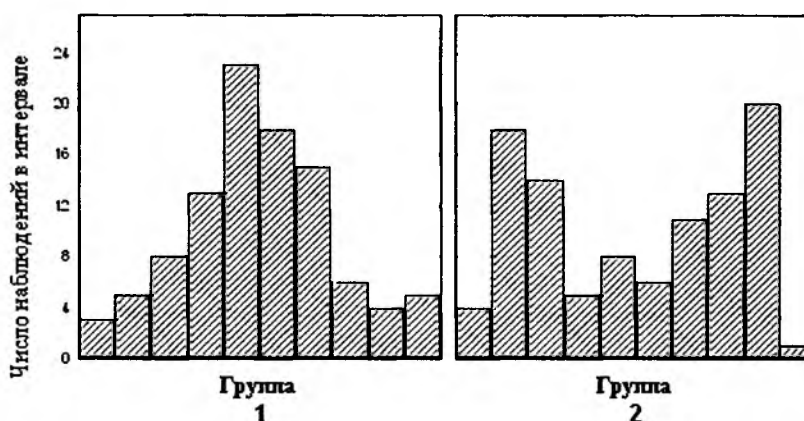


Рис.1. Гистограммы распределения двух сравниваемых групп.

STATISTICA. Wald-Wolfowitz Runs Test (Nazarenko.STA)
NONPAR By variable VAR2
STATS Group 1: 1 Group 2: 2

Valid N Group 1	Valid N Group 2	Mean Group 1	Mean Group 2	Z	p-level	Z adjstd
100	100	9,965234	9,965234	-3,68623	0,000228	3,615340

Рис. 2. Тест Вальда—Вольфовица

S-PLUS 2000. Two-Sample Kolmogorov-Smirnov Test

data: x: V1 in Nazarenko.SAV, and y: V2 in Nazarenko.SAV

$D_{KS} = 0,22$, $p\text{-value} = 0,0156$

alternative hypothesis:

cdf of x: V1 in Nazarenko.SAV does not equal the

cdf of y: V2 in Nazarenko.SAV for at least one sample point.

Summary Statistics for data in:
Nazarenko.SAV

	Group 1	Group 2
Min (Минимум)	7,526174	7,931365
1-st Qu (1-я квартиль)	9,358504	8,554207
Mean (Среднее)	9,965234	9,965234
Median (Медиана)	9,931250	10,101192
3-rd Qu (3-я квартиль)	10,550065	11,350258
Max (Максимум)	12,374363	11,931135
Total N (Общее число наблюдений)	100	100
Std Dev (Стандартное отклонение)	1,006245	1,393181

Рис. 3. Результаты применения двухвыборочного теста Колмогорова—Смирнова в пакете S-PLUS 2000.

щих 50 наблюдений это же самое число вычли (см. http://www.biometrika.tomsk.ru/lib/k_s.htm). Полученные при этом новые значения признака VAR1 и составили вторую группу. В результате этой операции обе группы имеют идентичные средние (9,965234) и равные суммы значений (996,5234) переменной VAR1 в каждой из этих групп. Отличие же их заключается в том, что первая группа имеет нормальное распределение с унимодальной гистограммой, тогда как вторая группа имеет ярко выраженное антимодальное распределение (провал в центре и по одной моде слева и справа) (рис. 1).

Проведем сравнение этих групп, используя непараметрические статистики. Тест Вальда—Вольфовица указывает на статистически значимое различие двух функций распределения (рис. 2).

О таком же различии говорят и результаты применения двухвыборочного теста Колмогорова—Смирнова в пакете S-PLUS 2000 (рис. 3).

Однако все сравнения средних указывают на их идентичность, т.е. нулевая гипотеза равенства генеральных средних принимается при отвержении гипотезы идентичности функций распределения! Таким образом, утверждение авторов этого доклада о том, что «... приблизительно у 70% обследованных работников СХК и жителей г. Северска отмечается существенное повышение частоты хромосомных aberrаций перадиационной природы» вполне может быть ошибочным! Это может быть любым другим различием, напри-

мер, различием дисперсий, различием мер формы и т.д. После разбора этого примера в разделе «Кунсткамера» сайта «Биометрика» (http://www.biometrika.tomsk.ru/kk/index_4.htm#64) один из авторов доклада потребовал у нас публичного извинения за моральный ущерб, однако мое предложение вынести нашу переписку по данному вопросу на суд читателей отверг.



ЧТО ЖЕ ДЕЛАТЬ, ЧТОБЫ ИЗБЕЖАТЬ ОШИБОК?

Свежесть бывает только одна — первая, она же и последняя.

А если осетрина второй свежести, то это означает, что она тухлая!

М.А. Булгаков (Мастер и Маргарита)

Уже в XX веке стало ясно, что наука становится все более междисциплинарной. В особенности это проявляется в биологии и медицине, куда все больше проникают методы точных наук. Поэтому наиболее радикальное средство избежать ошибок в использовании статистических методов — выполнение этого этапа медицинских исследований профессиональным биостатистиком. Медик или биолог, занятый экспериментальными исследованиями, конечно, должен владеть основными идеями современной прикладной статистики. Столь же очевидно, что он должен владеть и терминологией данной предметной области. Это облегчит ему общение с профессиональным биостатистиком. Всему этому должны учиться исследователей на этапе аспирантуры и докторантуры. Но, с другой стороны, и профессиональный статистик, выполняющий анализ биомедицинских данных, также должен частично проникать в соответствующую предметную область медика или биолога. Такой симбиоз возможен лишь в том случае, когда общение происходит не эпизодически, а постоянно, и является рабочей необходимостью. Достичь же этого возможно только путем создания специализированных лабораторий биостатистики в медицинских НИИ и вузах. Именно по такому пути и идут медицинские исследовательские центры за рубежом [7].

Попытки медика самостоятельно овладеть огромным инструментарием, который доступен сегодня профессиональному статистiku, в конечном счете обречены на неудачу. Статистика столь же узкоспециальная отрасль знания, как и кардиология, гинекология, психиатрия, физиология и т.д.

На последней странице книги «Теория эксперимента» (М: Наука 1971) В.В. Налимов пишет: «... Почему так часто язык статистики используется столь

вульгарно? ... Все дело в том, что язык математической статистики еще не стал родным языком экспериментатора. Экспериментатору пользоваться им так же трудно, как каждому из нас пользоваться мало-знакомым иностранным языком. ... Мы, конечно, далеки от мысли о том, что каждый экспериментатор может стать и специалистом по математической статистике. В будущем, по-видимому, нужно стремиться к созданию двухступенчатой системы организации исследований. Экспериментатор, понимающий идейную сторону теории эксперимента, должен получить

возможность постоянно взаимодействовать с консультантом. Возникает необходимость в появлении специалистов нового профиля — статистиков-консультантов». Именно такой междисциплинарный союз в дальнейшем и позволит читателю статьи или диссертации не только видеть текст, но и адекватно понимать его.

Полные описания работ, содержащих рассмотренные выше ошибки, приведены в разделе «Кунсткамера» сайта «Биометрика» (<http://www.biometrika.tomsk.ru/kk.htm>).

Литература

1. Рейтман М. Наука «с кондачка» и жизнь. URL: <http://www.lebed.com/1998/art439.htm>.
2. Логинов В.С., Дорошенко С.И. Мошенничество в науке. Химия и жизнь 1992;3:22—25.
3. Ермолаева Н.И. Еще раз о «гороховых законах». Яровизация 1939;2:79—86.
4. Колмогоров А.Н. Об одном новом подтверждении законов Менделя. Доклады Академии Наук СССР 1940;27. URL: <http://www.biometrika.tomsk.ru/kolmogorov/kolmogor1.htm>.
5. Лысенко Т.Д. По поводу статьи академика А.Н. Колмогорова. Доклады Академии Наук СССР 1940;28.
6. Леонов В.П. Долгое прощание с лысенковщиной: Статистика в опале. URL: <http://www.biometrika.tomsk.ru/lis/index10.htm>.
7. Леонов В.П. Искушение Знанием или «Сколько стоит биометрика?». URL: http://www.biometrika.tomsk.ru/price_biom.htm.
8. Боярский А.Я. Статистические методы в экспериментальных медицинских исследованиях. М: Медгиз 1955.
9. Dunn H.L. Application of statistical methods in physiology. Physiological Reviews 1929;9:275—398.
10. Greenwood M. What is wrong with the medical curriculum? Lancet 1932;1269—1270.
11. McGuigan S.M. The use of statistics in the British Journal of Psychiatry. Br.J. Psychiatry 1995;167:689—690.
12. Welch G.E., Gabbe S.G. Review of statistics usage in the American Journal of Obstetrics and Gynecology. Am J Obstet Gynecol 1996;175(5):1138—1141.
13. Porter A.M. Misuse of correlation and regression in three medical journals. J R Soc Med 1999;92(3):123—128.
14. White S.J. Statistical errors in papers in the British Journal of Psychiatry. British Journal of Psychiatry 1979;135:336—342.
15. Lewis J.A., Facey K.M. Statistical shortcomings in licensing applications. Statistics in Medicine 1998;17:1663—1673.
16. Banik N. Biostatistical methodology: Part 2. Clinical trials in the 21st century — a means of knowledge generation or self-fulfilling prophecy? URL: <http://www.drugdev123.com/editorial/archive/DataProcessing/159.shtml>
17. Bailer J.C. Bailer's laws of data analysis. Clin Pharmacol Ther 1976;20:113—120.
18. Rosenfeld R.M., Rockette H.E. Biostatistics in otolaryngology journals. Arch Otolaryngol Head Neck Surg 1991;117:1172—1176.
19. Wang Q., Zhang B. Research Design and Statistical Methods in Chinese Medical Journals. JAMA 1998;280:283—285.
20. Emerson J.D., Coldit G.A. Use of statistical analysis in the New England Journal of Medicine. N Engl J Med 1979;309:709—713.
21. Gardenier J.S. The Misuse of Statistics: Concepts, Tools, and a Research Agenda. URL: <http://ori.dhhs.gov/multimedia/acrobat/papers/gardenier.pdf>.
22. Kirkwood J.M., Ibrahim J., Sondak V.K., Ernstoff M.S., Flaherty L., Haluska F.J. Use and Abuse of Statistics in Evidence-Based Medicine. J Clin Oncol 2002;20(7):4118—4125.
23. Statistics with confidence. 2 ed. London: British Medical Journal 2000.
24. Altman D.G. Statistics in medical journals: some recent trends. Statistics in Medicine 2000;19:3275—3289.
25. Altman D.G. Statistics in medical journals: developments in the 1980s. Statistics in medicine 1991;10:1897—1913.
26. Altman D.G. Statistics in medical journals. Statistics in medicine 1983;1:59—71.
27. Altman D.G. Statistics and ethics in medical research. Misuse of statistics is unethical. BMJ 1980;281:1182—1184.
28. Altman D.G., Gore S.M., Gardner M.J., Pocock S.J. Statistical guidelines for contributors to medical journals. BMJ 1983;286:1489—1493.
29. Altman D.G. The scandal of poor medical research. BMJ 1994;308:283—284.
30. Altman D.G., Gooman S.N., Schroter S. How Statistical Expertise Is Used in Medical Research. JAMA 2002;287:2817—2820.
31. Altman D.G., Cates C. Authors should make their data available. BMJ 2001;323:1069.
32. Леонов В.П. Применение методов статистики в кардиологии (по материалам журнала «Кардиология» за 1993—1995 гг.). Кардиология 1998;1:55—58.
33. Altman D.G. Poor-quality of medical research: What can journal do? JAMA 2002;287:2765—2767.
34. Гланц С.А. Медико-биологическая статистика. М: Практика; 1999.
35. Квалификационная характеристика специалиста с высшим биологическим образованием. Приложение 15 к приказу Минздрава РФ №380 от 25.12.97. Здравоохранение 1998;7:91—102.
36. Bostanci A. Research ethics. Germany gets in step with scientific misconduct rules. Science 2002; 296(5574):1778.
37. Hagmann M. Panel Finds Scores of Suspect Papers in German Fraud Probe. Science 2000;288:23—26.
38. Леонов В.П. Долгое прощание с лысенковщиной. Меметический анализ описаний методов статистики. URL: <http://www.biometrika.tomsk.ru/lis.htm>
39. Hutchon D.J.R. Publishing raw data and real time statistical analysis on e-journals. BMJ 2001;322:530.
40. Леонов В.П. Три «Почему ...» и пять принципов описания статистики в биомедицинских публикациях. URL: <http://www.biometrika.tomsk.ru/principals.htm>.
41. Edited by Thomas A.Lang, Michelle Secic. How to report statistics in medicine: a guide for authors, editors and reviewers. American College of Physicians; 1997.
42. Жерлов Г.К. Арефлюксные анастомозы при операциях на желудке (клинико-экспериментальное исследование). Дис. ... д-ра мед. наук. Томский медицинский институт, Томск 1991.

43. *Лакин Г.Ф.* Биометрия. Москва: Высшая школа 1990.
44. *Розов С.М.* Дарвинизм и эпистемология: генетика и мемутика. На теневой стороне. Материалы к истории семинара М.А. Розова по эпистемологии и философии науки в Новосибирском Академгородке. Новосибирск: Новосибирский университет 1996;311—338.
45. *Зырянов Б.Н., Тузиков С.А., Тихонов В.И.* Адьювантная химиотерапия рака желудка. Томск: ТГУ 1996.
46. *Лукашова Л.В.* Лечение хронического описторхоза препаратом из растительного сырья (попутрилом) и оценка его эффективности. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск 1995.
47. *Красноженов Е.П.* Влияние некоторых стимуляторов центральной нервной системы на неспецифическую иммунобиологическую реактивность организма. Дис. ... канд. мед. наук. Томский государственный медицинский институт, Томск 1969.
48. *Завьялова Н.Г.* Ультразвуковая и радионуклидная диагностика патологии гепатобилиарной системы у детей дошкольного возраста. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск 1997.
49. *Мочалова Д.В.* Клиника и дифференциальная диагностика инфекционных экзантем у детей. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет 1998.
50. WHO. Case definitions for surveillance of severe acute respiratory syndrome (SARS). URL: <http://www.who.int/csr/sars/casedefinition/en/> 2003. Доступность проверена 11.11.2003.
51. *Bliss C.I.* The calculation of the time-mortality curve. *Ann Appl Biol* 1937;24(4):815—852.
52. Статистический словарь. Москва: Финансы и статистика; 1989.
53. Компьютерная биометрика. Под ред. В.Н.Носова. Москва: МГУ 1990.
54. *Браунли К.А.* Статистическая теория и методология в науке и технике. М: Наука 1977.
55. *Барабаш Н.А.* Клинико-диагностические критерии нарушений сердечно-сосудистой системы у детей с тимомегалией. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск 1996.
56. *Боровиков В.* STATISTICA: Искусство данных на компьютере. Для профессионалов. Ст-Петербург: Питер 2001.
57. *Васильченко М.И.* Арсфлюксные гастростомы и еюностомы. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск 1996.
58. *Камалтынова Е.М.* Гормонально-иммунологические параллели у детей с аллергическим поражением кожи и органов дыхания. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск 1994.
59. *Доненко Ф.В., Мороз Л.В.* Феномен стабильности роста опухоли в организме хозяина. Новый подход к концентрации роста и лечения опухоли. Вестник РАМН 1995;4:14—16.
60. *Спиридонова М.Г., Степанов В.А., Пузырев В.П., Карпов Р.С.* Анализ взаимосвязи полиморфизма C677T гена метилентетрагидрофолатредуктазы с клиническими проявлениями атеросклероза. Генетика 2000;9:1269—1273.
61. *Кривова Н.А.* Механизмы образования и дегенерации надэпителиального слизистого слоя пищеварительного тракта. Дис. ... канд. мед. наук. НИИ биологии и биофизики Томского государственного университета, Томск 1994.
62. *Харьков Е.И.* Исходы нестабильной стенокардии на догоспитальном и госпитальном этапах. Дис. ... канд. мед. наук. Красноярский медицинский университет, Красноярск 1987.
63. *Артюхов В.Г., Башарина О.В., Пантак А.А., Свекло Л.С.* Влияние излучения He-Ne лазера на ферментативную активность и оптические свойства каталазы. Бюллетень экспериментальной биологии и медицины 2000;6:633—636.
64. *Шварц Я.Ш., Зубахин А.А., Душкин М.И.* Депрессия кровотока при C14 - индуцированном гепатофиброзе: роль системной эндотоксикемии. Бюллетень экспериментальной биологии и медицины 2000;8:172—175.
65. *Новицкий В.В., Семин И.Р., Рязанцева Н.В.* Поверхностная архитектура эритроцитов периферической крови у психически больных. Бюллетень экспериментальной биологии и медицины 2000;10:429—432.
66. *Климов В.В.* Клиническая оценка реакции нейтрофилов при острой пневмонии у детей. Дис. ... канд. мед. наук. Томский медицинский институт, Томск 1989.
67. *Пузырев В.П., Назаренко С.А., Попова Н.А.* Цитогенетические эффекты ядерно-химического производства. Медицинские и экологические эффекты ионизирующей радиации (к 15-летию аварии на Чернобыльской АЭС). Материалы 1-й международной научно-практической конференции, 21—22 июня 2001 г., Северск—Томск. Ред. Р.М. Тахауов, Л.В. Каплевич, А.Б. Карпов. 2001.
68. *Афанасьева А.Н.* Роль поджелудочной железы в регуляции антитрипсиновой и трипсиновой активности крови. Дис. ... канд. мед. наук. Томск 1990.
69. *Васильев В.Н.* Роль трипсина в регуляции моторной функции тонкой кишки. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск, 1992.
70. *Грибовская Е.А.* Роль поджелудочной железы в регуляции состояния калликреин-кининовой системы крови. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск 1993.
71. *Шумилов С.П.* Функционирование калликреин-кининовой системы крови в норме и при патологии. Дис. ... канд. мед. наук. Сибирский государственный медицинский университет, Томск 1995.
72. *Кривова Н.А.* Механизмы образования и дегенерации надэпителиального слизистого слоя пищеварительного тракта. Дис. ... канд. мед. наук. НИИ биологии и биофизики Томского государственного университета, Томск, 1994.
73. *Леонов В.П.* Когда нельзя, но очень хочется, или Еще раз о критерии Стьюдента. URL: <http://www.biometrika.tomsk.ru/student.htm>.