

УДК [57:61]:519.23

ИСПОЛЬЗОВАНИЕ СТАТИСТИКИ В РОССИЙСКОЙ БИМЕДИЦИНСКОЙ ЛИТЕРАТУРЕ

© 2008 г. А. М. Гржибовский

Национальный институт общественного здоровья, г. Осло, Норвегия
Архангельская международная школа общественного здоровья,
г. Архангельск

В настоящее время в зарубежной литературе опубликовано большое количество статей и обзоров, посвященных анализу качества применения статистических методов обработки данных в биомедицинских исследованиях. Многие журналы регулярно проводят анализ качества собственных публикаций, что в совокупности с введением статистического рецензирования всех присылаемых рукописей и обеспечило наблюдаемый на сегодняшний день тренд увеличения количества опубликованных статей, в которых используются статистические методы обработки данных, параллельно с уменьшением доли работ, содержащих грубые статистические ошибки. Это отнюдь не означает, что все зарубежные публикации стопроцентно безошибочны, однако вероятность обнаружения статистической бессмыслицы в престижных изданиях достаточно невелика.

Анализ методологического качества публикаций за рубежом чаще всего заключается в оценке применения статистических методов, причем подобный анализ проводится не только в наиболее крупных и престижных журналах [34], но и в большинстве индексируемых в «Медлайн» журналах независимо от специальности [28–30, 32, 33, 38, 40, 42, 44, 50–52]. Хочется отметить, что оценка качества публикуемых статей с последующим опубликованием такой оценки в том же журнале характерна не только для всемирно известных журналов, издаваемых в Великобритании и США (в них это уже устоявшаяся традиция), но и для «небольших» журналов, издаваемых в Испании [43], Бразилии [31], Хорватии [41], Китае [49], и т. д. В России эта традиция пока не приживается. Большинство редакторов относятся без особого энтузиазма к идее оценки качества опубликованных в их журналах работ, хотя и здесь не обошлось без исключений. Так, в 1998 г. в журнале «Кардиология» была опубликована совсем нелестная статья о применении статистических методов в кардиологии, основанная на анализе работ, опубликованных в том же журнале в 1993–1995 гг. [20]. Повлияла ли эта публикация на изменение редакционной политики журнала «Кардиология» и изменилось ли качество публикаций после проведенной оценки, автору неизвестно.

Целью данной работы является ознакомление читателей журнала «Экология человека» с некоторыми из обзоров, описывающих качество применения статистических методов обработки данных в отечественных и зарубежных биомедицинских журналах. Кроме того, будет представлено, какие методы наиболее часто использовались авторами журнала «Экология человека» в 2000–2005 гг. Поскольку целью публикации является ознакомление будущих авторов с наиболее часто встречающимися ошибками, а не с авторами, которые эти ошибки допускают, ссылки на отдельные публикации намеренно не приводятся.

Вероятно, первой отечественной публикацией, посвященной качеству обработки данных биомедицинских исследований, была книга А. Я. Боярского «Статистические методы в экспериментальных медицинских исследованиях», выпущенная в 1955 г. (цит. по [19]), в которой автор приходит

Представлен несистематический обзор литературы о наиболее часто встречающихся ошибках и неточностях при использовании статистических методов в российской и зарубежной литературе, а также ошибки при интерпретации результатов применения статистических критериев. Приводится краткое описание способов обработки данных, применяемых авторами журнала «Экология человека» в 2000–2005 гг. Основными характеристиками публикаций можно считать использование простейших типов исследований и методов обработки данных, игнорирование проблемы множественных сравнений, редкий учет и оценка влияния вмешивающихся факторов, а также декларирование выводов, которые не вытекают из имеющихся данных или результатов анализа. Все это приводит к тому, что результаты многих исследований можно считать сомнительными с точки зрения принципов доказательной медицины. Больше всего беспокоит увеличение количества статей, в которых имеет место несоответствие между методами, описанными в соответствующем разделе, и фактически использованными методами обработки данных. Представлены примеры наиболее часто встречающихся ошибок, а также предложения, как их избежать.

Ключевые слова: тип данных, статистические критерии, распределение.

в выводу, что статистическая обработка данных является наиболее слабым местом отечественных медицинских исследований, что отражает недостаточную степень подготовленности исследователей в этой области.

В 1987 г. А. И. Орлов предложил провести критический анализ сложившейся практики обработки данных в медико-биологических исследованиях с последующим созданием службы статистических консультаций в научно-исследовательских учреждениях [25]. Первое из предложений активно реализовывалось в 1990-е гг. [2, 16–24, 27].

Наибольшее количество работ, посвященных оценке качества публикаций, опубликовано В. П. Леоновым и В. В. Власовым, причем если второй больше акцентирует внимание на дизайн исследований [4, 48], то первый детально описывает применение статистических методов, а также ошибки при их использовании или описании, причины ошибок, историю использования статистики в биомедицинских исследованиях в СССР и России, а также возможные пути выхода из сложившейся ситуации [16–24, 27]. Многие из публикаций В. П. Леонова представлены на созданном им сайте доказательной биологии и медицины «Биометрика» <http://www.biometrica.tomsk.ru/index.htm>. Пропагандой доказательной медицины, грамотного планирования исследований, обработки данных и представления результатов в Украине активно занимается К. П. Воробьев [5–8], чьи работы находятся в свободном доступе в интернете на сайте <http://www.vkr.dsip.net/Main.htm>. Читателям, особенно аспирантам, будет полезно узнать, что на сайте <http://www.biometrica.tomsk.ru/index.htm> существует специальный раздел «Кунсткамера», в котором В. П. Леонов «выставляет» наиболее неграмотные диссертации с детальным анализом ошибок обработки и представления данных. Ознакомление с этим разделом рекомендуется всем будущим авторам «Экологии человека» без исключения, хотя бы с целью профилактики попадания в него, а также как стимул для ознакомления со специальной статистической литературой или приглашения биостатистика для совместного планирования исследования и обработки данных для будущей диссертации.

Всего В. П. Леоновым было проанализировано более 1 500 статей, опубликованных в 1987–1997 гг. в таких журналах, как «Кардиология», «Бюллетень экспериментальной биологии и медицины», «Вестник РАМН», «Проблемы эндокринологии», «Клиническая медицина», «Радиационная биология. Радиоэкология» и др., а также более 250 диссертаций биомедицинской тематики. Автор отмечает, что примерно в четверти (!) работ, в которых использовался статистический анализ, вообще не сообщается о примененных методах, но в то же время эти статьи содержат большое количество выражений типа « $p < 0,05$ », что позволило автору говорить о «самозарождении p » [16]. Оставшиеся три четверти работ, в которых описывается применение статистических методов, автор условно делит на 3 группы [19, 21]. В первую попадает чуть меньше половины работ, в которых просто констатируется сам факт использования статистических методов, например, «результаты обработаны статисти-

чески»; для второй группы, составляющей почти треть статей, характерно указание некоторых деталей (упоминание справочника по статистике, применяемого критерия или статистического пакета); в третью группу, (приблизительно четверть работ) В. П. Леонов объединяет статьи, содержащие «безграмотные, а подчас и просто абсурдные либо противоречивые сочетания перечисляемых терминов или критериев» (цит. по [19]). Таким образом, половина статей содержит либо безграмотное описание статистических методов, либо вообще не содержит никакого описания. Оставшаяся половина также в целом имеет проблемы с описанием, так как просто ссылки на справочник и простого упоминания критерия недостаточно для аргументированного использования того или иного метода.

Большое внимание В. П. Леонов уделяет меметическому анализу описания методов статистики, особенно так называемым «камуфляжным мемам», которые представляют собой ошибочное, но с использованием большого количества различных статистических терминов описание методов обработки данных, причем целью такого описания является не подробное разъяснение читателям, какие методы применялись, а всего лишь «обнаучивание» нередко ошибочных выводов авторов. Часто встречающееся неверное использование реально существующих терминов, таких, как, например, «доверительная вероятность», «значимость», «уровень значимости», или изобретение авторами безграмотных комбинаций позволили В. П. Леонову обобщить сложившуюся ситуацию цитатой из М. В. Ломоносова: «Смутно пишут о том, о чем смутно представляют» (цит. по [19]).

Из применяемых российскими авторами статистических критериев или терминов наиболее часто встречается описательная статистика, чаще всего в виде $M \pm m$, выражения типа « $p < 0,05$ » и критерий Стьюдента. Очень редко используется критерий хи-квадрат, вероятно, потому что многие исследователи необоснованно сравнивают доли с помощью критерия Стьюдента. Также редко используется дисперсионный анализ, опять же, вероятно, потому что в тех ситуациях, где он должен бы применяться, используется критерий Стьюдента. Еще одним важным наблюдением является то, что в российской биомедицинской литературе чрезвычайно редко используются многомерные методы обработки данных, что свидетельствует о доминировании в отечественной биомедицинской науке так называемой «сдвиговой парадигмы» [16, 24], то есть ситуации, когда исследователи ограничиваются простым сравнением двух средних, разумеется, с помощью критерия Стьюдента.

Краткий анализ статей, опубликованных в журнале «Терапевтический архив» в 1999 г., показал очень схожие результаты. Из критериев, предназначенных для проверки гипотез, чаще всего применялся критерий Стьюдента, а связи между переменными чаще всего оценивались с помощью корреляционного анализа с расчетом коэффициента корреляции Пирсона [10], причем практически ни в одной из статей не проверялись ограничения применения того или иного критерия, а результаты анализа были зачастую представлены не так, как рекомендуется [9].

Отдельной проблемой является употребление термина «достоверность». В большинстве отечественных статей используется термин «достоверность» вместо термина «значимость», что может приводить к искажению восприятия научной публикации в сторону большего доверия к ней читателя. С подробным наукометрическим, лингвистическим и семиологическим анализом необоснованного употребления термина «достоверность» [12] можно ознакомиться на сайте <http://www.biometrika.tomsk.ru/index.htm>.

Сложившейся ситуации с применением статистических методов в российских биомедицинских исследованиях, а также опыту практической помощи исследователям в области обработки данных был посвящен целый номер «Международного журнала медицинской практики» в 2006 г., после прочтения которого складывается впечатление, что биостатистика является наименее научной областью из всей биомедицинской науки [26]. Это неудивительно, так как статистика не может чего-либо доказать, она лишь может оценить степень неопределенности выводов. Авторы сходятся во мнении, что большинство врачей имеют недостаточный уровень знаний в области биостатистики, работают в основном по шаблону, предложенному научным руководителем, не всегда могут определить конкретные цели и задачи исследования, статистические методы используются для подгонки результатов под теорию, а величина r является неким ритуально-камуфляжным атрибутом [1, 3, 11, 13, 15, 18, 26]. Вместе с тем совершенно справедливо отмечается, что ошибки статистического анализа представляют собой лишь вершину айсберга, потому что никакой, даже самый совершенный, статистический анализ неспособен исправить ошибки, допущенные при планировании исследования, поэтому необходима комплексная подготовка исследователей в области планирования и проведения научного исследования, а сотрудничество со специалистом в области биостатистики должно начинаться еще на этапе планирования исследования. В противном случае ситуация еще долго будет напоминать существующую сейчас, когда аспирант обращается к статистику с просьбой «посмотреть и найти хоть что-нибудь». А. Г. Барт с соавторами обращают внимание на то, что включение России в Болонский процесс может в ближайшее время привести к тому, что будет обращено внимание на недостаточное образование российских врачей в области биостатистики [1]. Так, из 123 диссертаций, защищенных в СПбГМУ им. И. П. Павлова в 1998–2004 гг., только в 13 % методы обработки данных соответствовали поставленным в исследовании задачам.

В качестве выхода из сложившейся ситуации большинство авторов предлагают создавать центры биостатистики в медицинских вузах и научно-исследовательских институтах [1, 13, 18], как это предлагал более 20 лет назад профессор Орлов [25]. Следует отметить, что подобные центры или лаборатории уже существуют в Санкт-Петербурге, Екатеринбурге и некоторых других городах России. Кроме того, ситуацию невозможно изменить без ужесточения требований к рукописям, представляемым в биомедицинские журналы, и введения статистического рецензирования как статей, так и

диссертаций на уровне ВАК РФ [13], что неоднократно предлагалось сторонниками доказательной медицины (см. www.biometrika.tomsk.ru). Преподаванию биостатистики для аспирантов также должно придаваться большое значение. Обучение аспирантов основам статистики уже несколько лет существует, например, в Санкт-Петербургской академии постдипломного образования.

Вероятно, у читателей сложилось впечатление, будто все, что публикуется в России, — плохо, а за рубежом — хорошо, однако так ли это?

Несомненно, количество используемых статистических методов и уровень их сложности постепенно возрастает, причем это отмечено не только для журналов, выпускаемых в Западной Европе и США, но и, например, китайских изданий [49].

Для некоторых зарубежных изданий, кстати индексированных в «Медлайн», характерны те же особенности, что и для многих российских. Так, в испанском журнале «Revista Clinica Espanola» в 1990-е гг. в подавляющем большинстве статей использовалась только описательная статистика или простейшие статистические критерии, правда, достаточно часто использовались непараметрические критерии [43]. Приблизительно такая же ситуация характерна и для бразильского журнала «Arquivos Brasileiros de Oftalmologia», но в 2000-е гг., причем отмечается, что почти каждая вторая статья содержит те или иные статистические ошибки [31].

В «American Journal of Tropical Medicine and Hygiene» в 1988 г. почти три четверти статей содержали минимум одну статистическую ошибку, причем наиболее популярным было неправильное использование понятий «стандартная ошибка» и «стандартное отклонение» [32]. Также отмечалось, что для понимания публикаций журнала достаточно лишь базового курса по статистике, так как большинство статей было написано с использованием простейших статистических процедур.

В «Journal of Infectious Diseases» в 1982 г. в 95 % статей, где использовались статистические методы, они не были описаны, а об их существовании читателям предлагалось догадываться по наличию величин r в тексте. Наиболее частой ошибкой было отсутствие коррекции на многократность проверки гипотез [42].

В журнале «Infection and Immunity» 54 % статей, опубликованных в первой половине 2002 г., имели либо ошибки статистического анализа (20 %), либо ошибки описания анализа и представления результатов (22 %), либо их комбинацию (12 %) [44]. Наиболее частыми ошибками анализа были отсутствие коррекции на множественные сравнения, декларирование выводов без подтверждения их расчетами и использование статистических тестов, подразумевающих нормальное распределение данных в ситуациях, когда данные имеют выраженное скошенное распределение. Из ошибок описания наиболее часто встречалось неграмотное использование мер рассеяния, а также неполное описание проводимых статистических процедур.

Анализ использования статистики авторами офтальмологических журналов показал, что статистика в том или ином виде использовалась в двух третях статей, причем центральные тенденции описывались в 65 %

статей, меры рассеяния — в 50 % работ, а самым популярным методом для проверки гипотез был критерий Стьюдента, встретившийся в 20 % публикаций в изучаемых журналах в 1990 г. [38].

Оценка фармакологических журналов показала [40], что практически во всех статьях используется статистика в том или ином объеме. В большинстве работ используется только дескриптивная статистика в виде описания долей (90 %), средних величин (74 %) и стандартных отклонений (58 %). Наиболее часто применяемым критерием был критерий хи-квадрат (33 % статей), за ним следовали критерий Стьюдента (26 %), корреляционный анализ с расчетом коэффициента корреляции Пирсона (18 %), дисперсионный анализ (14 %) и логистический регрессионный анализ (11 %).

Анализ американских радиологических журналов начала 1990-х гг. показал, что почти в половине статей статистика либо вообще не используется, либо используется только для описания долей или центральных тенденций, а наиболее популярным критерием для проверки гипотез является критерий Стьюдента [33]. Лишь в 14 % статей применялось три или более различных статистических методов, в 15 % статей методов было два, а в 27 % применялся только один статистический критерий. Исследования знаний врачей в области статистического анализа данных, проведенные в начале 1980-х гг., показали, что большинство врачей не в состоянии критически оценивать научные публикации, содержащие результаты применения статистических методов [53]. Однако это было более 25 лет назад, и с тех пор ситуация изменилась.

Интересно, что для 1980-х и начала 1990-х гг. была характерна выраженная ассоциация престижности журнала с долей опубликованных в нем статей, содержащих статистические выкладки. Сейчас эта связь значительно менее выражена, так как в подавляющем большинстве журналов приветствуются только те статьи, в которых выводы основываются на расчетах, однако в более престижных журналах чаще используются (причем правильно) более сложные методы анализа данных. При этом всегда следует помнить, что само по себе использование сложных методов анализа не гарантирует повышения практической ценности научной работы, особенно если этот анализ выполнен недостаточно качественно.

Для того чтобы понимать три четверти статей, опубликованных в самом престижном в мире «New England Journal of Medicine», врачам в 1983 г. было достаточно иметь представление об описательной статистике и критерии Стьюдента [34]. Сегодня сложно себе представить, что этот журнал примет к публикации оригинальную статью, в которой не используются многочисленные методы статистики.

Редакция американского журнала «Obstetrics and Gynecology» сначала провела оценку способностей собственных рецензентов оценить качество статистической обработки данных [45], после чего с 1993 г. ввела статистическое рецензирование рукописей. Параллельная оценка рецензентом по статистике и остальными рецензентами выявила, что практически каждая шестая статья, одобренная рецензентами по специальности, со-

держала недопустимые ошибки, связанные с обработкой данных [46]. Анализ опубликованных в том же журнале в 1994 г. статей показал, что более чем в половине статей допускались те или иные статистические ошибки, не позволяющие принять на веру результаты исследований [50]. Наиболее популярными критериями были критерий Стьюдента, критерий хи-квадрат, точный критерий Фишера, дисперсионный анализ и линейный регрессионный анализ. Повторный анализ шесть лет спустя показал: доля статей с адекватным представлением результатов статистической обработки данных возросла с 47 до 74 %, что авторы связывают с введением статистического рецензирования, внесением изменений в требования к рукописям и, возможно, с изменившейся «статистической культурой» авторов и редакторов журнала [51]. Наиболее частыми ошибками были недостаточно подробное описание проводимых статистических процедур; представление величины p без описания, с помощью какого критерия она была получена; использование параметрических методов статистики в ситуациях, когда распределение данных отличалось от нормального; использование односторонних и парных тестов в неподходящих для этого ситуациях, а также отсутствие оценки возможной мультиколлинеарности при проведении множественного регрессионного анализа [50].

Отмечается, что в большинстве журналов вероятность качественной методологической оценки всех рукописей далека от 100 %, что представляет собой серьезную проблему: только 27 % журналов, выпускаемых небольшим тиражом, имели в штате рецензента по статистике, в то время как такая должность имела в 82 % крупнотиражных (>25 000 экз.) журналов [35]. Правда, эти результаты отражают ситуацию десятилетней давности.

Практически во всех вышеупомянутых зарубежных публикациях рекомендовалось введение четких инструкций для авторов, ужесточение редакционной политики в отношении безграмотного использования статистических методов и представления результатов, обучение авторов и рецензентов основам (и не только основам) статистической обработки данных.

Опыт «Croatian Medical Journal» показывает, что введение в штат рецензента по статистике не является стопроцентной гарантией улучшения качества публикуемых работ [41]. Интересно, что если до введения в штат статистика в 1996 г. количество работ с ошибками составляло 54 %, то после введения такой должности ошибки статистического анализа регистрировались в 63 % публикаций. Редакторы журнала отмечают, что необходимы также четкая редакционная политика относительно методологической оценки всех рукописей, отслеживание повторно представленных работ и сотрудничество с профессиональными биостатистиками. Интересно, что чуть больше половины замечаний рецензентов-статистиков «Хорватского медицинского журнала» относятся к представлению авторами результатов анализа и лишь 11 % работ содержат комментарии относительно собственно анализа [41].

J. Ioannidis отмечает, что авторы часто недостаточно качественно описывают недостатки и ограничения собственных оригинальных исследований, что может

приводить к неверной их интерпретации [36]. Он приводит весомые математические аргументы в пользу того, что вообще большинство результатов публикуемых исследований ложны [37], выделяя при этом шесть особенностей, связанных с вероятностью того, что опубликованные результаты (подразумеваются значимые результаты) не являются истинными: 1) чем меньше исследование, тем выше вероятность того, что полученный результат будет ложным; 2) чем меньше эффект изучаемого воздействия на изучаемый исход, тем выше вероятность получения ложных результатов; 3) чем больше взаимосвязей проверяется в ходе исследования, тем выше вероятность получения ложных зависимостей; 4) чем больше неопределенности в определении изучаемого состояния, а также способов его изучения, тем выше вероятность получения ложного результата; 5) чем выше финансовая заинтересованность исследователей, а также чем прочнее устоявшееся мнение по поводу изучаемой проблемы, тем больше вероятность получить ложные результаты и 6) чем более актуальной и злободневной является тема исследования, тем больше вероятность получения ложного результата.

Еще одним примером того, что имеется серьезное смещение публикуемых результатов в пользу статистически значимых, может служить тот факт, что подавляющее большинство опубликованных исследований о различиях прогностических маркерах, относящихся к злокачественным новообразованиям, содержат статистически значимые результаты [39]. Такая ситуация может серьезно запутывать читателей и практикующих врачей, особенно если возникают противоречия между результатами, полученными в ходе обсервационных исследований, и результатами, полученными в ходе рандомизированных контролируемых испытаний [47].

Одним из наиболее полных и доступных обзоров статистических ошибок является переведенный на русский язык обзор Т. Ланг [14]. В нем отмечаются следующие 20 ошибок, которые может обнаружить любой, даже начинающий, исследователь: 1) представление количественных данных с излишней точностью; 2) представление непрерывных данных в виде порядковых без объяснения причин и способа преобразования; 3) представление групповых средних значений для парных данных без сообщения размера изменений внутри групп; 4) неправильное использование статистических характеристик данных; 5) использование стандартной ошибки средней величины для описательного анализа данных; 6) использование только величины r для описания различий между группами; 7) отсутствие подтверждения того, что анализируемые данные соответствуют предположениям, лежащим в основе использованных статистических методов; 8) использование линейной регрессии без подтверждения линейного характера связи; 9) включение в анализ не всех данных и не всех участников; 10) отсутствие поправок на многократность проверки гипотез; 11) ненужное сравнение исходных характеристик в рандомизированных контролируемых испытаниях; 12) отсутствие указания критериев нормы и отклонения от нормы при оценке эффективности диагностических методов; 13) отсутствие объяснения, каким образом неопределенные результаты учтены при вычислении чув-

ствительности и специфичности диагностического теста; 14) использование таблиц и рисунков для «хранения данных», а не для облегчения восприятия материала; 15) несоответствие вида графика данным, на основании которых он построен; 16) нечеткое определение понятия «объект исследования»; 17) интерпретация статистически незначимых результатов как отрицательных, а не как недостаточных; 18) непонимание различий между объяснительными и прикладными исследованиями; 19) представление результатов не в клинически важных единицах и 20) смешение понятий статистической и клинической значимости.

Таким образом, не только российские, но и многие зарубежные журналы публикуют статьи, содержащие те или иные ошибки статистического анализа, причем ошибки зачастую одинаковы, и в целом ситуация в России напоминает ситуацию за рубежом с той лишь разницей, что в крупных журналах Великобритании и США подобная картина наблюдалась несколько десятков лет назад, а в России и, например, в Бразилии она существует и сейчас.

В. В. Власов и К. Д. Данишевский [48] отмечают, что сегодня российские медицинские научные журналы находятся далеко позади зарубежных в вопросах качества рецензирования рукописей, независимости редакций и стандартов современной научной печати, причем низкое методологическое качество исследований, проводимых в России, не позволяет включать их в систематические обзоры по той или иной теме, что ставит под сомнение необходимость проведения таких исследований. Авторы делают вывод о существовании острой необходимости обучения отечественных исследователей основам научной методологии, что перекликается с рекомендациями исследователей, оценивавших качество солидных зарубежных изданий 25 лет назад [32, 34, 42].

Журнал «Экология человека» был основан в 1994 г. и с тех пор прошел путь от локального журнала с периодичностью один раз в 6 месяцев до ежемесячного, рекомендованного ВАК для публикации материалов докторских и кандидатских диссертаций, в котором публикуются авторы из различных регионов России, а также ближнего и дальнего зарубежья. Автором была предпринята попытка провести краткий анализ оригинальных статей, опубликованных в журнале в 2000–2005 гг. Всего было проанализировано 592 публикации. Оригинальные исследования составили 63 % от общего числа публикаций, причем доля оригинальных исследований увеличилась с 57 % в 2000 г. до 66 % в 2005-м. Статистические методы использовались в 60 % статей, опубликованных в 2000 г., и в 82 % статей, опубликованных в 2005-м.

Из используемых методов и выражений наиболее часто встречались выражения типа « $p < \dots$ ». Величины p были представлены в 83 % оригинальных статей, однако в 56 % случаев не указывалось, с помощью каких критериев эти величины были получены, что соответствует ремарке В. П. Леонова о «самозарождении p » [19]. Положительным моментом является то, что доля работ, в которых величина p не имела ссылки на статистические критерии, уменьшилась с 79 % (2000) до 39 % (2005). Тем не менее наличие необъясненных p

в более чем трети статей нельзя считать допустимым.

Второе место по частоте встречаемости (77 %) занимает описательная статистика, в большинстве случаев представленная в виде $M \pm m$, причем менее чем в половине статей объяснялось, что подразумевается под этими обозначениями. Лишь в единичных случаях данные представлялись в виде средних арифметических и среднеквадратических отклонений.

На третьем месте располагается критерий Стьюдента, который встречался в 30 % публикаций, однако, учитывая большое количество статей с неизвестно откуда взявшимися величинами p , можно предположить, что данный критерий использовался гораздо чаще. Несмотря на то, что критерий Стьюдента предназначен для проверки гипотезы о равенстве средних двух групп, он часто использовался для сравнения большего количества групп, причем практически всегда без применения поправок на множественные сравнения, что является ошибкой и может привести к неверным выводам. Поправка Бонферрони применялась всего в трех статьях, две из которых были опубликованы в 2005 г. Также стоит отметить, что, несмотря на декларирование использования критерия Стьюдента, почти в каждой десятой статье не удалось найти величину p , полученную в результате применения этого критерия. В некоторых случаях непарный критерий Стьюдента применялся для сравнения парных выборок, что также является ошибкой. Особо хочется обратить внимание читателей и будущих авторов на то, что для применения критерия Стьюдента следует убедиться, что соблюдаются необходимые для этого условия, некоторые из которых, например нормальность распределение и равенство дисперсий, можно проверить. Проверка распределения перед применением критерия Стьюдента проводилась лишь в 5 % работ, а проверки соблюдения условия равенства дисперсий ни в одной из работ обнаружить не удалось. Таким образом, можно заключить, что самый популярный статистический критерий применялся в подавляющем большинстве работ «наугад», без проверки необходимых условий и без поправки на множественные сравнения при наличии более чем двух групп, что позволяет усомниться в правомочности выводов таких исследований.

Корреляционный анализ встречался почти в каждой четвертой оригинальной статье (24 %). Доля статей, в которых использовался корреляционный анализ, увеличилась с 17 % (2000) до 27 % (2005). Однако только в 30 % от всех случаев применения корреляционного анализа авторы сообщали, какой коэффициент корреляции рассчитывался (15 % — Пирсона, 15 % — Спирмена), а в 70 % работ было неизвестно, с расчетом какого коэффициента проводился анализ. В 31 % работ не только не описывалось, какой коэффициент корреляции рассчитывался, но даже не упоминался корреляционный анализ при описании методов исследования, а о его применении можно было догадаться при прочтении результатов или на основании представленных выражений типа « $r = \dots$ ». Учитывая, что для использования корреляционного анализа по Пирсону также необходимо соблюдение определенных условий, результаты применения его в большинстве

случаев можно считать сомнительными с точки зрения доказательной медицины.

Критерий хи-квадрат применялся в каждой десятой статье, причем доля публикаций с его использованием возросла с 7 % (2000) до 18 % (2005). В единичных случаях, правда, встречалось описание применения этого критерия для сравнения средних, что является ошибкой.

Прочие методы встречались менее чем в 5 % публикаций. Из редко встречавшихся методов можно упомянуть линейный регрессионный анализ, дисперсионный анализ, непараметрические критерии Манна — Уитни, Вилкоксона, Краскела — Уоллиса и логистический регрессионный анализ с биномиальной переменной отклика. Причем только в 3 % статей проводилась проверка нормальности распределения. Также следует отметить применение некоторыми авторами простого регрессионного анализа для прогнозирования заболеваемости в зависимости от календарного года. Учитывая, что для таких ситуаций не соблюдается условие независимости наблюдений, в чем легко убедиться с помощью критерия Дарбина — Уотсона, не выполняется рекомендация о прогнозировании только в той области, которая ограничивается фактическими наблюдениями, и т. п., можно считать такие прогнозы как минимум неточными, что, кстати, легко проверить, посмотрев на спрогнозированные несколько лет назад для сегодняшнего дня результаты и фактические данные. Практически ни для одного из применяемых методов не проверялось соблюдение необходимых условий, что ставит под сомнение целесообразность их применения.

Ссылка на используемые статистические пакеты приводилась лишь в 35 % оригинальных статей, причем доля публикаций с их использованием увеличилась с 17 % (2000) до 48 % (2005). Из упоминаемых пакетов чаще всего встречались ссылки на различные версии электронных таблиц MS Excel (50 %). Далее следуют Statistica (45 %), SPSS (13 %), Biostat (9 %) и Statgraphics (6 %). Настораживает факт частого упоминания нескольких статистических пакетов в некоторых статьях, где анализ данных не распространялся за пределы описательной статистики.

За период с 2000 по 2005 г. был достигнут заметный прогресс в описании методов обработки данных. Если в 2000 г. не встретилось ни одной статьи, где имелись бы как ссылки на используемые методы статистики, так и аргументация их применения, проверка распределения и ссылка на пакет статистических программ, то к 2005 г. такие статьи начали появляться чаще. Все статьи можно условно разделить по качеству описания проводимых статистических процедур на 3 группы: 1 — ссылок на применение методов статистического анализа нет вообще; 2 — ссылки на методы описания или обработки есть, но этого описания явно недостаточно для понимания того, что же на самом деле анализировалось, и 3 — есть попытки обосновать применение тех или иных методов статистики или представить более подробное описание методов, чем в предыдущей группе. Невзирая на всю условность такого деления доля статей, отнесенных к последней группе, увеличилась с 0 (2000) до 14 %

(2005). Общее число статей вообще без описания производимых статистических процедур уменьшилось с 81 % (2000) до 40 % (2005). Несмотря на значительный рост количества публикаций с более подробным описанием методов обработки данных, наличие 40 % оригинальных статей вообще без описания производимых статистических процедур в 2005 г. и еще 26 % статей с описанием методов обработки данных в 2–3 строчки, отнесенных к группе 2, говорит о неудовлетворительном описании методологии проведения исследования авторами и неудовлетворительном контроле этого описания рецензентами.

Попробуем оценить одну из статей с наиболее подробным описанием методов обработки данных: *«Статистическая обработка полученных результатов, оценка распределения показателей, определение границ нормального распределения проводилась с помощью пакета статистических программ Statistica (StatSoft, USA). Критический уровень значимости (p) при проверке статистических гипотез принимался за 0,05. Оценка соотношений генеральных дисперсий в большинстве сравниваемых выборок показала их неравенство. Также в большинстве выборок выявлено неправильное распределение. В связи с этим статистическая обработка данных проводилась непараметрическими методами. Учитывались средние значения (M), стандартные отклонения (SD), минимальные и максимальные значения в выборках (Min-Max). Для проверки статистической гипотезы разности средних значений использовался ANOVA, метод множественных сравнений, критерий наименьшей значимости».*

На первый взгляд такое описание (одно из самых детальных из всех проанализированных статей) может показаться образцовым, так как автор проверяет тип распределения данных, условие равенства дисперсий, определяет принятый в данном исследовании критический уровень значимости, называет применяемые описательные статистики, обозначает используемые сокращения, а также аргументирует применение выбранных статистических процедур и приводит название статистического пакета с указанием производителя. На основании подобного описания читателю легко формировать собственное суждение о качестве выполненного исследования и при необходимости воспроизвести анализ при наличии собственных данных. Отдавая должное автору за достаточно подробное описание методов обработки данных, следует, однако, отметить несколько ошибок как в описании, так и в выборе методов представления и обработки данных:

1. Неясно, что имеется в виду под определением границ нормального распределения, каким образом эти границы определялись и вообще зачем это делалось, так как если распределение признака подчиняется закону нормального распределения, то возможны значения от $-\infty$ до $+\infty$.

2. Неизвестно, каким образом проверялось распределение данных.

3. Автор отмечает «неправильное» распределение данных. Такого распределения не существует, равно как не существует правильного распределения.

4. Несмотря на то, что распределение данных не подчинялось закону нормального распределения, автор

описывает данные с помощью средних арифметических и стандартных отклонений, что не рекомендуется делать в подобных ситуациях.

5. Автор ссылается на использование непараметрических методов обработки данных, однако результатов применения этих методов, равно как и указания, какие конкретно непараметрические методы применялись, обнаружить не удалось, а увиделось применение дисперсионного анализа, который является параметрическим методом.

6. Несмотря на сообщение автора о том, что данные не подчиняются закону нормального распределения, а выборочные дисперсии различаются между группами, все равно используется дисперсионный анализ, хотя оба важнейших условия применения этого метода, проверенные автором, не соблюдаются. То есть автор сознательно идет на применение ошибочного способа обработки данных!

7. В качестве апостериорных критериев автор применяет критерий наименьшей значимости. Несколько десятилетий назад было показано, что этот критерий недостаточно контролирует ошибку первого типа при множественных сравнениях, то есть вероятность получения достоверных различий там, где их на самом деле нет, очень велика. Предполагается, что именно критерий наименьшей значимости используется в качестве декларируемого автором «метода множественных сравнений».

Таким образом, несмотря на заслуживающую всяческих похвал подробность описания методов обработки данных, позволяющую читателю судить о качестве исследования, были применены некорректные методы как описания, так и обработки данных, что вполне могло привести к ошибочным выводам исследования.

Следует также признать, что довольно часто встречались формулировки, лишённые всякого смысла. Ниже приводятся лишь наиболее типичные и особенно запомнившиеся, а также случаи, где описание методов расходится с представленными результатами.

1. *«Весь материал обработан методами вариационной статистики с использованием пакета прикладных программ Statistica 5.0 на PC-IBM PII-400».*

Подобные формулировки встречались довольно часто, что вызывает большую озабоченность, так как смысла в них нет. Мало того, что без вариации нет статистики и что от термина «вариационная статистика» было рекомендовано отказаться еще в 1973 г., отсутствие ссылки на конкретные статистические критерии не позволяет читателю оценить адекватность их применения. Ссылка на использование компьютера, вероятно, дается для придания большей надежности полученным данным, однако абсолютно неважно, на каком компьютере произведен анализ, если критерий выбран неправильно.

2. *«При анализе данных использованы общепринятые статистические методы и методы наглядного представления средних значений «паутинка» (А. А. Иванов)».*

Какие методы являются для авторов общепринятыми, непонятно, хотя в результатах встретились упоминание о дисперсионном анализе и достигнутый уровень значимости. «Наглядный метод представления данных» так

и не был обнаружен, так как единственная таблица, приводимая в статье, включала в себя существующие рекомендации по физической активности для больных и не содержала оригинальных данных. Ни упомянутой выше «паутинки», ни ссылки на рекомендации А. А. Иванова в списке литературы обнаружено не было. Упоминание «общепринятых» методов встречается часто, но не помогает читателю понять, как все-таки обрабатывались данные. Поэтому такой формулировки следует избегать.

3. *«Результаты обработаны в программе «Statistica» с использованием параметрических и непараметрических статистических методов».*

Хорошо, что авторы имеют представление о параметрических и непараметрических методах, однако и тех и других методов существует большое количество. Подобных размытых и ничего не говорящих формулировок также следует избегать. Необходимо четко указывать, с помощью каких критериев проверялись те или иные статистические гипотезы.

4. *«Результаты исследований обрабатывались методом вариационной статистики с использованием дискретивного, корреляционного и факторного вида анализа...».*

В работе не удалось обнаружить результатов ни корреляционного, ни факторного анализов, а только данные в виде $M \pm m$ для двух групп и достигнутые уровни значимости, что позволяет заподозрить применение критерия Стьюдента, который не упоминался при описании методов. Кроме того, «дискретивного» анализа не существует. Есть методы описательной статистики или дескриптивные методики (от англ. descriptive). Заслуживает внимания и существование, по мнению авторов, какого-то одного «метода вариационной статистики». Таким образом, в данной статье мы имеем туманное описание проводимого статистического анализа и, что особенно пугает, несоответствие фактически применяемых методов обработки данных описанию в разделе «Материал и методы». Последнее также не является единичным примером, причем эта досточно новая проблема стала встречаться все чаще и чаще.

5. *«Обработка полученных данных проводилась на персональной ЭВМ с процессором 80486/487 Intel Inside с использованием программ статистической обработки Excel 6.0 for Windows, SAS (Statistic Analytical System). Применялись стандартные статистические и математические функции автоматизированных программ».*

Данная формулировка по праву может считаться наиболее изощренным по своей неинформативности клише, так как не только не несет никакой полезной информации, но и путем ссылок на использование ЭВМ с представлением технической информации, серьезный пакет программ (в расшифровке аббревиатуры которого, однако, делается две ошибки) и «математические функции автоматизированных программ» создает у неподготовленного читателя ощущение глубокой научности исследования.

6. *«Рассчитывались среднее, коэффициент вариации, асимметрия, эксцесс, коэффициент относительной энтропии информации».*

Безусловно похвально использование исследователями большого количества описательных статистик, однако, несмотря на значение коэффициента асимметрии 2,3, что свидетельствует о сильно скошенном вправо распределении (существенное отклонение от нормального), авторы все равно применяли критерий Стьюдента для сравнения выборочных средних, что в данной ситуации некорректно. В этой же работе критерий Стьюдента применялся для сравнения разных по объему групп на трех временных отрезках, что опять же некорректно. Таким образом, складывается впечатление, что расчет большого количества описательных статистик, включая асимметрию и эксцесс, производился просто для того, чтобы ошеломить читателя большим количеством расчетов, так как результаты этих расчетов не предотвратили использование неподходящих методов обработки данных.

7. *«Статистическая обработка данных проводилась на персональном компьютере с вычислением средних значений выборок, дисперсии, коэффициента корреляции. Достоверность параметров оценивали при помощи t-критерия Стьюдента».*

В данном случае, помимо описанных в предыдущих примерах проблем, появляется непонимание авторами семантики слов «достоверность» и «параметры», равно как и того, что «достоверность параметров» с помощью критерия Стьюдента проверить нельзя, так как он предназначен для проверки статистической гипотезы о равенстве средних при наличии двух независимых выборок (если нет указания на применение парного критерия).

Подобных примеров можно привести много, но были выбраны лишь те, которые в той или иной форме наиболее часто встречались в анализируемых статьях. При значительном увеличении объема описания методов обработки данных в статьях, опубликованных в 2005 г., качество большинства таких описаний все еще оставляет желать лучшего. Далеко не всегда подробно описывается выборка, а также процедура отбора участников исследования и тип исследования.

Помимо недостаточного описания применяемых методов сбора и обработки данных вызывает озабоченность представление результатов и особенно их интерпретация. Наиболее бросающейся в глаза проблемой является ограниченное использование авторами большинства статей таблиц и рисунков. Во многих статьях результаты представляются лишь в текстовом формате, что не дает оптимальной картины происходящего. Что касается интерпретации результатов, то наиболее часто встречающейся ошибкой является заключение о наличии «достоверных», а то и причинно-следственных связей между изучаемыми признаками. В подавляющем большинстве исследований на основании имеющихся данных такие выводы сделать невозможно. Лишь в единичных случаях авторы обсуждают достоинства и недостатки исследований, а также предлагают, какие еще исследования необходимо провести для получения ответа на поставленные вопросы. Следует отметить, что раздел «Обсуждение результатов» является наиболее важным для читателей при оценке результатов и выводов исследования, так как именно в нем обсуждается

вероятность наличия систематических ошибок и вмешивающихся факторов, а также сравнение настоящего исследования с подобными исследованиями, проведенными другими авторами. Такая оценка встречалась, к сожалению, лишь в единичных случаях.

Таким образом, можно констатировать, что качество выполнения и представления результатов статистического анализа в журнале «Экология человека» не превышает таковое в центральных журналах, описанное В. П. Леоновым [16–24], и вполне соответствует тому, что наблюдалось в зарубежных журналах пару десятилетий назад. Однако после того, как подробный отчет о качестве статей, опубликованных в «Экологии человека» в 2000–2005 гг., был представлен в редакцию в начале 2007 г., в журнале произошли значительные перемены в соответствии с рекомендациями как зарубежных [32, 34, 42], так и российских [1, 3, 13] авторов. Были введены новые единые требования к рукописям, представляемым в журнал, соответствующие рекомендациям Международного комитета редакторов медицинских журналов; была разработана новая система рецензирования рукописей, которая теперь включает не только рецензирование по специальности, но и оценку методологического качества исследования; в 2008 г. в структуру журнала была введена рубрика «Практикум», в которой читатели и будущие авторы знакомятся с некоторыми простейшими методами обработки данных; в Северном государственном медицинском университете, с которым так или иначе сотрудничает значительная доля авторов журнала, регулярно проводятся курсы как по основам статистики, так и по многомерным методам обработки данных, а также по эпидемиологии и планированию исследований. Приведут ли эти меры к улучшению качества публикаций, покажет сравнительный наукометрический анализ статей, опубликованных до и после введения вышеперечисленных изменений.

Список литературы

1. Барт А. Г. О состоянии дел и перспективах обучения статистическому анализу медицинских данных / А. Г. Барт, Е. В. Вербицкая, В. Н. Солнцев // Международный журнал медицинской практики. — 2006. — № 2. — С. 39–42.
2. Бащинский С. Е. Статистика умеет много гитик / С. Е. Бащинский // Международный журнал медицинской практики. — 1998. — № 4. — С. 13–15.
3. Вараксин А. Н. Статистический анализ биологической и медицинской информации: проблемы и решения / А. Н. Вараксин // Международный журнал медицинской практики. — 2006. — № 2. — С. 35–38.
4. Власов В. В. Эпидемиология в современной России / В. В. Власов // Международный журнал медицинской практики. — 2001. — № 2. — С. 27–31.
5. Воробьев К. П. Доказательная медицина — новая методология медицинской практики. Ч. I. Мотивации врача и исследователя при изучении доказательной медицины / К. П. Воробьев // Украинский медицинский альманах. — 2004. — № 5. — С. 41–45.
6. Воробьев К. П. Доказательная медицина — новая методология медицинской практики. Ч. II. Сущность доказательной медицины / К. П. Воробьев // Украинский медицинский альманах. — 2005. — № 6. — С. 142–146.
7. Воробьев К. П. Научные исследования и доказательная медицина: взаимоотношения и место в процессе получения

новых знаний / К. П. Воробьев // Біль, знеболювання і інтенсивна терапія. — 2005. — № 2. — С. 2–5.

8. Воробьев К. П. Проблемные вопросы представления материалов клинических исследований / К. П. Воробьев // Біль, знеболювання і інтенсивна терапія. — 2005. — № 2. — С. 71–80.

9. Гржибовский А. М. Корреляционный анализ в медицинских исследованиях / А. М. Гржибовский // Бюллетень СГМУ. — 2000. — № 2. — С. 22–23.

10. Гржибовский А. М. Применение статистики в терапии: критический анализ публикаций / А. М. Гржибовский // Бюллетень СГМУ. — 2000. — № 2. — С. 21–22.

11. Дюк В. А. Преподавание прикладной статистики в Санкт-Петербургской медицинской академии последипломного образования / В. А. Дюк // Международный журнал медицинской практики. — 2006. — № 2. — С. 30–31.

12. Зорин Н. А. О неправильном употреблении термина «достоверность» в российских научных психиатрических и общемедицинских статьях / Н. А. Зорин // Биометрика: журнал для медиков и биологов, сторонников доказательной медицины www.biometrica.tomsk.ru

13. Комарова М. В. Статистический анализ биомедицинских данных: проблемы и пути решения / М. В. Комарова // Международный журнал медицинской практики. — 2006. — № 2. — С. 28–29.

14. Ланг Т. Двадцать ошибок статистического анализа, которые вы сами можете обнаружить в биомедицинских статьях / Т. Ланг // Международный журнал медицинской практики. — 2005. — № 1. — С. 21–31.

15. Левандовский А. Б. Доказательная медицина молодому ученому / А. Б. Левандовский // Международный журнал медицинской практики. — 2006. — № 2. — С. 23–24.

16. Леонов В. П. Долгое прощание с лысенковщиной / В. П. Леонов // Биометрика: журнал для медиков и биологов, сторонников доказательной медицины www.biometrica.tomsk.ru

17. Леонов В. П. Об использовании прикладной статистики при подготовке диссертационных работ по медицинским и биологическим специальностям / В. П. Леонов, П. В. Ижевский // Бюллетень ВАК РФ. — 1997. — № 5. — С. 56–61.

18. Леонов В. П. Обучение медиков статистике: попытка системного подхода к проблеме / В. П. Леонов // Международный журнал медицинской практики. — 2006. — № 2. — С. 17–22.

19. Леонов В. П. Ошибки статистического анализа биомедицинских данных. / В. П. Леонов // Международный журнал медицинской практики. — 2007. — № 2. — С. 19–35.

20. Леонов В. П. Применение методов статистики в кардиологии (по материалам журнала «Кардиология» за 1993–1995 гг.) / В. П. Леонов // Кардиология. — 1998. — № 1. — С. 55–58.

21. Леонов В. П. Применение статистики в статьях и диссертациях по медицине и биологии. Ч. 1. Описание методов статистического анализа в статьях и диссертациях / В. П. Леонов, П. В. Ижевский // Международный журнал медицинской практики. — 1998. — № 4. — С. 7–12.

22. Леонов В. П. Применение статистики в статьях и диссертациях по медицине и биологии. Ч. 2. История биометрики и ее применения в России / В. П. Леонов // Международный журнал медицинской практики. — 1999. — № 4. — С. 7–19.

23. Леонов В. П. Применение статистики в статьях и диссертациях по медицине и биологии. Ч. 3. Проблемы взаимодействия автор — редакция — читатель /

В. П. Леонов // Международный журнал медицинской практики. — 1999. — № 12. — С. 7–13.

24. Леонов В. П. Применение статистики в статьях и диссертациях по медицине и биологии. Ч. 4. Наукометрия статистической парадигмы экспериментальной биомедицины / В. П. Леонов // Международный журнал медицинской практики. — 2002. — № 3. — С. 6–10.

25. Орлов А. И. О применении статистических методов в медико-биологических исследованиях / А. И. Орлов // Вестник Академии медицинских наук СССР. — 1987. — № 2. — С. 88–94.

26. Плавинский С. Л. О людях и цифрах. Обучение статистике: чему, кого и как учить? / С. Л. Плавинский // Международный журнал медицинской практики. — 2006. — № 2. — С. 9–16.

27. Фадеев В. В. Доказательная медицина и отечественная медицинская наука / В. В. Фадеев, В. П. Леонов, О. Ю. Реброва, Г. Л. Мельниченко // Проблемы эндокринологии. — 2003. — № 5. — С. 55–60.

28. Altman D. // Statistics in Medicine. — 1991. — Vol. 10. — P. 1897–1913.

29. Altman D. G. // Statistics in Medicine. — 1983. — Vol. 1. — P. 59–71.

30. Altman D. G. // Statistics in Medicine. — 2000. — Vol. 10. — P. 1897–1913.

31. Crosta F., Nishiwaki-Dantas M., Silvino W., P. Dantas // Arquivos Brasileiros de Oftalmologia. — 2005. — Vol. 68. — P. 339–345.

32. Cruess D. // The American Journal of Tropical Medicine and Hygiene. — 1989. — Vol. 41. — P. 619–626.

33. Elster A. D. // American Journal of Roentgenology. — 1994. — Vol. 163. — P. 711–715.

34. Emerson J. D., Colditz G. A. // New England Journal of Medicine. — 1983. — Vol. 309. — P. 709–713.

35. Goodman S. N., Altman D., George S. L. // Journal of General Internal Medicine. — 1998. — Vol. 13. — P. 753–756.

36. Ioannidis J. // Journal of Clinical Epidemiology. — 2007. — Vol. 60. — P. 324–329.

37. Ioannidis J. // PLoS Medicine. — 2005. — Vol. 2. — P. 696–701.

38. Juzych M., Shih D., Seyedsadr M. et al. // Archives of Ophthalmology. — 1992. — Vol. 110. — P. 1225–1229.

39. Kyzas P., Denaxa-Kyza D., Ioannidis J. // European Journal of Cancer. — 2007. — Vol. 43. — P. 2559–2579.

40. Lee C. M., Soin H. K., Einarson T. R. // Annals of Pharmacotherapy. — 2004. — Vol. 38. — P. 1412–1418.

41. Lukic, I. K., Marusic M. // Croatian Medical Journal. — 2001. — Vol. 42. — P. 500–503.

42. MacArthur R. D., Jackson G. G. // Journal of Infectious Diseases. — 1984. — Vol. 149. — P. 349–354.

43. Mora I., Ripoll R., Canela I., Sorel J. // Revista Clinica Espanola. — 1997. — Vol. 197. — P. 23–27.

44. Olsen C. // Infection and Immunity. — 2003. — Vol. 71. — P. 6689–6692.

45. Pitkin R. M. // Obstetrics and Gynaecology. — 1994. — Vol. 83. — P. 1043–1044.

46. Pitkin R. M., Burmeister L. F. // Obstetrics and Gynaecology. — 1995. — Vol. 86. — P. 124–125.

47. Tatsioni A., Bonitsis N., Ioannidis J. // Journal of the American Medical Association. — 2007. — Vol. 298. — P. 2517–2526.

48. Vlassov V. V., Danishevskiy K. D. // Emerging Themes in Epidemiology. — 2008. — Vol. 5. — P. 15.

49. Wang Q., Zhang B. // Journal of the American Medical Association. — 1998. — Vol. 280. — P. 283–285.

50. Welch G. E., Gabbe S. G. // American Journal of Obstetrics and Gynecology. — 1996. — Vol. 175. — P. 1138–1141.

51. Welch G. E., Gabbe S. G. // American Journal of Obstetrics and Gynecology. — 2002. — Vol. 186. — P. 584–586.

52. White S. J. // British Journal of Psychiatry. — 1979. — Vol. 135. — P. 336–342.

53. Wulff H., Andresen B., Brandenhoff P., Guttler F. // Statistics in Medicine. — 1987. — Vol. 6 — P. 3–10.

USE AND MISUSE OF STATISTICS IN RUSSIAN BIOMEDICAL LITERATURE

A. M. Grjibovski

National Institute of Public Health, Oslo, Norway
Arkhangelsk International School of Public Health,
Arkhangelsk, Russia

The paper presents a non-scientific and non-systematic review of literature on the most common errors associated with the use statistical methods in Russian and international biomedical literature as well as errors in interpretation of the results. Moreover a short analysis of statistical methods used by the authors of publications in the Human Ecology journal is presented. Simple study designs and basic statistical tests, failure to correct for multiple comparisons, failure to adjust for confounders, uncritical interpretation of the results and overstepping the data are among the other most common features of the Russian biomedical literature, which makes most of the findings (with some exceptions) questionable from the evidence-based medicine point of view. The author is mostly concerned about the recent increase in disparities between methods description and results, which may confuse both reviewers and readers and result in the overestimation of the scientific merits of these papers. Examples of the most common errors are given with suggestion for improvements.

Key words: type of data, statistical tests, distribution.

Контактная информация:

Гржибовский Андрей Мечиславович — доктор медицины, магистр международного общественного здоровья, старший советник Национального института общественного здоровья, г. Осло, Норвегия; директор Архангельской международной школы общественного здоровья при Северном государственном медицинском университете

Адрес: Nasjonalt folkehelseinstitutt, Pb 4404 Nydalen, 0403 Oslo, Norway

Тел.: +47 21076392, +47 45268913; e-mail: angr@fhi.no

Статья поступила 01.12.2008 г.