

М. И. Мусатов

НИИ клинической иммунологии СО РАМН
ул. Ядринцевская, 14, Новосибирск, 630099, Россия
Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия
E-mail: lugamus@yandex.ru

СТРУКТУРА ОПИСАНИЯ МЕТОДОВ СТАТИСТИЧЕСКОГО АНАЛИЗА В ПУБЛИКАЦИЯХ ИММУНОЛОГИЧЕСКИХ ЖУРНАЛОВ*

Проведено ретроспективное наукометрическое исследование 3 169 оригинальных статей в отечественных и зарубежных журналах за периоды 1980–1991, 1992–1999 и 2000–2007 гг. Анализ качества описаний статистических методов в российских публикациях показал, что за счет их отсутствия, несоответствия фактически проведенной статобработке и, в большей мере, смысловой неопределенности уровень доказательности публикаций существенно снижается. Лишаются доказательности 3–4 % публикаций, переходя в категорию статей, не использующих статистическую обработку. Только описательные статистики использовались в 636 публикациях, при этом в 40–75 % случаев применялись неинтерпретируемые выражения. В 874 статьях использовали только выражения типа « $p < \dots$ », из них 208 за счет неопределенности описаний. В результате доля тех отечественных иммунологических публикаций, где использовались методы статистического анализа, на 56–59 % состоит из статей, использующих только описательные статистики и выражения типа « $p < \dots$ ». Это превышает частоту аналогичных недостатков в международных иммунологических журналах в 22–33 раза. Более того, такая структура отечественных работ является постоянной и не менялась с 1980 по 2007 г.

Ключевые слова: биостатистика, статистика в медицине, иммунология.

Обоснованность требований редакций многих научных журналов биомедицинского и тем более медицинского, клинического профиля к обязательной статистической обработке результатов исследования очевидна. Ошибочно оцененные результаты на практике могут привести к повышению риска ущерба здоровью и даже жизни пациентов. Тем не менее часто описание статистических методов исчерпывается фразами типа «полученные результаты обрабатывались статистически», хотя по смыслу это тоже самое, как если бы методы обследования больных сократить до сообщения, что пациенты «обследовались терапевтически». Некорректность в описании методов статистического анализа в отечественных биомедицинских публикациях является воспроизводимым явлением при наукометрических исследованиях. Несомненно, это нечто большее, чем просто погрешность оформ-

ления. Описание статистических методов в научной статье – существенная часть всего статистического анализа. Однако вопрос о том, как измерить последствия ошибочных описаний, представляется недостаточно изученным.

Материал и методы

Дизайн исследования – наукометрическое (библиометрическое) ретроспективное. Анализировались отечественные и зарубежные публикации в области иммунологии. Рассматривались только оригинальные работы (в том числе и краткие сообщения), структурированные формально или по контексту на разделы, обязательно содержащие описание методов исследования, полученных результатов и их обсуждение.

Из анализа исключались обзорные, редакционные, дискуссионные и другие аналогичные публикации, а также тезисы кон-

* Работа выполнена при финансовой поддержке РГНФ в рамках научно-исследовательского проекта «Наукометрический анализ статистических парадигм отечественных и зарубежных иммунологических исследований» № 08-03-006117а.

ференций и симпозиумов. Анализ включал публикации за 27 лет в отечественных журналах «Иммунология» (ИМ), «Russian Journal of Immunology» (РJI), «International Journal of Immunorehabilitation» (последние два с 1996 г.) и зарубежных изданиях «Journal of Immunology», «Clinical and Vaccine Immunology», «Clinical and Diagnostic Laboratory Immunology», «International Immunology» (последний с 1999 г.). Для анализа использовались как поисковые системы сайтов журналов и анализ электронных вариантов публикаций, так и поиск в «ручном режиме» по всем отечественным журналам и «Journal of Immunology» за период с 1980 по 2002 г.

Проанализировано 3 272 публикации в отечественных и 64 368 – в зарубежных журналах (рис.). В основной анализ включено 3 169 публикаций, 103 статьи из «International Journal of Immunorehabilitation» (ИИ) рассмотрены отдельно. Основные данные анализа были условно разделены на 3 периода: 1980–1991 гг. (первый) – год начала издания ИМ – год прекращения существования СССР, два периода по 7 лет до 2007 г. включительно (в соответствии с рамками настоящего проекта).

Для целей количественного анализа использовалась следующая типология описаний статистических методов:

1) в публикации статистический анализ не используется;

2) используются только описательные статистики, в том числе и явно не обозначенные (выражения типа « $X \pm x$ » или «усы» на рисунках);

3) для оценки гипотез используются конкретные статистические методы;

4) информации об использованных статистических критериях не содержится ни в разделе «Материалы и методы», ни в пояснениях к таблицам и рисункам, ни в тексте статьи, однако авторами используются выражения типа « $p < \dots$ ». Информация может отсутствовать просто потому, что нет описания, либо из-за бессодержательности последнего.

Проверка гипотезы о равенстве относительных частот двух показателей проводилась по исходным абсолютным значениям с помощью критерия χ^2 с поправкой Йетса.

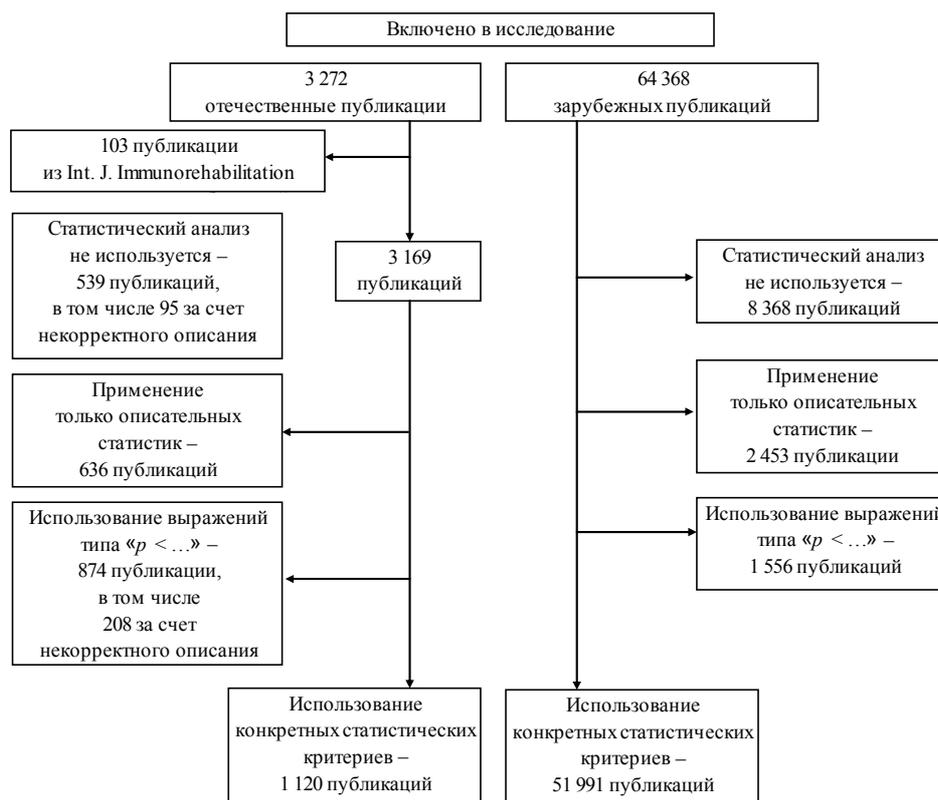
При анализе динамики показателей за три периода использовался метод χ^2 для тенденции. Отношение относительных частот выражалось как отношение шансов (ОШ) с 95 % доверительным интервалом (ДИ). Относительные частоты представлены со стандартной ошибкой. Для большей наглядности в целях обсуждения полученных данных часть результатов приводится также в процентах. Расчеты выполнены с помощью статистического калькулятора Statcalc из пакета программ EpiInfo версии 3.5.1 (2008).

Результаты исследования и обсуждение

Форма описаний статистических методов отечественными авторами заслуживает самостоятельного исследования. Это отчасти было сделано в более ранних исследованиях [1], но разработанная его авторами классификация является скорее классификацией ошибочных типов описаний. Используемая в настоящей работе рубрификация, несмотря на нетривиальность задачи, основывается на анализе содержания и более прагматична. Приведем несколько примеров.

«Полученные значения клинико-иммунологических параметров до и после комплексного лечения обработаны методом вариационной статистики с определением среднеарифметической величины (M) и ее ошибки (m) в целом для группы и индивидуально для каждого конкретного больного» (ИМ. 2004. Т. 25, № 6. С. 371). Эта работа относится к категории «использование только описательных статистик».

«Подсчет среднего количества бактерий $M \pm m$ проводили по методу Стьюдента» (ИМ. 1983. № 2. С. 56). В работе не обнаруживаются никаких признаков использования данного критерия, поэтому это тоже работа из группы 2. К ней же относится цитируемая далее работа, поскольку в ней кроме приводимого описания присутствуют только «усы» на рисунках (трех градаций длины). «Доверительный интервал средних определяли для 95 % доверительного интервала. Достоверность различий высчитывали по критерию Стьюдента» (ИМ. 1983. № 2. С. 43).



Дизайн исследования

Следующая работа относится к категории 3: «Достоверность различий сравниваемых показателей находили по критерию Стьюдента. Учитывая, что показатели контрольных и подопытных животных исследовали в стандартных условиях (животных в обеих группах одного возраста и пола исследовали одновременно), при оценке статистической значимости данных пользовались порогом вероятности безошибочных прогнозов, употребляемых при обычной ответственности, равной 0,95» (ИМ. 1989. № 1. С. 32).

К этой же категории относится и работа, использовавшая «criterion Maik – Wittny – Wemnacson» (RJ. 1998. Vol. 3, № 2. P. 167). В данном случае легко узнается несколько устаревшее название критерия Манна – Уитни – Вилкоксона – «критерий Манна – Уитни».

Следующий пример относится к категории 4, поскольку в тексте присутствуют только выражения « $p < \dots$ ». Появление в тексте таких выражений без каких-либо прелиминариев В. П. Леонов [2] назвал «самозарождение p »

«Для статистической обработки материала применен пакет прикладных программ Microsoft Office (номер продукта 54521-750-6140064-17912)» (IJ. 2003. Т. 5, № 1. С. 28).

Несмотря на явное указание статистического критерия, следующая работа относится к категории публикаций 1, где статистический анализ не используется, поскольку в тексте отсутствуют не только значения t и уровень доверительной вероятности, но и выражения типа « $X \pm x$ » или их графический аналог.

«Статистический анализ проводился по методу t -критерия Стьюдента при сравнении снижения МСВ на 15 % после ФВ» (IJ. 1998. № 10. С. 168).

В среднем статистические методы использовались в 83 % отечественных и 88 % зарубежных публикациях, что согласуется с данными сравнительного анализа биомедицинских публикаций за 1987–1997 гг., где выявлено соответственно 82 и 87 % описаний [3]. Отсутствие в публикации статисти-

ческого анализа – нормальное явление для работ, чьи результаты носят качественный характер. Однако доля публикаций, не использующих статистический анализ, в отечественной выборке несколько завышена из-за вынужденного отнесения к ним части статей. Среди отечественных публикаций за весь период анализа относительная частота таких работ с формально присутствующим описанием статистических методов составила $0,0361 \pm 0,0008$, или 3,61 %. В абсолютных значениях (см. рис.) – 95 работ. Таким образом, доля публикаций, использующих статистический анализ, могла бы составить 86,6 %, как в зарубежных работах.

Встречаемость некорректных описаний использования статистических методов в анализируемых англоязычных публикациях оказалась весьма редким явлением. Так, за период 1980–2002 гг. в J. Immunol. при ручном анализе были обнаружены только 3 публикации, сообщавшие, что обработка данных проводилась с помощью статистических программ. Учитывая, что относительная частота таких работ составила $0,0002 \pm 0,00013$, в дальнейшем они были исключены из анализа.

Как известно, статистический анализ подразумевает раздельное или сочетанное применение двух основных приемов – использование описательных статистик и проверку гипотез с помощью статистических критериев. Последнее в основном выходит за рамки настоящей публикации, но мы рассмотрим такую своеобразную часть работ, которые содержат «самозародившееся p », т. е. подразумевают проведение анализа с помощью неких статистических критериев.

Следует отметить, что в результате некорректности описания статистических методов основная масса таких публикаций выросла почти на четверть. Среди статей, где описание статистических методов формально присутствовало, доля таких работ (с неидентифицируемыми критериями) составила $0,238 \pm 0,0321$, т. е. 208 публикаций (см. рис.).

Анализ структуры используемых описательных статистик показал, что существенную долю в них составляют выражения типа « $X \pm x$ » или «усы» на рисунках и графиках. По выделенным периодам относительная частота подобного «показателя»

составила $0,406 \pm 0,042$, $0,753 \pm 0,039$ и $0,423 \pm 0,072$ соответственно. Доля же обозначенных описательных статистик по мере убывания выстраивается в следующий ряд: ошибка средней арифметической (примерно $2/3$), среднеквадратическое отклонение и доверительный интервал. Даже если вычислить на основе имеющихся данных доверительные интервалы вместо ошибок средних, то при малых выборках и неизвестном соответствии их нормальному распределению доказательность исследований, результаты которых анализировались таким образом, представляется невысокой.

Примечательно, что после некоторого (и статистически значимого) повышения использования неидентифицируемых описательных статистик в 90-е гг., частота таких показателей в 2000–2007 гг. вернулась к уровню 1980–1991 гг.

Нами представлены данные по относительным частотам использования в публикациях только описательных статистик (табл.). Анализ отечественных данных с помощью критерия χ^2 для тенденции по отдельности для частот использования описательных статистик и для частот применения выражений « $p < \dots$ » показал в период 1992–1999 гг. статистически значимое повышение частоты последних и также значимое снижение частоты использования описательных статистик (причем за счет обозначенных), «нейтрализующее» повышение показателя « $p < \dots$ ». Суммарная частота использования в отечественных публикациях показателей « $p < \dots$ » и «описательные статистики» не обнаруживает наличия тренда за три периода ($\chi^2 = 0$, $p = 0,997$).

Для суммы этих показателей для зарубежных публикаций по трем периодам по критерию χ^2 для тенденции выявлено наличие линейного тренда ($\chi^2 = 227,74$; $p < 0,0001$) за счет показателя для периода 2000–2007 гг. Такой же тренд выявляется и для анализа каждого из них для зарубежных публикаций.

Таким образом, в зарубежных иммунологических публикациях в течение всего периода анализа наблюдается небольшое, но статистически значимое снижение использования в работах только описательных статистик (ОШ = 1,38, 95 % ДИ = 1,25–1,53)

Относительные частоты использования описательных статистик и выражений типа « $p \leq \dots$ » в отечественных и зарубежных иммунологических публикациях

Показатель	Годы публикаций		
	1980–1991	1992–1999	2000–2007
Отечественные публикации			
Только описательные методы статистики	0,254 ± 0,023	0,1395 ± 0,0360	0,240 ± 0,036
Выражения типа « $p \leq \dots$ »	0,294 ± 0,022	0,3472 ± 0,0310	0,3033 ± 0,034
Всего	0,548 ± 0,018	0,487 ± 0,028	0,543 ± 0,054
Зарубежные публикации			
Только описательные методы статистики	0,0489 ± 0,0074	0,062 ± 0,008	0,0361 ± 0,0070*
Выражения типа « $p \leq \dots$ »	0,0313 ± 0,0074	0,0403 ± 0,0080	0,0210 ± 0,0072*
Всего	0,080 ± 0,002	0,1000 ± 0,0025	0,0574 ± 0,0020*

Примечание: * – статистически значимое снижение частоты, обуславливающее наличие линейного тренда по критерию χ^2 для тенденции.

и выражений типа « $p < \dots$ » (ОШ = 1,52, 95 % ДИ = 1,33–1,73).

Сравнение отечественных и зарубежных публикаций по частоте использования описательных статистик и выражений типа « $p < \dots$ » показало, что в период с 1980 по 1991 г. в отечественных работах эти показатели использовались в 21–28 раз чаще (ОШ = 24,62, 95 % ДИ = 21,45–28,26; $\chi^2 = 3790,57$; $p < 0,0001$) и в период 2000–2007 гг. еще чаще – в 22–33 раза (ОШ = 27,33, 95 % ДИ = 22,61–33,04; $\chi^2 = 2482$; $p < 0,0001$), за счет снижения этих показателей в иностранных публикациях.

Заключение

Проведенный анализ качества описаний статистических методов в российских иммунологических публикациях показал, что за счет их отсутствия, несоответствия фактически проведенной статобработке и, в большей мере, смысловой неопределенности уровень доказательности публикаций существенно снижается:

- 3–4 % публикаций лишаются доказательности, переходя в категорию статей, не использующих статистическую обработку;

- невысокая доказательность описательных статистик, фактически необходимых для описания, а не для оценок альтернативных групп, снижается еще больше, на

40–75 %, подменяясь неопределенными показателями;

- безусловно некорректная категория публикаций, опирающаяся в качестве основных аргументов на выражения типа « $p < \dots$ », на 20–27 % формируется за счет дефектности описаний.

В результате доля тех отечественных иммунологических публикаций, где использовались методы статистического анализа, на 56–59 % состоит из статей, использующих только описательные статистики и выражения типа « $p < \dots$ ». Это превышает частоту аналогичных недостатков в международных иммунологических журналах в 22–33 раза. Более того, такая структура отечественных работ является постоянной и не менялась в течение 27 лет, с 1980 по 2007 г.

Список литературы

1. *Леонов В. П., Ижевский П. В.* Применение статистики в статьях и диссертациях по медицине и биологии. Часть 1: Описание методов статистического анализа в статьях и диссертациях // *Международ. журн. мед. практики.* 1998. № 4. С. 7–12.

2. *Леонов В. П.* Долгое прощание с лысенковщиной [Электронное издание]. Режим доступа: <http://www.biometrica.tomsk.ru> (дата обращения 10.12.2009).

3. *Леонов В. П.* Применение статистики в статьях и диссертациях по медицине и биологии. Часть 4: Наукометрия статистической парадигмы экспериментальной био-

медицины // *Междунар. журн. мед. практики.* 2002. № 3. С. 6–10.

Материал поступил в редколлегию 14.12.2009

M. I. Musatov

The Structure of Describing The Methods of Statistical Analysis Used in Publications in Immunology Journals

In our retrospective scientometric study, we examined and compared 3 169 original articles published in Russian journals («Immunology», «Russian Journal of Immunology») and foreign journals («Journal of Immunology», «Clinical and Vaccine Immunology», «Clinical and Diagnostic Laboratory Immunology», «International Immunology») for the periods of 1980–1991, 1992–1999 and 2000–2007. Our analysis of the quality of describing the statistical methods used in Russian immunology publications points to lack of validity of research results in those publications. This is mainly due to lack of application of statistical methods, choice of incompatible statistical techniques, or interpretative ambiguity of obtained outcomes. In particular: 1. 3–4 % of publications lose vindictive power, essentially becoming articles without the use of statistical techniques (95 out of 539 papers); 2. Descriptive statistics were only used in 636 publications, whilst in 40–75 % cases authors used non-interpretative phrases like «mean ± something else»; 3. Only statements like 874 publications used « $p < \dots$ », while 208 of those exploited ambiguity of statements; 4. Hence, the share of Russian immunology publications, where authors applied statistical methods, is represented by 56–59 % of the articles, in which authors resorted exclusively to descriptive statistics and statements like « $p < \dots$ ». This frequency is 22–33 times higher than comparable frequency of similar drawbacks in research papers in foreign journals. Furthermore, aforementioned research structure of local publications has remained static and hasn't changed in past 27 years – from 1980 to 2007.

Keywords: biostatistics, statistics in medicine, immunology.