

Гнусные цифры

Дуглас Арнольд и Кристин Фаулер*
Перевод: [Сергей Кузнецов](#), ИСП РАН

Оригинал: Douglas N. Arnold, Kristine K. Fowler. [Nefarious Numbers](#). arXiv.org, Cornell University Library, eprint, last revised 17 Nov 2010 (this version, v4)

Аннотация: Из-за (неуместного) акцентирования важности импакт-факторов повышение значения этого показателя становится целью, к достижению которой стремятся редакторы и издатели журналов. В свою очередь, это становится новым источником проблем. Закон Гудхарта (Goodhart) предупреждает нас, что "когда достижение некоторого показателя становится целью, он перестает быть хорошим показателем". В точности так обстоит дело с импакт-факторами. Их ограниченная полезность еще больше компрометируется манипулированием импакт-факторами, такими действиями над этим пресловутым показателем качества журналов, которые увеличивают значение показателя, но не повышают (а в действительности – понижают) качество журналов.

Введение

Широко распространено использование импакт-фактора в качестве показателя качества журнала. Импакт-факторы применяются в библиотеках при принятии решений о подписке на периодические издания; исследователями, решающими, где стоит публиковаться и что стоит читать; университетскими рекомендательными комитетами, работающими в предположении, что публикация в журнале с более высоким значением импакт-фактора лучше характеризует специалиста; редакторами и издателями как средство оценки и продвижения своих журналов. Импакт-фактор для журнала в заданном году вычисляется ISI (Thomson Reuters) как среднее число появившихся в этом году ссылок на статьи данного журнала, опубликованные за два предыдущих года. Этот показатель подвергается критике с самых разных позиций^{1,2,3,4}:

*Дуглас Арнольд является Макнайтовским (McKnight) президентским профессором математики в университете Миннесоты и президентом Общества промышленной и прикладной математики (Society for Industrial and Applied Mathematics, SIAM). Кристин Фаулер – сотрудник математической библиотеки университета Миннесоты. Авторы выражают признательность Сьюзан Лори (Susan K. Lowry), которая разработала и поддерживала базу данных, использованную в этом исследовании, а также Молли Уайт (Molly T. White).

¹ P. O. Seglen, Why the impact factor of journals should not be used for evaluating research. *BMJ* 314 (1997), 498-502.

² J. Ewing, Measuring journals. *Notices of the AMS* 53 (2006), 1049-1053.

³ R. Golubic, M. Rudes, N. Kovacic, M. Marusic, and A. Marusic, Calculating impact factor: how bibliographical classification of journal items affects the impact factor of large and small journals. *Sci. Eng. Ethics* 14 (2008), 41-49.

⁴ R. Adler, J. Ewing, and P. Taylor, Citation statistics. *Statistical Sciences* 24 (2009), 1-14.

- Распределение ссылок на журнал не определяет его качество.
- Импакт-фактор – это грубая статистика, доносящая только часть информации о распределении ссылок.
- Эта статистика дефектна. С одной стороны, распределение ссылок по статьям является сильно скошенным, так что усреднение по всем статьям журнала обычно бывает недостоверным. С другой стороны, в импакт-факторе учитываются ссылки, появившиеся только за два первых года после публикации (недостаток, особенно существенный для математики, где около 90% ссылок появляется позже).
- Используемая база данных дефектна, содержит ошибки и включает предвзятую подборку журналов.
- Игнорируются многие побочные факторы, например, тип статьи (редакционные, обзорные статьи, письма в противопоставление оригинальным исследовательским статьям), наличие нескольких авторов, самоцитирование, язык публикации и т.д.

Несмотря на эти сложности, для многих оказывается неотразимой привлекательность импакт-фактора как отдельного, легко доступного числового показателя – не требующего сложного разбирательства или привлечения экспертов, но претендующего на характеристику качества журнала. В 2000 г. в статье для информационного бюллетеня для редакторов журналов Амин (Amin) и Мейд (Made) писали⁵, что "в последние годы импакт-фактор из невразумительного библиометрического показателя превратился в основной количественный показатель качества журнала, его исследовательских статей, исследователей, пишущих эти статьи, и даже организаций, в которых они работают". Обычным явлением для журналов стало распространение абсурдных объявлений, рекламирующих их импакт-факторы, подобных тому, которое было разослано по всему миру издательством World Scientific, издающим *International Journal of Algebra and Computation* (IJAC): "Импакт-фактор IJAC вырос с 0,414 в 2007 г. до 0,421 в 2008 г.! Поздравляем редакционный совет и авторов IJAC". В этом случае возрастание импакт-фактора на 1,7% представляет появление одной дополнительной ссылки на одну из 145 статей, опубликованных в журнале за два предыдущих года.

Из-за (неуместного) акцентирования важности импакт-факторов повышение значения этого показателя становится целью, к достижению которой стремятся редакторы и издатели журналов. В свою очередь, это становится новым источником проблем. Закон Гудхарта (Goodhart) предупреждает нас, что "когда достижение некоторого показателя становится целью, он перестает быть хорошим показателем"⁶. В точности так обстоит дело с импакт-

⁵ M. Amin and M. Mabe, Impact factors: use and abuse. *Perspectives in Publishing* 1 (2000), 1-6.

⁶ Эта сжатая формулировка взята из статьи M. Strathern "'Improving Ratings': Audit in the British University System", *European Review* 5 (1997), 305-321.

факторами. Их ограниченная полезность еще больше компрометируется манипулированием импакт-факторами, такими действиями над этим пресловутым показателем качества журналов, которые увеличивают значение показателя, но не повышают (а в действительности – понижают) качество журналов.

Возможно множество форм манипулирования импакт-факторами. В 2007 г. в очерке о вредоносных воздействиях манипулирования импакт-факторами Макдональд (Macdonald) и Кэм (Kam)⁷ иронически заметили, что "благоразумный редактор выращивает кадры постоянных авторов, которые надежно способствуют повышению измеряемого качества журнала путем самоцитирования и цитирования друг друга". Широко распространены жалобы авторов рецензируемых рукописей, которых редакторы просят или вынуждают сослаться на другие статьи из того же журнала. Поскольку судьба публикации статьи автора зависит от решения редактора, эта практика граничит с вымогательством, даже если соответствующее пожелание подается в форме всего лишь намека. В большинстве случаев можно лишь догадываться о наличии подобного давления, но в 2005 г. уже поступали сообщения о явных случаях принуждения от Монастырского (Monastersky)⁸ в *Chronicle of Higher Education* и Бегли (Begley)⁹ в *Wall Street Journal*. Третий установившийся метод, за счет которого редакторы повышают значения импакт-факторов своих журналов, состоит в публикации обзорных заметок с большим числом ссылок на данный журнал. Например, главный редактор *Journal of Gerontology A* взял за правило писать и публиковать в каждом январском номере журнала обзорную статью, опирающуюся на публикации предыдущих двух лет; в 2004 г. 195 из 277 ссылок указывали на *Journal of Gerontology A*. Хотя ущерб, наносимый этими ненаучными методами научной литературе, время от времени вызывает тревогу, многие люди считают влияние подобных методов минимальным или полагают их настолько легко опознаваемыми, что оказываемым ими воздействием можно пренебречь. Для обоснования наличия причин для серьезного беспокойства следует привести контрпример.

Пример IJNSNS

В области прикладной математики имеется наглядный пример, позволяющий исследовать такое искажение импакт-фактора. В последние несколько лет в списках импакт-факторов категории "Mathematics, Applied" доминирует

⁷ S. Macdonald and J. Kam, Aardvark et al.: quality journals and gamesmanship in management studies. *Journal of Information Science* 33 (2007), 702-717.

⁸ R. Monastersky, The number that's devouring science. *Chronicle of Higher Education* 52 (2005).

⁹ S. Begley, Science journals artfully try to boost their rankings. *Wall Street Journal*, 5 June 2006, B1.

International Journal of Nonlinear Sciences and Numerical Simulation (IJNSNS). В 2006-2009 гг. этот журнал занимал первое место, обычно со значительным отрывом, а в 2005 г. находился на втором месте. Однако, как мы увидим, более тщательное изучение показывает, что IJNSNS никоим образом не является ведущим в данной области. Попробуем понять происхождение высокого значения его импакт-фактора.

В 2008 г. (этот год мы рассмотрим наиболее детально) импакт-фактор IJNSNS составлял 8,91, будучи, безусловно, наивысшим среди 175 журналов, относимых к категории прикладной математики в Отчетах о цитируемости журналов (*Journal Citation Report, JCR*) ISI. Для контроля мы также взглянем на два журнала из той же категории со вторым и третьим значениями импакт-фактора: *Communications on Pure and Applied Mathematics (CPAM)* и *SIAM Review (SIREV)*, у которых в 2008 г. значения импакт-фактора составляли 3,69 и 2,80 соответственно. CPAM тесно связан с Курантовским институтом математических наук (*Courant Institute of Mathematical Sciences*), а SIREV является ведущим журналом Общества промышленной и прикладной математики (*Society for Industrial and Applied Mathematics, SIAM*)¹⁰ У обоих журналов имеется превосходная репутация.

Наилучшей альтернативой оценки качества журналов по показателям цитируемости является экспертная оценка. У метода тщательного ранжирования журналов экспертами имеются собственные потенциальные проблемы, но, скорее всего, он обеспечивает намного более точную и целостную оценку качества журнала, чем импакт-фактор и подобные ему показатели. В математике, как и во многих других областях, у исследователей близко совпадают мнения относительно того, какие журналы являются наилучшими в их специальности. Австралийский исследовательский совет (*Australian Research Council*) недавно опубликовал результаты такого сравнения, в котором качественно оценены 20000 рецензируемых журналов разных дисциплин. Этот список разрабатывался на основе широкомасштабного процесса рецензирования с участием научных академий (таких как Австралийская академия наук (*Australian Academy of Science*)), сообществ представителей отдельных дисциплин (таких как Австралийское математическое общество (*Australian Mathematical Society*)) и многих исследователей и экспертов¹¹. Этот рейтинг будет использоваться в 2010 в рамках инициативы повышения уровня научных исследований в Австралии (*Excellence in Research Australia, ERA*), и его принято называть Списком журналов ERA 2010 (*ERA 2010 Journal List*). Рейтинг качества журнала, который оценивает его "общее качество", представляется одной из четырех оценок:

¹⁰ Первый автор данной статьи в настоящее время является президентом SIAM.

¹¹ Australian Research Council, Ranked Journal List Development, http://www.arc.gov.au/era/journal_list_dev.htm.

- A*: один из лучших в своей области или подобласти;
- A: очень высокое качество;
- B: надежная, хотя и не выдающаяся репутация;
- C: не отвечает критериям высокого уровня.

В список ERA вошли 170 из 175 журналов, которым в 2008 г. в JCR был назначен импакт-фактор в категории "Mathematics, Applied". На рис. 1 показаны импакт-факторы журналов для каждого из четырех уровней рейтинга.

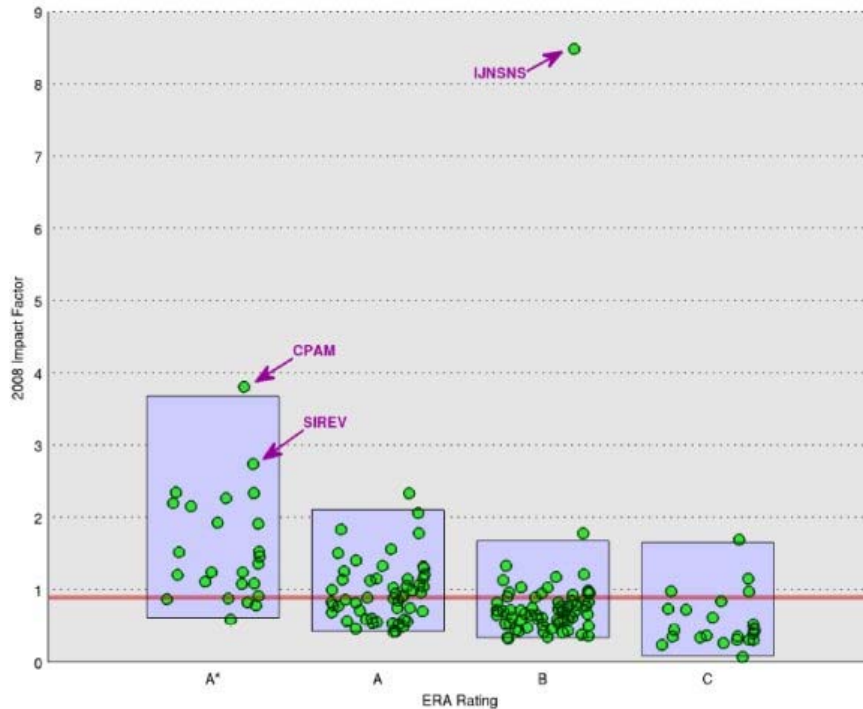


Рис. 1. Импакт-факторы в 2008 г. 170 журналов по прикладной математике, сгруппированные по своим уровням в рейтинге ERA. Для каждого уровня демонстрируется диапазон от 2,5-ой до 97,5-ой перцентили, что позволяет охарактеризовать средние 95%. Горизонтальные позиции точек данных внутри уровней выбраны случайным образом для улучшения наглядности. Красная линия соответствует 20-ой перцентили уровня A*.

Мы видим, что импакт-фактор слабо соответствует оценке экспертов. Имеется много примеров журналов с более высокими значениями импакт-фактора, чем у других журналов, занимающих в рейтинге позиции на один, два и даже на три уровня выше. Под красной линией находится 20% журналов уровня A*;

примечательно, что более высокое значение импакт-фактора имеют 51% журналов уровня A, 23% журналов уровня B и даже 17% журналов уровня C. Самый экстремальный выброс демонстрирует IJNSNS, который, несмотря на относительно огромное значение импакт-фактора, находится не на первом или втором, а скорее на третьем уровне рейтинга. В рейтинге ERA наивысшую оценку A* получили 25 журналов. В этой группе представлено большинство журналов с наивысшими значениями импакт-фактора, включая SPAM и SIREV, но из 10 журналов с наивысшими значениями импакт-фактора два получили оценку A, и только IJNSNS попал в группу B. Всего имелось 53 журнала с рейтингом A и 69 журналов с рейтингом B. Если допустить, что IJNSNS является лучшим среди журналов группы B, то имелось бы 78 журналов с более высоким рейтингом ERA; если же он был бы худшим, то его рейтинг упал бы до 147. Коротко говоря, рейтинг ERA показывает, что IJNSNS не только не является лучшим журналом в области прикладной математики, но что его рейтинг находится где-то в диапазоне 75-150. Это удивительное несоответствие между репутацией журнала и его импакт-фактором нуждается в разъяснении.

Создание высокого импакт-фактора

На первом шаге на пути к пониманию происхождения высокого импакт-фактора IJNSNS мы взглянем на то, сколько авторов внесло существенный вклад в возрастание числа ссылок, и кто такие эти авторы. Автором, который больше всего ссылался на IJNSNS в 2008 г., является главный редактор журнала Цзы-Хуан Хе (Ji-Huan He), сославшийся на журнал (на статьи, вышедшие в течение предыдущего двухлетнего окна) 243 раза. На втором месте со 114 ссылками находится Д.Д. Ганжи (D.D. Ganji), который также является членом редакционного совета, и на третьем месте – региональный редактор Мохамед эль Наши (Mohamed El Naschie) с 58 ссылками. Совместно эти трое несут ответственность за 29% ссылок, учтенных в импакт-факторе. Для сравнения, три автора, больше других сославшиеся на SIREV, сослались на журнал всего 7, 4 и 4 раза соответственно, что составляет менее 12% учтенных ссылок, и ни один из этих авторов не участвовал в редактировании журнала. Для SPAM три автора, наиболее активно сославшиеся на журнал (9, 8 и 8 ссылок), привнесли около 7% от общего числа ссылок, и опять же никто из них не состоял в редакционном совете. Другое заметное явление представляет уровень концентрации ссылок на статьи, опубликованные в пределах двухлетнего окна, которое используется для подсчета импакт-фактора. Наш анализ ссылок, появившихся в 2008 г., на статьи, опубликованные после 2000 г., показывает, что 16% ссылок на SPAM относится к статьям этого двухлетнего окна, а в случае SIREV таких ссылок всего 8%. По сравнению с этим, 71,5% ссылок на IJNSNS ведут на статьи, опубликованные в пределах двухлетнего окна.

В табл. 1 мы показываем импакт-факторы этих трех журналов в 2008 г., а также модифицированные значения импакт-фактора, соответствующие среднему числу ссылок в 2008 г. на статьи, опубликованные в журналах не в 2006-2007 гг., а за предыдущие шесть лет. Поскольку время полураспада цитирования (время, за которое появляется половина от окончательного числа ссылок на статью) для прикладной математики составляет около 10 лет¹², этот показатель, по крайней мере, имеет не меньше смысла, чем импакт-фактор. Кроме того, он независим (в отличие от пятилетнего импакт-фактора JCR), поскольку его период времени не перекрывается с тем, который используется при вычислении двухлетнего импакт-фактора. Заметим, что при использовании другого окна цитирования импакт-фактор IJNSNS стремительно падает, уменьшаясь в семь раз. В отличие от этого, импакт-фактор CPAM остается почти неизменным, а импакт-фактор SIREV заметно возрастает. Можно просто заметить, что, в отличие от контрольных образцов, в ссылках на IJNSNS в 2008 г. наблюдается значительное предпочтение статей, опубликованных точно в течение тех двух лет, которые используются при вычислении импакт-фактора.

Журнал	Импакт-фактор в 2008 г. при использовании обычного окна 2006-2007 гг.	Модифицированный "импакт-фактор" в 2008 г. при использовании окна 2000-2005 гг.
IJNSNS	8.91	1.27
CPAM	3.69	3.46
SIREV	2.8	10.4

Дальнейшее прозрение приходит при анализе журналов, а не авторов с большим числом ссылок. Учет журнальных самоцитирований в импакт-факторе часто подвергается критике, и в данном случае этот подход действительно используется. В 2008 г. в IJNSNS появились 102 ссылки на статьи этого же журнала, или 7% от общего числа ссылок, учитывавшихся при вычислении импакт-фактора. Соответствующими цифрами являются 1 ссылка (0,8%) для SIREV и 8 ссылок (2,4%) для CPAM. Расхождения для других недавних лет настолько же или еще более велики.

Однако наибольшее число ссылок на IJNSNS обеспечил Journal of Physics: Conference Series. В одном выпуске этого журнала содержались 294 ссылки на статьи, опубликованные в пределах окна подсчета импакт-фактора, что составляет более 20% от общего числа ссылок, на основе которых подсчитывался импакт-фактор. И что же это был за выпуск? Это были труды конференции, организованной главным редактором IJNSNS Хе в его родном

¹² В 2010 г. в JCP категории "Mathematics, Applied" было установлено совокупное время полураспада цитирования в 9,5 лет.

университете. Он отвечал за рецензирование статей этого выпуска. На втором месте по числу ссылок на IJNSNS находился журнал Topological Methods in Nonlinear Analysis, который обеспечил 206 ссылок (14%), и все ссылки были размещены в одном выпуске журнала. Это был специальный выпуск с Цзы-Хуан Хе в качестве приглашенного редактора; его соредатор Лан Ксу (Lan Xu) также является членом редакционного совета IJNSNS. Сам Ц.-Х. Хе в этом выпуске опубликовал краткую статью из трех страниц текста и 30 ссылок. Из них 20 ссылок вели на статьи IJNSNS, опубликованные в пределах окна подсчета импакт-фактора. Среди оставшихся ссылок имелось 8 ссылок на Хе и 2 на Ксу.

Продолжая спускаться по списку журналов, активно ссылающихся на IJNSNS, можно обнаружить еще один подобный случай: 50 ссылок из одного выпуска Journal of Polymer Engineering (как и IJNSNS, он публикуется издательством Freund), приглашенными редакторами которого является та же пара Цзы-Хуан Хе и Лан Ксу. Однако третье место занимает журнал Chaos, Solitons & Fractals (CS&F) со 154 ссылками, разбросанными по разным номерам. Это опять такие ссылки, которые можно считать обусловленными влиянием или контролем редакторов. В 2008 г. Цзы-Хуан Хе состоял в редакционном совете CS&F, а главным редактором являлся Мохамед эль Наши, который также был соредатором IJNSNS. При обстоятельствах, получивших широкую огласку, весь редакционный совет CS&F недавно был заменен, но эль Наши остался соредатором IJNSNS.

Многие другие ссылки на IJNSNS появлялись в статьях, публикуемых в журналах, редактором которых являлся Хе; например, 40 ссылок обеспечил журнал Zeitschrift für Naturforschung A. Таких журналов слишком много, чтобы можно было их здесь перечислить, поскольку Хе исполнял редакторские функции более чем в 20 журналах (а теперь назван главным редактором еще четырех журналов недавно образованного издательства Asian Academic Publishers). Еще одним источником многочисленных ссылок являлись статьи, написанные другими редакторами IJNSNS. В общей сложности, совокупность таких родственных редакторам ссылок, для выявления которых требуется много времени, включает более 70% ссылок, использованных для вычисления импакт-фактора IJNSNS.

Библиометрия для отдельных личностей

Библиометрия также используется для оценки отдельных людей, статей, институтов и даже государств. Наукометрическая база данных Ключевых научных показателей (Essential Science Indicators), поддерживаемая компанией Thomson Reuters, рекламируется как средство для ранжирования "ведущих стран, журналов, ученых, статей и институтов в разных исследовательских областях". Однако эти показатели, главным образом, основываются на тех же данных о ссылках, которые используются для подсчета импакт-фактора, и такими показателями так же легко манипулировать (на самом деле, это

делается одновременно). В специальном выпуске Journal of Physics: Conference Series, который редактировался Хе и собрал 243 ссылки на его журнал, также содержатся 353 ссылки на самого Хе. Он утверждает, что общее число ссылок на его статьи превышает 6800¹³. Достижение даже половинного значения этого показателя считается весьма примечательным событием, о чем свидетельствует следующее сообщение, опубликованное на сайте ScienceWatch.com¹⁴: "По данным недавнего анализа Ключевых научных показателей от Thomson Scientific, профессор Цзы-Хуан Хе назван восходящей звездой в области компьютерной науки... В Web of Science зарегистрированы его 137 статей, которые к настоящему времени цитировались 3193 раза". В составе дюжины ученых из всех областей науки Хе упоминался в списках ESI "Наиболее активных исследований 2007-2008 гг." (Hottest Research of 2007-8) и "Наиболее активных исследований 2009 г."

Другим популярным показателем продуктивности исследователей, основанным на учете ссылок, является h-индекс (индекс Хирша (J. Hirsch)). H-индекс отдельного человека – это наибольшее число его (или ее) статей, каждая из которых цитировалась не менее того же числа раз. На этот показатель также действует закон Гудхарта. Хе утверждает, что его h-индекс составляет 39, в то время как сам Хирш оценивает среднее значение своего индекса для Нобелевских лауреатов в области физики как 35¹⁵. Оценки отдельных исследователей или журналов, основанные на учете ссылок, не заменяют экспертных оценок качества.

Заключительные размышления

Несмотря на многочисленные недостатки, импакт-фактор широко используется как показатель качества журналов и даже статей и авторов. Это порождает стимул для манипулирования данным показателем. Более того, можно значительно повысить значение импакт-фактора вообще без повышения качества журнала. Огромное влияние на импакт-фактор может иметь деятельность всего лишь нескольких заинтересованных людей, причем для разоблачения этой деятельности требуются значительные усилия. Мы, главным образом, обсуждали один экстремальный пример, но вряд ли стоит сомневаться в том, что подобные методы применяются во многих журналах, хотя и в меньшей степени (и из-за этого их труднее обнаружить). Совокупным результатом недостатков общего подхода и использования методов

¹³ Это утверждение, а также то, что его h-индекс составляет 39, содержится в биографических сведениях к одной из его недавних статей (Nonl. Sci. Letters 1 (2010), page 1).

¹⁴ ScienceWatch.com, April 2008, <http://sciencewatch.com/inter/aut/2008/08-apr/08aprHe/>.

¹⁵ J. Hirsch, An index to quantify an individual's scientific research output. PNAS 102 (2005), 16569-16572.

манипулирования является то, что импакт-фактор обеспечивает очень неточное представление о качестве журналов. В более общем смысле, ссылки, образующие основу импакт-фактора, и различные другие библиометрические показатели ненадежны по своей природе.

Эта плачевная ситуация имеет огромные последствия. Неверным способом назначаются награды, научная литература и предметные области деформируются, вокруг всего этого растет уровень скептицизма. Что же делать? Как и в самих научных исследованиях, нужно противиться искушению принятия упрощений, когда это может нанести серьезный ущерб точности. Порицают ученые, поддающиеся искушению скрыть некоторые данные или исказить статистику, чтобы получить более понятные результаты. Мы должны придерживаться аналогичного уровня принципиальности при оценке результатов исследований. Администраторы, финансирующие организации, библиотекари и все прочие, кому требуются такие оценки, должны отказаться от упрощенных подходов и принимать важные решения на основе вдумчивости, здравого смысла и экспертизы.