

УДК 001 + 519.25

ББК 78.34

КАК РАБОТАЮТ НАУКОМЕТРИЧЕСКИЕ ПОКАЗАТЕЛИ: ВЫБОРОЧНОЕ ИССЛЕДОВАНИЕ УЧЁНЫХ-БИОЛОГОВ РОССИИ

Новочадов В. В.¹, Широкий А. А.²

(Федеральное государственное автономное образовательное учреждение высшего профессионального образования Волгоградский государственный университет, Волгоград)

В настоящей статье предпринимается попытка проанализировать эффективность применения классических наукометрических показателей, заложенных в открытую базу данных РИНЦ Электронной научной библиотеки eLibrary.ru, на большой репрезентативной выборке учёных-биологов России. В формате методов популяционной биологии предпринята попытка выделить и проанализировать сравнительные показатели в образовавшихся группах учёных-биологов. Сопоставление полученных данных с реальной ситуацией в современной биологии позволили критически отнестись к использованию накопительных показателей научной продуктивности (число публикаций, цитирований, индексы цитирования и Хирша) для оценки труда конкретного учёного-биолога.

Ключевые слова: наукометрия, наукометрические показатели, статистический анализ больших популяций, анализ распределений в выборках

¹ Новочадов Валерий Валерьевич, доктор медицинских наук, профессор (novovv@rambler.ru).

² Широкий Александр Александрович, кандидат физико-математических наук (mhwide@hotmail.com).

1. Введение

Несмотря на то, что наукометрию трудно отнести к новейшим революционным направлениям в исследовании человеческой деятельности, такие факторы, как:

1) включение ее результатов во что-то, напоминающее управляющие воздействия на развитие науки в России (см., например, работы [1], [2], [3]);

2) определённый прогресс в понимании роли и аксиоматическом построении наукометрических процедур за рубежом (см., например, работы [11], [13]);

3) возникновение практических инструментов для массовых исследований в самое последнее время (в частности, доступность надстроек электронной научной библиотеки eLibrary.ru),

актуализируют дискуссию по целому ряду нерешённых проблем. Необходимость такой дискуссии во многом определена парадигмой самой наукометрии и сводится к ответу на два ключевых вопроса: «можно ли и нужно ли измерять и размещать в открытом доступе именно ЭТО?» и «можно ли результаты этих измерений считать основой для принятия управленческих РЕШЕНИЙ?»

Отчетливо осознавая, что в рамках одной статьи невозможно охватить всё многообразие подходов и все противоречия в современной наукометрии (системность развития и междисциплинарные связи, продуктивность отдельных учёных и коллективов, проблема вклада в науку, сравнительная оценка отдельных направлений в науке, адекватность и эффективность инструментария, проблемы наполнения информационной базы, мониторинга эффективности управляющих воздействий и т. д.), мы позволили себе сосредоточиться на том, насколько классические индикаторы научной продуктивности, заложенные в систему РИНЦ электронной научной библиотеки России, отражают объективные закономерности, свойственные любой большой социальной группе, выделяемой по принципу сходства основного вида деятельности. Такие группы по определению должны хорошо описываться популяционными закономерностями

стями, а выбранные показатели – обеспечивать возможность дифференциации внутри такой «популяции» по ключевым различиям в условиях и характере деятельности. Такой подход типичен для современной биологии, медицины, социологии и ряда других наук.

Очевидно, что между ключевым условием принадлежности деятельности к науке в виде равноправного гласного обмена информацией и передачи полезных достижений результатов человечеству и публикацией статей в реферированных журналах не только нет эквивалентности, но второе даже с большим допущением нельзя признать частью первого. Применённые в системе РИНЦ показатели характеризуют, преимущественно, два качества научной деятельности конкретного ученого: публичную продуктивность в профессиональных изданиях достаточного уровня и востребованность этих продуктов (публикаций) аналогичными учеными, то есть участниками публикаций в изданиях того же уровня. Сюда, по факту формирования показателей, не попадает множество других сторон научной деятельности.

В известной работе Налимова и Мульченко [5], проводится аналогия между развитием науки и развитием биологического организма, с определённой периодизацией, информационным и вещественным обменом с окружающей средой. С учетом того, что в реальности мы имеем дело с большим количеством учёных, сгруппированных в динамически меняющиеся коллективы и команды, вполне допустимо (и даже целесообразно) применить к ним законы популяционной биологии. Мы вправе ожидать, что применение математического аппарата изучения популяций, столь продуктивное в различных областях биологии, даст не менее интересные и корректные результаты при анализе количественных показателей деятельности когорты самих учёных. При постановке такой задачи возникают два предположения, правомерность которых может быть проверена при помощи данного подхода.

Во-первых, если по мнению множества авторитетных экспертов существуют качественные различия в результатах научной деятельности учёных, позволяющие выделить среди них

определенные группы (например, молодые талантливые ученые, ведущие отечественные исследователи, мировые лидеры науки, и т. п.), то статистический анализ этих показателей должен выявить признаки этих различий [5]. Во-вторых, предлагаемые новые методы анализа научной деятельности (по крайней мере, новые наукометрические показатели) должны удовлетворять этим же требованиям. Не вполне решённым представляется и вопрос самой группировки учёных, который имеет множество ограничений, связанных с корректностью постановки задачи, ограниченностью информационного ресурса, а также этическими проблемами. Особой проблемой является трудность учёта временного фактора при оценке достигнутых значений наукометрических показателей.

С учетом вышеизложенного, мы попытались изучить при помощи статистического инструментария для больших выборок распределение величин классических наукометрических показателей среди учёных-биологов (по данным РИНЦ) и выработать на этой основе представления об их сравнительной информативности для выделения подгрупп с качественными отличиями в научной деятельности.

2. Анализ наукометрических показателей

2.1. ПРОЦЕДУРА ФОРМИРОВАНИЯ ВЫБОРКИ

Для формирования выборки использовалась база данных электронной научной библиотеки eLibrary.ru, интегрированной с РИНЦ. Выборка производилась следующим образом. Из списка российских журналов, зарегистрированных в eLibrary, были выбраны журналы по тематике «биология» (на момент сбора данных – 3406 журнала). После сортировки по убыванию импакт-фактора были последовательно открыты оглавления 2-го, 32-го и т. д. журнала до конца списка. В каждом случае было случайным образом открыто 6 номеров, вышедших в 2010 – 2012 годах (по 2 за каждый год), в которых также случайным образом выбрано по 3-5 авторов. В итоге была сформирована база данных 502 авторов, публикующихся в журналах по тематике «биология», содержащая следующие сведения: количество

публикаций в РИНЦ на момент проведения исследования, отдельно – число таких публикаций за 2010 – 2012 годы, аналогичным образом – число ссылок на публикации авторов в РИНЦ, а также индекс Хирша на момент исследования. По этическим соображениям персональные данные авторов при создании базы заменялись цифровыми индексами.

2.2. МЕТОДИКА ВЫЯВЛЕНИЯ ПОПУЛЯЦИЙ

Прежде чем перейти к анализу наукометрических показателей, опишем использование в настоящей работе методов популяционной биологии. Мы полагаем, что использование данного подхода можно считать корректным, поскольку «популяция» в широком смысле слова уже давно используется для обозначения любой большой общности участников, связанных закономерностями воспроизводства, существования (жизнедеятельности) и управления.

Для таких систем распределение значения показателя в однородной популяции близко к биномиальному, выявление расщепленных максимумов или нескольких пиков на кривой пиков и отклонения обычно свидетельствует о наличии нескольких субпопуляций, различающихся по данному признаку. При значительной асимметрии кривой или наличия явного цензурирования (справа или слева) мы вправе заподозрить активное влияние внешних управляющих воздействий (см., например, [4]).

Гипотетические кривые распределения мужчин и женщин по росту приведены на рис. 1. У мужчин максимум несколько сглажен, а дисперсия выше в силу биологической целесообразности: мужской пол варьирует генетически детерминированными признаками, отбираемыми женским полом для совместной репродукции.

Естественно, в обобщенной выборке без учета пола мы получим расщепление пика с максимумами для женщин, и для мужчин. Иными словами, признак «рост» позволяет выявить качественное различие между участниками выборки.

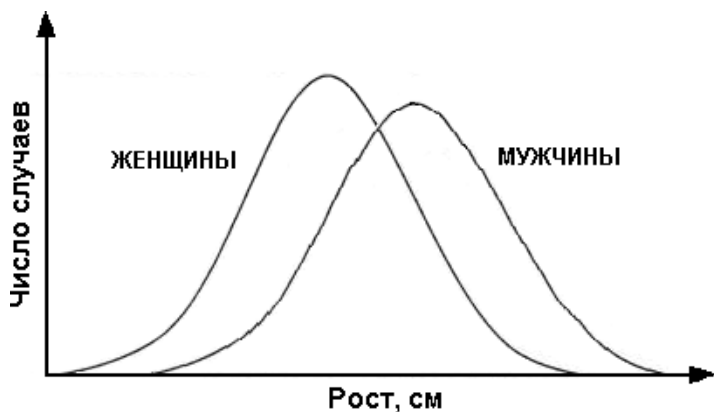


Рис. 1. Распределение взрослых женщин и мужчин по росту

Соответственно, и мы будем называть признак информативным, если в исследуемой популяции удаётся выделить как минимум две группы, различающиеся по этому признаку. В рассматриваемом примере признак «рост» является информативным для выборки в целом, но не для мужчин и женщин по отдельности.

2.3. АНАЛИЗ РАСПРЕДЕЛЕНИЙ ПО РАЗЛИЧНЫМ ПРИЗНАКАМ

Ранее уже отмечалось, что распределение индивидов в отдельно взятой популяции по какому-либо характеризующему признаку стремится к биномиальному. Рассмотрим распределение учёных-биологов из выборки, сформированной в п. 2.1 настоящей статьи, по индексу Хирша (см. рис. 2).

В литературе, в том числе в статьях из настоящего сборника, неоднократно отмечалось, что h -индекс не является информативным показателем. Приведённый на рис. 2 график является ещё одним подтверждением этого факта, поскольку выделить из него какие-либо группы не представляется возможным. Некоторые колебания графика и максимумы в точках 8, 14 и 16 обусловлены, по всей видимости, сравнительно небольшим объёмом выборки.



Рис. 2. Распределение учёных-биологов России по h-индексу

В терминах популяционной биологии это означает, что h-индекс является показателем, аналогичным, скажем, возрасту – то есть, характеризующий «популяцию» учёных в целом. Использовать такой показатель для сравнительной оценки эффективности, разумеется, нельзя.

Рассмотрим теперь распределение учёных-биологов из построенной ранее выборки по числу опубликованных статей, учитываемых РИНЦ (см. рис. 3). Ввиду того, что выборка была относительно небольшой, значения по горизонтали были сгруппированы в интервалы – это позволило получить более гладкую кривую распределения.

Исходя из поведения кривой на рис. 3, можно выделить четыре «популяции» учёных-биологов – видны отчётливые пики на интервалах «от 10 до 19», «от 80 до 89», «от 120 до 129». К четвёртой группе мы отнесём учёных с очень большим количеством публикаций – свыше 150 – таковых всего 52 в выборке (около 10%).



Рис. 3. Распределение учёных-биологов России по числу статей

Первая группа (обозначим её буквой А) самая многочисленная – в неё попало 380 человек из выборки (более 75%). Заметим, что кривая «цензурируется» слева минимальным числом публикаций в РИНЦ – в выборку попали только те учёные, у которых есть хотя бы одна. В то же время, у молодых учёных может не быть статей, опубликованных в индексируемых журналах. Поэтому, если абстрагироваться от смысла значений на горизонтальной оси и достроить левую часть кривой симметрично правой, то получится, что эта «популяция» ещё примерно на 40% многочисленнее. В этом предположении выходит, что 4/5 всех учёных-биологов относятся именно к ней.

Вторая и третья «популяции» (обозначим их буквами В и С) получились примерно одинаковыми по объёму (36 и 34 человека в выборке соответственно) – примерно по 7% от общего числа. Можно предположить, что это две группы профессоров с разной публикационной активностью. Возможно также, что группа на самом деле одна, а «провал» на графике в промежутке от 100 до 119 публикаций – следствие недостаточности объёма выборки.

График ещё одного распределения в плане выделения «популяций» менее информативен, но тоже в некотором смысле показателен (см. рис. 4).



Рис. 4. Распределение учёных-биологов России по приросту статей за 2 года (с 2010 по 2012)

Можно отметить некоторое сходство графиков на рис. 3 и рис. 4. Заметим, что для перемещения из середины группы *A*, в которую попадают молодые учёные, в соседнюю группу *B* в течение 25-30 лет (период активного карьерного роста в высшей школе) требуется издавать в среднем 3 статьи каждый год. В то же время у более чем половины всех учёных из рассматриваемой выборки (301 из 502) за два года вышло менее шести статей. Это означает, что более половины всех учёных так никогда и не достигнут уровня профессора.

В то же время, достаточную для попадания в группу *D* динамику показывают 100 из 502 учёных выборки. Но если среди них отобрать только учёных из группы *A*, то останется уже 40 человек. Напомним, что в группу *A* попало 380 учёных из выборки, что означает, что только у 10,5% молодых учёных-биологов просматривается потенциал для попадания в число ведущих специалистов отрасли. Отметим также известный факт,

что в большинстве направлений науки и, в частности, в биологии соотношение числа кандидатских диссертаций к числу докторских составляет примерно 10 к одному. Это может являть косвенным свидетельством того, что полученные результаты согласуются с реальным положением дел.

2.4 ОБОСНОВАНИЕ РАЗЛИЧИЙ ГРУПП

При разбиении популяции на группы часто встаёт вопрос о том, являются ли эти выборки разными, иначе говоря, требуется обоснование достоверности различий характеристик сравниваемых выборок. В таблице 1 приведены средние значения измеренных характеристик для групп A, B, C, D и величина стандартной ошибки.

Таблица 1. Таблица средних значений измеренных показателей (группы A, B, C, D)

	Группа A	Группа B	Группа C	Группа D
Число респондентов	380	36	34	52
Доля от общего числа	75,7%	7,2%	6,8%	10,4%
Число публикаций (среднее значение)	29,9 ± 0,05	92,3 ± 0,23	126,9 ± 0,31	269,8 ± 2,62
Число цитирований (среднее значение)	105,4 ± 0,41	322,0 ± 7,35	596,6 ± 11,15	1599,0 ± 30,26
Средний РИНЦ	3,1 ± 0,01	3,5 ± 0,08	4,8 ± 0,09	5,7 ± 0,07
Средний h-индекс	4,1 ± 0,01	8,0 ± 0,10	11,0 ± 0,128	16,8 ± 0,16
Публикационная активность за 2 года	4,6 ± 0,01	8,7 ± 0,23	12,5 ± 0,24	20,3 ± 0,38
Цитирований за 2 года	21,6 ± 0,07	46,3 ± 0,93	98,8 ± 2,06	202,4 ± 3,26

Для проверки альтернативной гипотезы для подобных выборок часто применяется критерий Вилкоксона-Манна-Уитни. В

таблице 2 приведены p -значения этого критерия для сравниваемых выборок A и B , B и C , C и D , уровень значимости был выбран наиболее типичный для анализа популяций – 0,05. Ввиду того, что выборки значений показателей по группам A и D значительно превосходят по объёму выборки по группам B и C , в выборках, соответствующих A были оставлены только элементы с номерами, кратными 10, а из выборок, соответствующих D были удалены элементы, кратные 3.

Таблица 2. Сравнение характеристик выборок значений различных показателей для групп A , B , C , D

	р-значение критерия Вилкоксона для выборок, соответствующих группам		
	A и B	B и C	C и D
Число цитирований (среднее значение)	0,000016	0,000028	0,000364
Средний РИНЦ	0,152815	0,015557	0,857535
Средний h -индекс	0,000008	0,000808	0,006836
Публикационная активность за 2 года	0,012211	0,025222	0,002055
Цитирований за 2 года	0,000188	0,000147	0,004019

Исходя из значений, приведённых в таблице 2, можно сделать вывод, что выборки различных групп различаются с вероятностью 95%, за исключением выборок РИНЦ для групп C и D , которые вероятнее всего совпадают, несмотря на заметно отличающееся среднее значение. Несмотря на то, что критерий Вилкоксона не является состоятельным (см., например, п. 3.1 в работе [6]), это является поводом задуматься об эффективности этого наукометрического показателя. Учитывая, что выборки РИНЦ для групп A , B , C всё же различаются, можно предположить, что применение этого показателя оправдано для оценки результатов деятельности молодых исследователей, но не для лидеров научного направления.

2.5. АНАЛИЗ ЗАВИСИМОСТЕЙ

Рассмотрим соотношение количества статей в РИНЦ (на 2012 год) к публикационной активности за два предыдущих года для группы А (см. рис. 5).

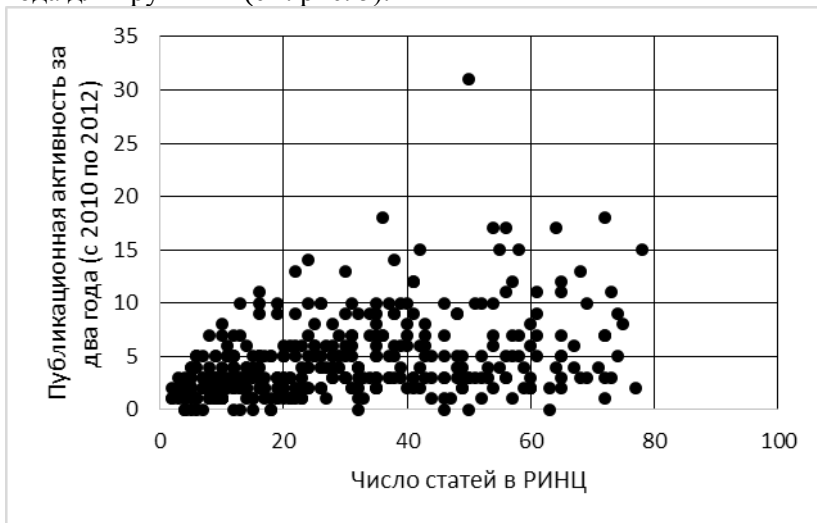


Рис. 5. Соотношение числа статей и публикационной активности (группа А)

Можно отметить, что прирост статей достаточно стабилен и составляет от 0 до 4 статей за два года у почти 60% учёных из группы. Прирост 10 и более статей за тот же период нехарактерен – чуть больше 10%.

Можно рассмотреть соотношение количества статей к числу ссылок (см. рис. 6). Видно, что с ростом количества статей разброс числа цитирований увеличивается. Хочется сделать вывод о зависимости между количеством статей и числом цитирований. Отметим, что у более чем половины респондентов в выборке с количеством статей менее десяти число цитирований не превосходит числа статей. В то же время практически у всех представителей этой группы с более чем сорока статьями число цитирований больше, либо равно числу статей. Это вполне согласуется с общепринятым мнением о том, что число цитиро-

ваний вначале отстаёт от числа статей, а затем обгоняет его. Мы ещё вернёмся к этому вопросу при анализе других «популяций».

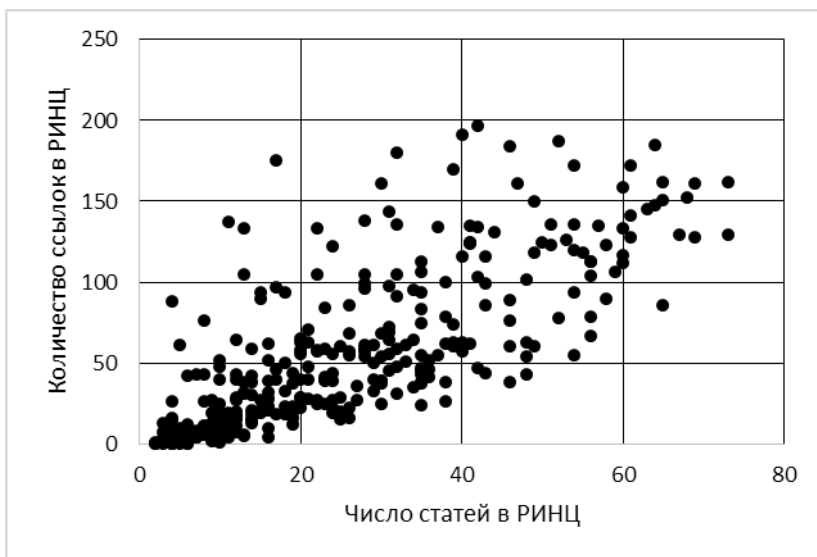


Рис. 6. Соотношение числа статей и цитирований (группа А)

Соотношение числа статей к приросту ссылок за 2 года приведено на рис. 7. Отметим, что на статьи 90% авторов ссылаются за два года до 50 раз, более, чем 70% – до 25 раз.

Авторам представляется, что с учётом специфики рассматриваемой области научного знания, а также учитывая высочайшие темпы развития данного направления науки, такой показатель, как число цитирований за последние несколько лет может оказаться весьма информативным. Дело в том, что вероятность появления ссылки на статью по биологии быстро убывает с её возрастом – приведённые данные быстро устаревают. Таким образом слабо выражен «накопительный эффект», когда на однажды написанную монографию или обзор ссылаются снова и снова, как это бывает в других предметных областях. Имеются основания полагать, что учёные-биологи активно ссылаются на

действительно качественные статьи в активно разрабатываемых предметных областях.

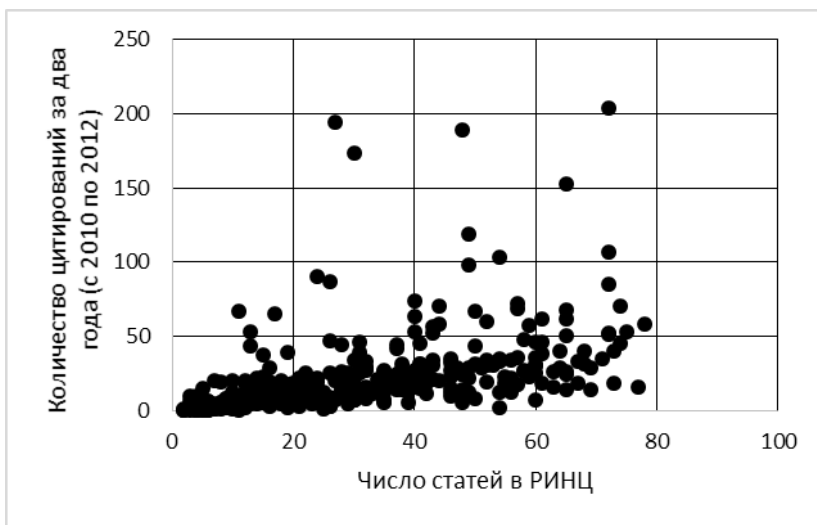


Рис. 7. Соотношение числа статей и цитирований за два года (группа А)

Перейдём теперь к рассмотрению других «популяций». В связи с небольшим количеством представителей (напомним, что в группу А попало 380 респондентов из 502) будем рассматривать графики для групп В, С, D одновременно.

На рис. 8 можно увидеть, какую публикационную активность демонстрируют представители групп В, С, и D с 2010 по 2012 год. Сразу можно отметить, что «продуктивность» учёных с большим количеством статей существенно выше, чем у представителей группы А. Почти половина респондентов демонстрируют публикационную активность за два года на уровне десяти и более статей, а более чем 30% представителей опубликовали за тот же период от десяти до девятнадцати статей. Кроме того, значения показателя публикационной активности 20 и более статей встречаются чаще, чем 1 – 9.

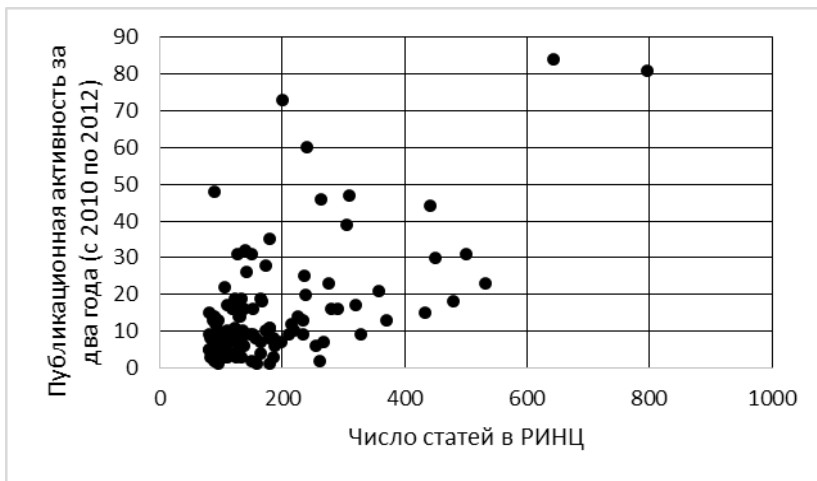


Рис. 8. Соотношение числа статей и публикационной активности (группы B, C, D)

Рассмотрим теперь соотношение числа статей к количеству цитирований (рис. 9).

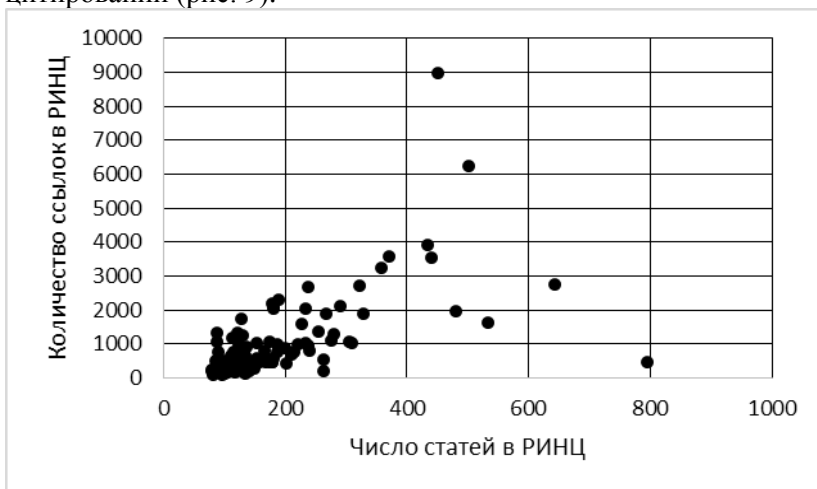


Рис. 9. Соотношение количества статей и цитирований (группы B, C, D)

Интересно отметить, что в выборку не попало ни одного респондента с числом цитирований меньше 78 и всего двое с числом цитирований меньше 100. Гипотеза о нелинейной зависимости числа цитирований от числа статей, по всей видимости, подтверждается – среднее число цитирований с увеличением числа статей растёт в этой группе быстрее. Что же касается количественных характеристик, то более 46% представителей имеют от 100 до 499 цитирований, чуть более четверти – от 500 до 999 (см. рис. 10).

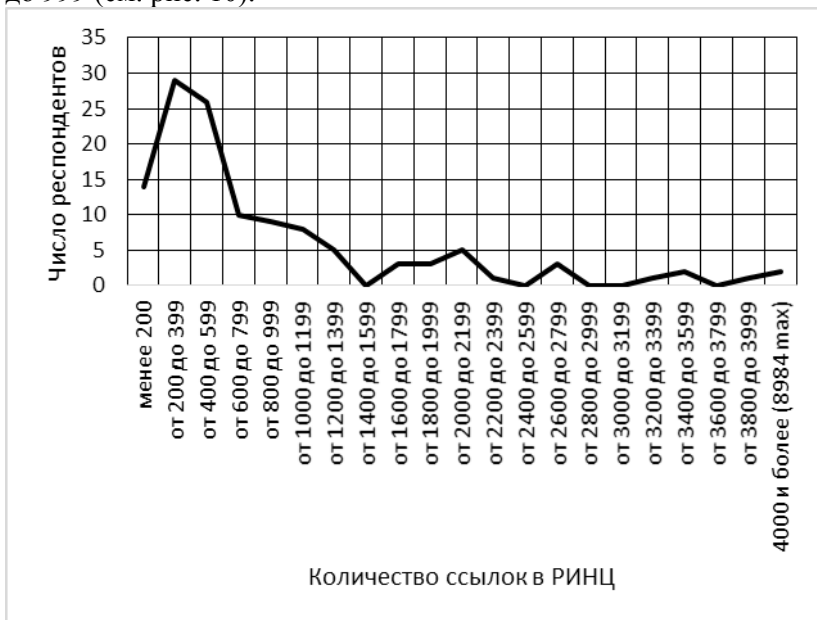


Рис. 10. Распределение учёных по количеству цитирований (группы B, C, D)

На рис. 11 изображено соотношение цитирований за два года и числа статей. Заметно, что респонденты групп B, C, D демонстрируют существенно более высокие значения показателя цитирования, чем представители группы A – учёных из этой группы цитируют чаще и охотнее, чем «молодых исследователей».

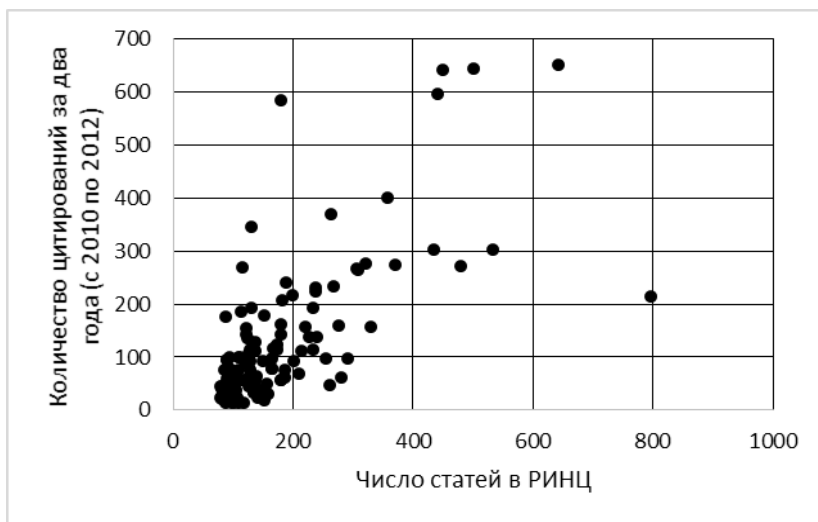


Рис. 11. Соотношение числа статей и цитирований за два года (группы В, С, D)

Так, прирост до 50 ссылок за 2 года демонстрируют только 30% респондентов, а семидесятипроцентный порог лежит в диапазоне от 125 до 149 ссылок.

3. Заключение

Попытаемся подытожить сказанное.

1) Необходимо разделять использование наукометрических индексов в целях объективного количественного описания явлений и процессов, происходящих в той или иной научной области (с учетом индивидуальных различий исследователей), и в целях принятия решений по управлению наукой.

2) Каждая отрасль науки отличается своей спецификой и любые «перекрестные» или, тем более, «сплошные» сравнения требуют учета этой специфики (традиции организации исследований, соотношение эксперимента и теории, характерная численность научных команд и коллективов, средняя активность, темпы научного и карьерного роста и т.д.). Специфика биологи-

ческого знания, технологический прорыв в методиках и текущий этап развития биологии (завершение периода эмпирического накопления и начало формирования базисных общебиологических теорий), по-видимому, определяют максимально высокую научную востребованность большинства научных продуктов (в том числе – журнальных публикаций) в первые 5-10 лет после их выхода в свет.

3) Применение подходов популяционной биологии и статистических методов для «автоматизированного» выделения типовых групп исследователей требует введения дополнительных предположений (иногда достаточно спорных) об их свойствах и структуре. С другой стороны, такие характеристики, как число статей и число их цитирований (а также производные от них показатели), сами по себе не всегда достаточны (или информативны) для решения подобной задачи. Но, несмотря на это, основывающийся на формальном анализе количественных показателей поиск популяционных закономерностей развития науки представляется нетривиальной и актуальной задачей.

Литература

1. ГРАНОВСКИЙ Ю. В. *Трудная судьба науковедения в России* // Науковедческие исследования. – 2010. – С. 110–124.
2. ИВАНЧЕВА Л. *Наукометрия сегодня: методологический обзор* // Международный форум по информации. – 2009. – Т. 34, № 2. – С. 3–8.
3. КОСТЮКОВА М. *Современное состояние и развитие Российского индекса научного цитирования* // Профессиональное образование. Столица. – 2011. – № 2. – С. 38–42.
4. ЛЕВОНТИН Р. *Человеческая индивидуальность: наследственность и среда* // Пер. с англ. – М.: Прогресс, 1993. – 208 с.
5. НАЛИМОВ В. В., МУЛЬЧЕНКО З. М. *Наукометрия. Изучение развития науки как информационного процесса*. – М.: Наука, 1969. – 192 с.

6. ОРЛОВ А. И. *О применении статистических методов в медико-биологических исследованиях* // Вестник Академии медицинских наук СССР. – 1987. – № 2. – С. 88–94.
7. ОРЛОВ А. И. *Прикладная статистика. Учебник.* – М.: Издательство "Экзамен", 2006. – 671 с.
8. ПИТЕРС Д., МАРИШ Р. *Rate my research dot com: измеряем то, что ценим, ценим, что измеряем* // Научная периодика: проблемы и решения. – 2011. – № 1. – С. 40–45.
9. ЯБЛОНСКИЙ А. И. *Математические модели в исследовании науки.* – М.: Наука, 1986. – 352 с.
10. GEISLER E. *The measurement of scientific activity: Research directions in linking philosophy of science and metrics of science and technology outputs* / E. Geisler // Scientometrics. – 2005. – Vol. 62, No. 1. – P. 269–284.
11. LARSEN P.O., VON INS M. *The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index* / P. O. Larsen, M. von Ins // Scientometrics. – 2010. – № 84 (3). – P. 575–603.
12. OUIMET M., BÉDARD P. O., GÉLINEAU F. *Are the h-index and some of its alternatives discriminatory of epistemological beliefs and methodological preferences of faculty members? The case of social scientists in Quebec* / M. Ouimet, P. O. Bédard, F. Gélneau // Scientometrics. – 2011. – № 88 (1). – P. 91–106.
13. PANNE G. *Issues in measuring innovation* / G. Panne // Scientometrics. – 2007. – Vol. 71, № 3. – P. 495–507.
14. VAN RAAN, A. F. J. *Properties of journal impact in relation to bibliometric research group performance indicators* / A. F. J. van Raan // Scientometrics. – 2012. – № 92 (2). – P. 457–469.

HOW SCIENTOMETRICAL INDICES WORK: SELECTIVE RESEARCH OF RUSSIAN BIOLOGISTS

Valery Novochadov, Institute of Natural Sciences of VolSU, Volgograd. MMD, professor (novovv@rambler.ru).

Alexander Shiroky, Institute of Natural Sciences of VolSU, Volgograd. Ph. D. in mathematics, teaching assistant (mhwide@hotmail.com).

Abstract: This article attempts to analyze the effectiveness of the classic scientometrical indices from the public database of electronic scientific library eLibrary.ru on a large representative sample of Russian biologists. We attempt to separate and analyze the comparative indices in the formed groups of biologists using methods of population biology. A comparison of the data with the real situation in modern biology have allowed to take critical look at usage of cumulative indicators of scientific productivity such as number of publications, citations, citation index and h-index to make an assessment of the work of the concerned biologists.

Keywords: scientometrics, scientometrical indices, statistical analysis of large populations, distribution analysis of samples