

Библиометрическая оценка результативности научно-исследовательской работы: к чему мы пришли?

Джованни Абрамо

Статья поступила
в редакцию
в декабре 2016 г.

Джованни Абрамо (Giovanni Abramo) — руководитель Лаборатории исследований в области оценки научной деятельности Института системного анализа и вычислительной техники (*IASI-CNR*) Национального научно-исследовательского совета Италии. Адрес: Via dei Taurini 19, 00185 Rome, Italy. E-mail: giovanni.abramo@uniroma2.it

Аннотация. В данной работе критически анализируются наиболее популярные библиометрические показатели и методики, применяемые для оценки результативности научно-исследовательской работы отдельных ученых и научных учреждений в целом. Цель такого анализа — показать риски, с которыми связано использо-

вание ненаучных интуитивных способов оценивания научной продуктивности и удобных стандартных ответов на непростой вопрос: как оценить научную деятельность? Предлагается методика библиометрической оценки результативности научно-исследовательской работы, которая представляется автору пригодной для ранжирования как отдельных ученых, так и учреждений, объединяющих исследователей, которые работают в разных предметных областях.

Ключевые слова: оценка исследовательской деятельности, продуктивность, относительная научная продуктивность (Fractional Scientific Strength, FSS), рейтинги университетов.

DOI: 10.17323/1814-9545-2017-1-112-127

В условиях сформировавшейся в современном глобализированном мире экономики знаний источниками конкурентного преимущества на уровне как отдельных производств, так и целых государств становятся инновации, основанные на научных исследованиях. Поэтому повышение эффективности и продуктивности национальной научной и технологической инфраструктуры значителен в числе основных приоритетов политической повестки дня для правительств многих стран. Университеты и государственные научные учреждения, будучи ядром этой инфраструктуры, играют важнейшую роль в генерировании и трансляции нового знания и новых научных достижений и, следовательно, приобретают решающее значение в производственной конкуренции, экономическом росте и решении проблемы занятости. Вместе

Abramo G. Bibliometric Evaluation of Research Performance: Where Do We Stand? (пер. с англ. Л. Трониной).

с тем затраты на научные исследования растут, а государственные бюджеты жестко ограничены — в этих условиях возникает необходимость внедрения более эффективных схем распределения ресурсов. Чтобы стимулировать постоянное повышение качества исследований, совершенствовать контроль и учет и эффективнее распоряжаться бюджетными средствами, все больше стран внедряют программы оценивания качества научно-исследовательской работы. Вместе с тем во многих государствах перешли от привычной системы распределения ассигнований — в зависимости от размера научного учреждения и характера исследований — к другой, основанной на результативности научных исследований. Оценивание качества научной работы вносит вклад в достижение пяти принципиальных целей, которые (или часть из которых) встают перед властями многих стран, — это: 1) стимулирование эффективности производства; 2) избирательное распределение ресурсов; 3) сокращение информационной асимметрии между продавцом и покупателем на рынке знаний; 4) формулирование политики и стратегии проведения научных исследований; и последнее — по счету, но не по значению 5) демонстрация эффективности и общественной полезности инвестиций в научные исследования. Как следствие сформировавшегося социального запроса моментально возросла востребованность все более и более точных индикаторов результативности научной работы и методов ее оценивания. За последние годы наукометристы предложили самые разные способы оценки и бесчисленное множество индикаторов и их модификаций, и модификаций модификаций (наукометристы уже задействовали весь алфавит и для обозначения новых индикаторов и их модификаций теперь используют индексы). Количество предлагаемых способов оценивания растет, и это некоторым образом дезориентирует тех, кто отвечает за принятие решений: им уже не под силу соотнести плюсы и минусы тех или иных показателей и методов, разрабатывая конкретную программу оценки. Доказательство тому — все возрастающее число экспертных комиссий и рабочих групп, которые создаются в учебных заведениях, на государственном и межгосударственном уровнях для того, чтобы рассмотреть и рекомендовать к использованию тот или иной показатель или набор показателей, ту или иную методологию оценивания научной деятельности. Национальные и международные рейтинги, отражающие результативность научной деятельности, торжественно публикуются в прессе, оказывая влияние на общественное мнение и на фактический выбор. У автора этой статьи, к сожалению, складывается впечатление, что составители этих рейтингов — «небиблиометристы» (*THE* 2016¹;

¹ THE — Times Higher Education (2016) World University Rankings 2016–2017.

SJTU 2016²; QS 2016³ и т. д.) и даже библиометристы (Лейденский университет⁴, *SCImago*⁵ и т. д.) — опираются в основном на то, что легко подсчитать, а не на то, что следует принимать в расчет.

В данной работе я провожу критический анализ наиболее популярных библиометрических показателей и методологий оценивания результативности научной деятельности отдельных ученых и научных организаций в целом. Моя цель — разогнать туман, которым окружено использование этих показателей, и помочь специалистам-практикам осознать риски, с которыми сопряжено применение ненаучных, интуитивных методик оценивания и удобных стандартных ответов на непростой вопрос: как оценить научную работу? В этой статье вы не найдете принципиально новой информации, отличной от той, что содержится в предыдущих моих работах. Я также прошу прощения за то, что буду много цитировать самого себя и, соответственно, поместил в список литературы в конце статьи много своих работ. Надеюсь, читатель сочтет, что с ними стоит ознакомиться. Новое в этой работе — это систематический обзор современных практик библиометрической оценки результативности научной работы. Комментируя эти практики, я буду прямо высказывать свои критические замечания, как и следует ожидать от того, чьи «идеи в корне отличны от магистрального направления наукометрической мысли» [Waltman, 2016]. Идеи, которые я изложу в этой статье, — результат совместной научно-исследовательской работы, которую вели в течение нескольких лет сотрудники лаборатории, основанной мной вместе с моим коллегой Чириако Андреа Д'Анжело. Разумеется, в ответе за эти идеи я один, однако заслуга осуществления работы, в результате которой они были сформулированы, принадлежит всему научному персоналу и аспирантам — бывшим и нынешним сотрудникам нашей лаборатории. Конечно, критикой существующих методик я не ограничусь — это было бы пустой затеей, а предложу методику библиометрической оценки результативности научно-исследовательской работы, которая представляется мне корректной. В следующей главе я рассмотрю методики оценки, которые не рекомендовал бы использовать, и неадекватные показатели результативности научно-исследовательской работы. Далее я опишу корректный,

<http://www.timeshighereducation.co.uk/world-university-rankings/2016-17/world-ranking>.

² SJTU — Shanghai Jiao Tong University (2016) Academic Ranking of World Universities 2016 <http://www.shanghairanking.com/ARWU2016.html>.

³ QS — Quacquarelli Symonds (2016) World University Rankings 2016. <http://www.topuniversities.com/university-rankings/world-university-rankings>.

⁴ CWTS Leiden Ranking 2016 <http://www.leidenranking.com/>

⁵ <http://www.scimagoir.com/>

с точки зрения сотрудников нашей лаборатории, подход к оценке научной деятельности. В заключительном разделе я представляю выводы.

До сих пор в библиометрической научной литературе предлагаются индикаторы и методы оценки результативности научных исследований, в значительной мере неадекватные с микроэкономической точки зрения. Здесь я критически проанализирую наиболее широко используемые из них. Вероятно, самый яркий пример — индикатор продуктивности научно-исследовательской работы. Под продуктивностью библиометристы традиционно понимают количество публикаций исследователя и рассматривают ее отдельно от степени влияния научной работы, которое измеряется количеством цитат. Честно говоря, мне не удалось проследить, какой ученый первым предложил данное определение, однако еще в 1926 г. Альфред Дж. Лотка использовал такой показатель, как количество публикаций, в своем эпохальном труде [Lotka, 1926], где сформулировал закон, который теперь называют законом научной продуктивности Лотки. Увы, с экономической точки зрения такое определение почти не имеет смысла. Оно было бы приемлемым, если бы все научные публикации имели одинаковую ценность, или степень влияния, а это, конечно же, не имеет ничего общего с истиной. С таким же успехом можно сравнивать две автомобилестроительные компании, одна из которых выпускает *Fiat 500*, а другая — *Ferrari 488*, и утверждать, что они имеют одинаковую продуктивность, поскольку выпускают одинаковое количество автомобилей в день и все производственные факторы у этих компаний тождественны, или измерять ВВП страны количеством произведенных продуктов, не учитывая их рыночную стоимость.

Другая категория неадекватных показателей измеряет среднее число цитирований в расчете на одну публикацию; самый известный ее представитель — средняя нормализованная цитируемость (*Mean Normalized Citation Score, MNCS*). *MNCS* заявлена как индикатор результативности научной работы, измеряющий среднее количество цитирований публикаций отдельного ученого или научного учреждения, нормализованное по предметной области и году публикации [Waltman et al., 2011]. Аналогично доля публикаций ученого или научного учреждения, относящихся к верхнему 1% (10% и т. д.) высокоцитируемых статей (*highly cited articles, HCA*), соотношенная с другими публикациями в той же научной области за тот же год, считается еще одним индикатором результативности научной работы. Свои возражения по поводу данных индикаторов мы уже высказывали [Abramo, D'Angelo, 2016a; 2016b]. Если взять два университета одинакового размера с одинаковыми ресурсами, работающих в одних

Неадекватные библиометрические показатели и рейтинги

и тех же предметных областях, — какой показывает результаты лучше: тот, у которого из 100 статей каждая удостоилась 10 цитирований, или тот, у которого из 200 статей 100 получили по 10 цитирований каждая, а другие 100 — по пять? Университет с 10 высокоцитируемыми статьями из 100 публикаций или с 15 из 200? В соответствии с *MNCS*, или долей высокоцитируемых статей, второй университет показывает результаты хуже, чем первый (на 25% хуже). Но всякому здравомыслящему человеку очевидно, что второй университет более успешен, так как демонстрирует более высокую отдачу от затрат на научные исследования (на 50% больше). Исходя из базовых экономических законов, при равных ресурсах более эффективный исполнитель — тот, кто больше производит, при равном объеме выпуска — тот, кто затрачивает меньше ресурсов. В действительности *MNCS*, доля высокоцитируемых статей (*HCA*), и все остальные индикаторы, измеряющие среднее число цитирований на публикацию, являются неадекватными индикаторами результативности, потому что нарушают аксиому теории производства: если при тех же исходных ресурсах объем выпуска растет, производительность не может считаться падающей. Научное учреждение или отдельный ученый оказываются в парадоксальной ситуации: их рейтинг по *MNCS* снижается, если учреждение или ученый выпускает еще одну статью, нормализованный показатель научного влияния которой оказывается хоть сколько-нибудь ниже значения *MNCS* предыдущей.

Еще один широко известный показатель результативности — *h*-индекс, предложенный в 2005 г. аргентино-американским физиком Хорхе Хиршем [Hirsch, 2005]. *H*-индекс отражает максимальное количество (*h*) работ ученого, которые были процитированы как минимум *h* раз каждая. Хирш вывел обобщенный показатель (целое число), отражающий одновременно и количество работ ученого, и степень их научного влияния, и это было важным достижением. Однако *h*-индекс и большинство его модификаций не учитывают влияние научных работ с количеством цитирований менее *h* и количество цитирований основных работ (*h*), превышающее *h*. Кроме того, *h*-индекс не предусматривает нормализации количества цитирований по предметной области и не учитывает количество соавторов и место каждого в списке авторов. Наконец, последнее, но не менее важное: поскольку интенсивность цитирования научных публикаций в разных предметных областях различается, рейтинг продуктивности на основе *h*-индекса следует выстраивать для каждой предметной области отдельно [Abramo, D'Angelo, 2007], однако на практике довольно часто проводится сравнение *h*-индексов исследователей из разных предметных областей. Каждая из предложенных модификаций *h*-индекса, корректируя одну из его многочисленных погрешностей, не учитывает все остальные, поэтому

ни одну из этих модификаций нельзя считать полностью отвечающей требованиям.

Каждый год мы становимся свидетелями публикации международных рейтингов, которые составляют разные исследовательские организации. Однако прежде чем формировать какое-либо мнение или принимать, основываясь на нем, какие-либо решения, людям, ответственным за их принятие, следует обратить особое внимание на предполагаемые показатели результативности, на которых строятся такие рейтинги. Например, Лейденский рейтинг (*CWTS Leiden Ranking*) 2016 г. формируется на основе таких неадекватных индикаторов, как суммарное количество публикаций и доля высокоцитируемых статей, а до 2015 г. он строился на основании *MNCS*. Тем же недостатком отмечен рейтинг научно-исследовательских организаций *SCImago* (*SCImago Institutions Ranking*) 2016 г.: главным индикатором научной продуктивности в нем служит нормализованный показатель научного влияния, который представляет собой соотношение среднего показателя научного влияния публикаций данного учреждения с общемировым средним показателем научного влияния публикаций, вышедших в тот же период, в той же предметной области, того же издательского формата. Я не буду подробно рассматривать какие-либо из множества ежегодных международных рейтингов научных организаций, составляемых небиблиометристами (*THE* 2016; *SJTU* 2016; *QS* 2016 и т. д.). В этих рейтингах при определении позиции университета индикаторам результативности придаются разные веса. Однако при их применении возникают искажения, обусловленные отсутствием нормирования по предметной области и сильной зависимостью результата рейтингования от размера университета. Например, Шанхайский академический рейтинг университетов мира (*SJTU*) печально известен тем, что лежащий его в основе показатель результативности деятельности вуза более чем на 90% зависит от размера университета. Неудивительно, что такие ненаучные рейтинги широко освещаются в популярных и рекламных СМИ, в то время как в научных изданиях в их адрес звучит много критики.

Что касается национальных программ, созданных для сравнительной оценки результативности университетов и научных организаций, как минимум в 15 странах (Китай, Австралия, Новая Зеландия и 12 стран Евросоюза) такая оценка проводится регулярно, и от ее результатов зависит государственное финансирование [Hicks, 2012]. Развитие библиометрических методик в последнее время привело к тому, что их стали использовать — там, где они применимы, — наряду с более традиционными методиками экспертной оценки. В 2014 г. в Великобритании была введена Программа достижения выдающихся результатов в исследованиях (*Research Excellence Framework, REF*), пришедшая

на смену Экспертной программе оценки качества исследовательской работы в учебных заведениях (*Research Assessment Exercise, RAE*). *REF* стала первой британской программой информированной оценки, в которой при составлении экспертного заключения учитывались данные о цитировании научных работ и другие количественные показатели. Проблема заключается в том, что в программах экспертной оценки или информированной экспертной оценки, проводимых на государственном уровне, заключение делается на основании анализа лишь части научного продукта — из-за дефицита времени и финансовых ограничений. А если программа оценки основана на библиометрических методиках и индикаторах, такие ограничения отсутствуют. Библиометрическая методология дает как минимум два очевидных преимущества: 1) позволяет при оценке результативности отдельных ученых или научных организаций исключить искажения, связанные с неэффективным отбором научного продукта и 2) позволяет исключить искажения, вызванные тем, что оценивается лишь часть научного продукта. Мы впервые дали количественную оценку этих искажений на примере первой итальянской программы оценки эффективности научной деятельности — *VTR 2004–2006* [Abramo, D'Angelo, Caprasecca, 2009]. Мы, в частности, рассчитали погрешность, обусловленную отбором научного продукта для анализа продуктивности ученого или научной организации в области точных наук: результаты указывают на то, что максимальный показатель результативности ухудшается на 23–32% по сравнению с показателем эффективной выборки [Abramo, D'Angelo, Di Costa, 2014]. Мы также провели анализ чувствительности рейтинга результативности к доле оцениваемого научного продукта [Abramo, D'Angelo, Viel, 2010]. Мы продемонстрировали, что с точки зрения точности, робастности, валидности, функциональности, временных и финансовых затрат библиометрические методы имеют преимущества перед экспертной оценкой [Abramo, D'Angelo, Di Costa, 2011]. Однако чиновники и часть академического сообщества решительно сопротивляются замене экспертной оценки библиометрической (там, где она применима) при реализации широкомасштабных программ оценивания.

Корректная методика библиометрической оценки результативности отдельных ученых и научных учреждений

Вместе с моим коллегой Чириако Андреа Д'Анжело мы сформулировали показатель основного индикатора эффективности любой производственной единицы — продуктивности. Несколько лет с его помощью мы измеряли и ранжировали результативность деятельности итальянской профессуры и исследовательских институтов. В отдельной статье мы дали рабочее определение нашего прокси-индикатора продуктивности и объяснили, как его применять [Abramo, D'Angelo, 2014]. В этой главе

я расскажу об основных его аспектах, а за более подробной информацией отсылаю читателя к вышеупомянутой работе.

Научно-исследовательские организации ничем не отличаются от любых других производственных систем. Они используют ресурсы (производственные факторы), чтобы произвести продукт (новое знание). В микроэкономической теории это соотношение выражено хорошо известной производственной функцией: $Q = F(K, L)$, где Q — объем продукции, L — труд, а K — совокупность прочих производственных факторов, помимо L . Для измерения продуктивности в научно-исследовательских системах необходимо учесть природу этих систем, т. е. принять некоторые упрощения и допущения как «на входе», так и «на выходе». Новое знание, т. е. продукт научной работы, нематериально. Поскольку измерить можно лишь то, что удастся представить количественно, в качестве прокси объема произведенных знаний библиометристы используют публикации (которые индексируются в библиометрических базах данных, например *WoS* или *Scopus*). Отсюда напрямую следует, что в тех дисциплинах (прежде всего в сфере искусств и гуманитарных наук), где охват научного продукта библиометрическими базами данных ограничен, библиометрические методики не могут применяться для оценки результативности научной работы. Публикации различаются научной значимостью, т. е. степенью влияния на развитие науки, и это влияние библиометристы в некотором приближении измеряют количеством цитирований. Необходимо отметить, что импакт-фактор научного журнала ни в коем случае не следует использовать в качестве эквивалента показателя цитируемости или в сочетании с ним; это допустимо, только если окно цитирования очень узкое [Abramo, D'Angelo, DiCosta, 2011; Abramo, D'Angelo, DiCosta, 2010a; Levitt, Thelwall, 2011; Stern, 2014; Abramo, D'Angelo, 2016c]. Поскольку традиции цитирования в разных предметных областях различаются, мы нормализуем количество цитирований каждой публикации в соответствии со средним распределением цитирований всех цитируемых публикаций, индексируемых в том же году в той же предметной области⁶. Интенсивность публикаций в разных предметных областях также различается, поэтому необходимым условием свободной от искажений оценки результативности является классифицирование каждого научного работника в рамках только одной предметной области [Abramo et al., 2013a]. Научное исследование зачастую проводит группа ученых, и тогда они фиксируются как соавторы публикации — в этом случае мы принимаем во внимание относительный вклад каждого ученого в данный научный продукт, и ме-

⁶ Среднее распределение цитирований всех цитируемых публикаций в том же году в той же предметной области — наиболее эффективный фактор взвешивания [Abramo, Cicero, D'Angelo, 2012c].

сто, которое конкретный ученый занимает в списке авторов, может указывать, каков же этот вклад.

Дополнительные трудности в оценивании результативности научной работы создает необходимость учета производственных факторов, и здесь неизбежны определенные упрощения. Идентифицировать иные производственные факторы, помимо труда, а также вычислить их полезность и их долю в каждой предметной области — задача очень сложная. (Как оценить количественно полезность накопленного знания или научных инструментов, если они — общие для ряда структурных единиц?) Во многих странах даже определить принадлежность ученого к конкретному институту может быть непросто, не говоря уже о том, чтобы отнести его к одной предметной области. В Италии мы используем базу данных Министерства образования, университетов и научных исследований, где зафиксированы все научные работники, и о каждом содержится следующая информация: принадлежность к научной организации, ученая степень и область научно-исследовательской деятельности. Отнесенность каждого профессора к одной определенной научно-исследовательской области — это, по-видимому, уникальная особенность итальянской системы высшего образования. Области научных исследований формально определены и называются разделами научной дисциплины (*Scientific Disciplinary Sectors, SDSs*) — всего их 370; они объединены в 14 областей университетских дисциплин (*University Disciplinary Areas, UDAs*).

Из-за недостатка информации о капитале (K), находящемся в распоряжении каждого отдельного специалиста или структурной единицы, измерить совокупную производительность факторов производства, как правило, вообще невозможно. В связи с этим зачастую приходится исходить из допущения, что ресурсы, доступные отдельному специалисту или структурной единице, в рамках одной предметной области одинаковы. Далее принимается еще одно допущение (опять же если не имеется особых данных): что количество времени, затраченного на научно-исследовательскую работу, у каждого ученого более или менее одинаково. Наконец, для оценки результатов работы имеет значение и полезность научных работников, а она неодинакова, дифференцирована, что находит отражение в разной стоимости их труда как внутри одного структурного подразделения, так и в разных подразделениях. Если данные о стоимости труда доступны, то объем производимого продукта следует нормализовать по этому показателю.

В процессе оценивания научной продуктивности необходимо конкретизировать и такие параметры, как период публикации и окно цитирования. Выбор периода публикации должен обеспечивать выполнение двух часто противоречащих друг другу требований: гарантировать надежность результатов оценивания и одновременно давать возможность многократного проведения оценки. О целесообразном выборе периода публикации

см.: [Abramo, Cicero, D'Angelo, 2012a], об окне цитирования, которое оптимизирует соотношение между точностью рейтингования и временным интервалом, который охватывает процедура оценки, см.: [Abramo, Cicero, D'Angelo, 2012b].

Наш индикатор — прокси-показатель среднегодовой продуктивности за некоторый период времени — мы назвали показателем относительной научной продуктивности (*Fractional Scientific Strength, FSS*). В случае с отдельным ученым (R) мы измеряем FSS_R , учитывающий стоимость его труда, следующим образом:

$$FSS_R = \frac{1}{W_R} \cdot \frac{1}{t_R} \sum_{i=1}^N \frac{C_i}{\bar{c}} f_i,$$

где: W_R — средняя годовая заработная плата ученого;

t — количество лет работы ученого в рассматриваемом периоде;

N — количество публикаций ученого в рассматриваемом периоде;

C_i — количество цитирований на публикации i ;

\bar{c} — среднее значение распределения цитирований всех процитированных публикаций в том же году и в той же тематической области, к которой относится публикация i ;

f_i — относительный вклад ученого в подготовку публикации i .

Относительный вклад рассчитывается как величина, обратно пропорциональная числу авторов данной публикации, в тех научных областях, где принято располагать авторов просто в алфавитном порядке. Однако в других случаях учитывается степень участия каждого автора в подготовке данной публикации. В сфере медико-биологических наук в Италии широко распространена практика располагать фамилии в списке авторов в порядке, отражающем степень их участия в публикуемой научно-исследовательской работе. Для медико-биологических *SDS* мы присваиваем соавторам публикации разные веса согласно их позиции в списке авторов и характеру участия (очное или заочное) [Abramo et al., 2013b]. Если первый и последний автор относятся к одному университету, каждому присваивается по 40% цитирований, остальные 20% делятся между остальными авторами. Если первые два и последние два автора относятся к разным университетам, по 30% цитирований присваивается первому и последнему авторам, по 15% — второму и предпоследнему, оставшиеся 10% делятся между всеми остальными⁷.

⁷ В других странах могут быть иные подходы, и относительный вклад ученого в научно-исследовательскую работу может оцениваться соответственно с ними.

В Италии сложилась следующая практика: к первичным данным из *WoS* применяется комплексный алгоритм, чтобы обеспечить точность аффилиации ученого с конкретной научной организацией и устранить возможную неопределенность при идентификации его личности, в результате каждая публикация закрепляется за ее автором (авторами) [D'Angelo, Giuffrida, Abramo, 2011]. Благодаря такому алгоритму мы можем составлять рейтинги научной продуктивности отдельных ученых в государственном масштабе. На основе значений FSS_R мы получаем для каждого *SDS* рейтинговый список, который можно представить в виде процентильной шкалы 0–100 (от худших к лучшим) или как отношение FSS_R к средней продуктивности всех итальянских ученых-коллег, работающих в данном *SDS*, с продуктивностью выше нуля⁸. Таким образом, можно сравнить результативность всех итальянских научных работников независимо от того, в каком *SDS* они работают.

При необходимости оценить или сравнить результативность деятельности институций, объединяющих ученых, которые работают в разных предметных областях, — научных дисциплин, кафедр, университетов, регионов, стран — мы сталкиваемся с задачей свести в единый показатель индикаторы продуктивности ученых из разных областей. Мы увидели, что результативность отдельных ученых может быть представлена в виде шкалы процентильных рангов или нормализована в соответствии со средним показателем результативности в данной предметной области. Процентильные ранги не следует складывать или усреднять, так как процентиль — это величина, которая представляет собой результат неравноинтервальных измерений [Thompson, 1993]. К тому же процентильный ранг чувствителен к величине предметной области и распределению показателя результативности. Поэтому мы используем стандартизированный FSS , который учитывает разброс показателей продуктивности отдельных ученых. Формула продуктивности (FSS_U) междисциплинарного научно-исследовательского структурного подразделения (*U*) за определенный период выглядит следующим образом:

$$FSS_U = \frac{1}{RS} \sum_{j=1}^{RS} \frac{FSS_{Rj}}{\overline{FSS_R}},$$

где *RS* — научный персонал подразделения в рассматриваемом периоде;

FSS_{Rj} — продуктивность научного работника *j* в подразделении;

⁸ Среднее распределение продуктивности всех ученых с продуктивностью выше нуля — самый эффективный коэффициент шкалирования для сравнения результативности ученых из разных научных областей [Abramo, Cicero, D'Angelo, 2012c].

\overline{FSS}_R — общенациональный средний показатель продуктивности всех продуктивных научных работников того же SDS , к которому относится научный работник j .

Библиометрические показатели и построенные на их основе рейтинги в подавляющем большинстве имеют два принципиальных ограничения: недостаточная нормализация показателя «на выходе» в соответствии с показателями «на входе» и отсутствие классификации ученых по предметным областям. Без нормализации невозможно измерить продуктивность, которая является главным индикатором деятельности всякой производственной единицы; а если не распределять ученых по предметным областям, рейтинги междисциплинарных научно-исследовательских структурных подразделений неизбежно будут искаженными, поскольку в разных предметных областях интенсивность публикаций разная. Отсюда напрямую следует, что невозможно корректно сравнить продуктивность научных учреждений на международном уровне. По сути дела, международных стандартов классификации ученых не существует, к тому же нам неизвестны государства, кроме Италии и Скандинавских стран, в которых существовали бы свои, национальные классификации ученых по предметным областям. Эту проблему можно отчасти решить, классифицировав научного работника косвенным образом — в соответствии с тем, к каким категориям относят его научный продукт *WoS* или *Scopus*, и выделив затем основную для этого ученого, доминирующую категорию. Относительная научная продуктивность (FSS) — это прокси-показатель продуктивности, позволяющий рассчитывать ее для разных организационных уровней. Безусловно, сам этот показатель и связанные с ним методы оценки могут быть усовершенствованы, тем не менее с точки зрения экономической теории производства они выглядят вполне приемлемыми. Другие показатели (и построенные на их основе рейтинги) — такие как количество (или доля) публикаций научно-исследовательского подразделения или средний нормализованный показатель научного влияния — сами по себе не могут быть основанием для оценки результативности, однако их имеет смысл использовать в сочетании с истинными показателями продуктивности. Если некое научно-исследовательское подразделение достигает среднего уровня продуктивности, это может означать, что оно произвело среднее количество научного продукта средней степени научного влияния или большой объем продукта с низким показателем научного влияния, или малый объем продукта, который имеет высокую научную значимость. Дополнительные показатели продуктивности, такие как количество публикаций и средний нормализованный показатель научного влияния, помогут в этом случае понять, на какой харак-

Выводы

теристике (количество или влияние) научного продукта следует сосредоточить внимание, чтобы повысить эффективность производства.

Кому принадлежит высказывание «Не все, что можно сосчитать, следует принимать в расчет, и не все, что следует принимать в расчет, можно сосчитать» — Альберту Эйнштейну или Уильяму Кэмерону, — вопрос спорный, однако в том, что оно справедливо и чрезвычайно важно для наукометрии, сомнений нет. Всякому, кто занимается оценкой научно-исследовательской работы, надлежит помнить эту мудрость и считать только то, что следует принимать в расчет.

Литература

1. Abramo G., Cicero T., D'Angelo C.A. (2012a) What Is the Appropriate Length of the Publication Period over Which to Assess Research Performance? // *Scientometrics*. Vol. 93. No 3. P. 1005–1017.
2. Abramo G., Cicero T., D'Angelo C.A. (2012b) A Sensitivity Analysis of Researchers' Productivity Rankings to the Time of Citation Observation // *Journal of Informetrics*. Vol. 6. No 2. P. 192–201.
3. Abramo G., Cicero T., D'Angelo C.A. (2012c) Revisiting the Scaling of Citations for Research Assessment // *Journal of Informetrics*. Vol. 6. No 4. P. 470–479.
4. Abramo G., Cicero T., D'Angelo C.A. (2013a) Individual Research Performance: A Proposal for Comparing Apples to Oranges // *Journal of Informetrics*. Vol. 7. No 2. P. 528–539.
5. Abramo G., D'Angelo C.A. (2007) Measuring Science: Irresistible Temptations, Easy Shortcuts and Dangerous Consequences // *Current Science*. Vol. 93. No 6. P. 762–766.
6. Abramo G., D'Angelo C.A. (2014) How Do You Define and Measure Research Productivity? // *Scientometrics*. Vol. 101. No 2. P. 1129–1144.
7. Abramo G., D'Angelo C.A. (2016a) A Farewell to the MNCS and Like Size-Independent Indicators // *Journal of Informetrics*. Vol. 10. No 2. P. 646–651.
8. Abramo G., D'Angelo C.A. (2016b) A Farewell to the MNCS and Like Size-Independent Indicators: Rejoinder // *Journal of Informetrics*. Vol. 10. No 2. P. 679–683.
9. Abramo G., D'Angelo C.A. (2016c) Refrain from Adopting the Combination of Citation and Journal Metrics to Grade Publications, as Used in the Italian National Research Assessment Exercise (VQR2011–2014) // *Scientometrics*. Vol. 109. Iss. 3. P. 2053–2065.
10. Abramo G., D'Angelo C.A., Caprasecca A. (2009) Allocative Efficiency in Public Research Funding: Can Bibliometrics Help? // *Research Policy*. Vol. 38. No 1. P. 206–215.
11. Abramo G., D'Angelo C.A., Di Costa F. (2010a) Citations versus Journal Impact Factor as Proxy of Quality: Could the Latter Ever Be Preferable? // *Scientometrics*. Vol. 84. No 3. P. 821–833.
12. Abramo G., D'Angelo C.A., Di Costa F. (2011) National Research Assessment Exercises: A Comparison of Peer Review and Bibliometrics Rankings // *Scientometrics*. Vol. 89. No 3. P. 929–941.
13. Abramo G., D'Angelo C.A., Di Costa F. (2014) Inefficiency in Selecting Products for Submission to National Research Assessment Exercises // *Scientometric*. Vol. 98. No 3. P. 2069–2086.

14. Abramo G., D'Angelo C.A., Rosati F. (2013b) The Importance of Accounting for the Number of Co-Authors and their Order when Assessing Research Performance at the Individual Level in the Life Sciences // *Journal of Informetrics*. Vol. 7. No 1. P. 198–208.
15. Abramo G., D'Angelo C.A., Viel F. (2010b) Peer Review Research Assessment: A Sensitivity Analysis of Performance Rankings to the Share of Research Product Evaluated // *Scientometrics*. Vol. 85. No 3. P. 705–720.
16. D'Angelo C.A., Giuffrida C., Abramo G. (2011) A Heuristic Approach to Author Name Disambiguation in Bibliometrics Databases for Large-Scale Research Assessments // *Journal of the American Society for Information Science and Technology*. Vol. 62. No 2. P. 257–269.
17. Hicks D. (2012) Performance-Based University Research Funding Systems // *Research Policy*. Vol. 41. No 2. P. 251–261.
18. Hirsch J. E. (2005) An Index to Quantify an Individual's Scientific Research Output // *Proceedings of the National Academy of Sciences*. Vol. 102. No 46. P. 16569–16572.
19. Levitt J. M., Thelwall M. (2011) A Combined Bibliometric Indicator to Predict Article Impact // *Information Processing and Management*. Vol. 47. No 2. P. 300–308.
20. Lotka A. J. (1926) The Frequency Distribution of Scientific Productivity // *Journal of the Washington Academy of Sciences*. Vol. 16. No 12. P. 317–324.
21. Stern D. I. (2014) High-Ranked Social Science Journal Articles Can Be Identified from Early Citation Information // *PLoS ONE*. Vol. 9. No 11. P. 1–11.
22. Thompson B. (1993) GRE Percentile Ranks Cannot Be Added or Averaged: A Position Paper Exploring the Scaling Characteristics of Percentile Ranks, and the Ethical and Legal Culpabilities Created by Adding Percentile Ranks in Making "High-Stakes" Admission Decisions. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA, November 12, 1993.
23. Van Raan A. F. J. (2005a) Fatal Attraction: Conceptual and Methodological Problems in the Ranking of Universities by Bibliometric Methods // *Scientometrics*. Vol. 62. No 1. P. 133–143.
24. Waltman L. (2016) Special Section on Size-Independent Indicators in Citation Analysis // *Journal of Informetrics*. Vol. 10. No 2. P. 645.
25. Waltman L., Van Eck N. J., Van Leeuwen T. N., Visser M. S., Van Raan A. F. J. (2011) Towards a New Crown Indicator: Some Theoretical Considerations // *Journal of Informetrics*. Vol. 5. No 1. P. 37–47.

Bibliometric Evaluation of Research Performance: Where Do We Stand?

Author **Giovanni Abramo**

Head of the Laboratory for Studies in Research Evaluation, Institute for System Analysis and Computer Science (IASI–CNR), National Research Council of Italy. Address: Via dei Taurini, 19, 00185 Roma, Italy. E-mail: giovanni.abramo@uniroma2.it

Abstract This work provides a critical examination of the most popular bibliometric indicators and methodologies to assess the research performance of individuals and institutions. The aim is to raise the fog and make practitioners more aware of the inherent risks in do-it-myself practices, or cozy out-of-the-shelf solutions to the difficult question of how to evaluate research. The manuscript also proposes what we believe is the correct approach to bibliometric evaluation of research performance.

Keywords research evaluation, productivity, FSS, university rankings.

- References**
- Abramo G., Cicero T., D'Angelo C.A. (2012a) What Is the Appropriate Length of the Publication Period over Which to Assess Research Performance? *Scientometrics*, vol. 93, no 3, pp. 1005–1017.
- Abramo G., Cicero T., D'Angelo C.A. (2012b) A Sensitivity Analysis of Researchers' Productivity Rankings to the Time of Citation Observation. *Journal of Informetrics*, vol. 6, no 2, pp. 192–201.
- Abramo G., Cicero T., D'Angelo C.A. (2012c) Revisiting the Scaling of Citations for Research Assessment. *Journal of Informetrics*, vol. 6, no 4, pp. 470–479.
- Abramo G., Cicero T., D'Angelo C.A. (2013a) Individual Research Performance: A Proposal for Comparing Apples to Oranges. *Journal of Informetrics*, vol. 7, no 2, pp. 528–539.
- Abramo G., D'Angelo C.A. (2007) Measuring Science: Irresistible Temptations, Easy Shortcuts and Dangerous Consequences. *Current Science*, vol. 93, no 6, pp. 762–766.
- Abramo G., D'Angelo C.A. (2014) How Do You Define and Measure Research Productivity? *Scientometrics*, vol. 101, no 2, pp. 1129–1144.
- Abramo G., D'Angelo C.A. (2016a) A Farewell to the MNCS and Like Size-Independent Indicators. *Journal of Informetrics*, vol. 10, no 2, pp. 646–651.
- Abramo G., D'Angelo C.A. (2016b) A Farewell to the MNCS and Like Size-Independent Indicators: Rejoinder. *Journal of Informetrics*, vol. 10, no 2, pp. 679–683.
- Abramo G., D'Angelo C.A. (2016c) Refrain from Adopting the Combination of Citation and Journal Metrics to Grade Publications, as Used in the Italian National Research Assessment Exercise (VQR2011–2014). *Scientometrics*, vol. 109, iss. 3, pp. 2053–2065.
- Abramo G., D'Angelo C.A., Caprasecca A. (2009) Allocative Efficiency in Public Research Funding: Can Bibliometrics Help? *Research Policy*, vol. 38, no 1, pp. 206–215.
- Abramo G., D'Angelo C.A., Di Costa F. (2010a) Citations versus Journal Impact Factor as Proxy of Quality: Could the Latter Ever Be Preferable? *Scientometrics*, vol. 84, no 3, pp. 821–833.
- Abramo G., D'Angelo C.A., Di Costa F. (2011) National Research Assessment Exercises: A Comparison of Peer Review and Bibliometrics Rankings. *Scientometrics*, vol. 89, no 3, pp. 929–941.

- Abramo G., D'Angelo C.A., Di Costa F. (2014) Inefficiency in Selecting Products for Submission to National Research Assessment Exercises. *Scientometrics*, vol. 98, no 3, pp. 2069–2086.
- Abramo G., D'Angelo C.A., Rosati F. (2013b) The Importance of Accounting for the Number of Co-Authors and their Order when Assessing Research Performance at the Individual Level in the Life Sciences. *Journal of Informetrics*, vol. 7, no 1, pp. 198–208.
- Abramo G., D'Angelo C.A., Viel F. (2010b) Peer Review Research Assessment: A Sensitivity Analysis of Performance Rankings to the Share of Research Product Evaluated. *Scientometrics*, vol. 85, no 3, pp. 705–720.
- D'Angelo C.A., Giuffrida C., Abramo G. (2011) A Heuristic Approach to Author Name Disambiguation in Bibliometrics Databases for Large-Scale Research Assessments. *Journal of the American Society for Information Science and Technology*, vol. 62, no 2, pp. 257–269.
- Hicks D. (2012) Performance-Based University Research Funding Systems. *Research Policy*, vol. 41, no 2, pp. 251–261.
- Hirsch J. E. (2005) An Index to Quantify an Individual's Scientific Research Output. *Proceedings of the National Academy of Sciences*, vol. 102, no 46, pp. 16569–16572.
- Levitt J. M., Thelwall M. (2011) A Combined Bibliometric Indicator to Predict Article Impact. *Information Processing and Management*, vol. 47, no 2, pp. 300–308.
- Lotka A. J. (1926) The Frequency Distribution of Scientific Productivity. *Journal of the Washington Academy of Sciences*, vol. 16, no 12, pp. 317–324.
- Stern D. I. (2014) High-Ranked Social Science Journal Articles Can Be Identified from Early Citation Information. *PLoS ONE*, vol. 9, no 11, pp. 1–11.
- Thompson B. (1993) *GRE Percentile Ranks Cannot Be Added or Averaged: A Position Paper Exploring the Scaling Characteristics of Percentile Ranks, and the Ethical and Legal Culpabilities Created by Adding Percentile Ranks in Making "High-Stakes" Admission Decisions*. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans, LA, November 12, 1993.
- Van Raan A. F.J. (2005a) Fatal Attraction: Conceptual and Methodological Problems in the Ranking of Universities by Bibliometric Methods. *Scientometrics*, vol. 62, no 1, pp. 133–143.
- Waltman L. (2016) Special Section on Size-Independent Indicators in Citation Analysis. *Journal of Informetrics*, vol. 10, no 2, pp. 645.
- Waltman L., Van Eck N. J., Van Leeuwen T. N., Visser M. S., Van Raan A. F.J. (2011) Towards a New Crown Indicator: Some Theoretical Considerations. *Journal of Informetrics*, vol. 5, no 1, pp. 37–47.