ПРОФЕСОРСКИ ПЛЕНУМ

ВИСТИНАТА ЗА ИМПАКТ ФАКТОРОТ

Избор од критички web-материјали за потребите на Професорскиот пленум

Зима 2014-5

ПРОФЕСОРСКИ ПЛЕНУМ

ВИСТИНАТА ЗА ИМПАКТ ФАКТОРОТ

Избор од критички web-материјали за потребите на Професорскиот пленум

зима 2014-5

Содржина:

I.

Impact Factor Statement

The European Association of Science Editors (EASE)

– ◊ –

The EASE Statement on Inappropriate Use of Impact Factors was published in November 2007 after a consultation exercise. It records how, although the journal impact factor was developed as a means to measure the impact of scientific journals, its use has been extended to measuring the quality of scientific journals, the quality of individual articles and the productivity of individual researchers.

EASE recommends that journal impact factors are used only – and cautiously – for measuring and comparing the influence of entire journals, but not for the assessment of single papers, and certainly not for the assessment of researchers or research programmes either directly or as a surrogate.

2.

ЗА МИТОТ НАРЕЧЕН „ИМПАКТ ФАКТОР“

Катерина Колозова

Истражувањата во општествените науки во Македонија: Состојби, предизвици и препораки, Скопје 2013

– ◊ –

Законот за високото образование определува што подразбира поимот „импакт-фактор“, а тоа се неколку комерцијални рејтинг-бази на академски списанија, кои ја регистрираат цитираноста на списанијата и статиите внатре во самите нив (секоја база за себе одделно). Во своите последни измени и дополнувања, Законот за високото образование („Службен весник на Република Маке-донија“, бр. 15/2013) ги именува базите чиј фактор на влијание треба да се вреднува при изборите, реизборите и акредитирањата за ментори и менторки.

Базите се посредувачки, односно не се создадени, ниту се сопственост на академските издавачки куќи. Зад нив не стојат меѓународни научни тела, релевантни национални и меѓународни асоцијации, академска заедница, која би доделила научен легитимитет на рангирањето. Квантитетот на цитираноста, сам по себе претставува мерило на научната вредност на трудот. Станува збор за комерцијални бази – наедноставно речено: компании што се вградени во Законот: Emerald, Scopus и Thomson Reuters. Привилигирањето на определени бизнис субјекти со слово на Законот е спротивно на самата дефиниција на поимот „закон“, кој треба да ги определува принципите на дејствувањето колку што е можно погенерички, поуниверзално, а не да се заснова на постоењето и работата на конкретни и контингентни бизнис субјекти. Ги нарекуваме „контингентни“, затоа што е прашање на историска контингенција дали веќе утре ќе постојат или, ако постојат, дали воопшто ќе поседуваат референтност.

Значајно е да се напомене, дека ниту една од глобално релевантните издавачки куќи, какви што се, на пример, Sage или Palgrave MacMillan, не учествува во овие посреднички бази и поседува свои, кои не ја индексираат цитираноста. За целите на ова истражување направивме проверка на присуството во базите на индексирање на списанијата од општествените и од хуманистичките науки издавани од штотуку споменатите куќи, како и од уште некои други: Routledge, Oxford University Press, Edinburgh University Press. Не пронајдовме, буквално, ниту едно од нивните позначајни списанија од областите што нè засегаат во оваа анализа. Значи, не само што е проблматично самото привилигирање на деловни субјекти во рамки на еден закон туку и, кога станува збор за општествените и за хуманистички науки, овие три комерцијални бази споменати во последните измени и дополнувања на Законот за високото образование не се ни најмалку референтни.

Референтноста, пак, во поглед на цитирањето и точноста на сликата за релевантност на определена студија или списание, што ја поседуваат овие бази е крајно проблематична. Се мери цитирањето на трудовите на списанија присутни внатре во базите, додека разните комбинаторики на заемно цитирање на издавачите и авторите се појава што сериозно ја поткопува секоја претензија на релевантност на овие бази и нивните индексирања.

Од овие причини, како и од причини што определени дисциплини по самата своја природа и нивната улога во глобалната пазарна економија – а не и на „пазарот“ на научната иновација наложуваат поголема цитираност на списанието од областа и претпазливо [наш курзив] – за мерење и споредување на влијанието на цели списанија, но не и за вреднување на поединечните трудови и, секако, не за вреднување на истражувачите и истражувачките програми“ 1. Советот за финансирање на високото образование на Англија ги потсети Панелите за оцена на истражувачката успешност на универзитетите и поединечните истражувачи дека се должни да го оценуваат квалитетот на секој од трудовите независно од угледот на списанието во кое е издаден. 2. Германската фондација за наука во 2010 издаде слична, па и построга препорака, а, исто така, и Националната научна фондација на САД.

Кога ќе се споредат политиките во поглед на феноменот „импакт-фактор“ во Македонија со оние на развиените академски средини, станува очигледно дека овие вториве настојуваат по пат на препораки и подзаконски акти да се заштитат од академската произволност на базите, додека кај нас се фаворизираат како неприкосновен критериум.

Конечно, привилигирањето објавување статии наспроти книги е во целосна спротивност со општоприфатените скалила на вреднувањето на категориите труд во општествените и во хуманистичките науки. Имено, во овие области, за разлика од природните и од техничките науки, најзначајно постигнување е издавањето монографија, а не статија што се сведува на извештај од спроведено истражување.

Очигледна е претпазливоста на државата во поглед на критериумите, па и на интегритетот на академската заедница, чии претставници треба заемно да го вреднуваат сопствениот труд. Доколку оваа претпоставка е точна, можеме да заклучиме дека законодавецот прави обид да ги обезбеди критериумите на развиените академски средини по пат на закон и, со тоа, да жртвува дел од академската слобода и автономија во име на унапредување на истражувачката извонредност.

Со ова се прави наивен и недоречен потег од гледна точка на градењето ефектни политики – обид бирократски и технички да се обезбеди унапредувањето на академската продукција во Македонија. Наивитетот се состои во претпоставката дека постојат универзални параметри, кои функционираат неприкосновено непристрасно, по пат на автоматизирани „детектори на академската вредност“ – во случајов базите Emerald, Scopus и Thomson Reuters. Сакале или не, арбитрарноста е неизбежна компонента во вреднувањето на што и да е. Академското вреднување претставува вреднување на создавањето знаење, кое е секогаш, според својата природа, непредвидливо во поглед на крајните резултати на неговата вредност и значајност, знаење што е производ на експеримент и ризик. Специфичноста на вреднувањето што треба да се спроведува, наложува неговите критериуми да ги воспоставува, а судовите да ги носи, заедницата на зналци. Со други зборови, рецензиите и специфичните критериуми, кои комисијата или институцијата ќе ги воспостави, мора да бидат главната алатка на процената.

Ако законодавецот има цел да го реши проблемот со парохијалноста, локалната самодоволност и сојузништвото во заемното промовирање на просечноста поради кариеристички конформизам, тогаш препорачуваме наместо вградување контингентни категории во Законот (како називи на бази на цитирање), овозможување на:

- подигање на критериумите за академска извонредност по пат на обуки на академскиот кадар, што учествува во рецензиските процеси, кои ќе ги спроведуваат академски лица од високоразвиени научноистражувачки и високообразовни средини;

- враќање на автономијата на воспоставување критериуми на вреднување назад во рацете на академската средина, придружено со законско решение што пропишува формирање тела за формулирање правилници за избори и реизбори, во кои ќе учествуваат барем една третина претставници на високоразвиените академски заедници во Европа;

- бришење на членовите од законите за наука и за високо образование што ги воспоставуваат цитирањата во базите на индексирање и бројот на статии како критериуми на избор и вреднување на научните работници;

- воведување поопшти, генерички критериуми како „издавање трудови во меѓународни списанија на еден од следниве светски јазици [...]“ и/или „издавање трудови во меѓународни списанија надвор од регионот на Југоисточна Европа“;

- општите критериуми од Законот да стимулираат високи стандарди, кои поконкретно ќе бидат преточени во правилници на институциите, кои нужно ќе се сообразат со овие стандарди;

- да се препознае и да се уважи посебноста на општествените и на хуманистички со тоа што ќе се овозможи повисоко вреднување на монографијата во споредба со статијата во списание;

- да се овозможи воведување добри академски практики и стандарди на извонредност и од други напредни академски средини во Европа, како француската и германската, наспроти привилигирањето на англосаксонската.

3.

Импакт факторот и неговата злоупотреба

Ивица Милевски

ИГЕО-портал, 31.12.2014

– ◊ –

Деновиве една од најчесто застапените теми низ медиумите се импакт факторот (ИФ) или т.н. фактор на влијание и индексирачки бази на научни списанија, цитирања, истражувачи и сл. Во најголем број од медиумите, само површно, паушално, па дури и тенденциозно се изнесуваат наведените поими. Причина за тоа е што со импакт факторот и индекс-базите, наеднаш почнаа да се занимаваат тие што воопшто нити ги засегало, нити биле во „допир“ или прв пат слушнале за спомената проблематика. Затоа ќе се обидеме многу концизно да појасниме што е импакт фактор и што се научни индекс-бази, меѓу кои и фамозниот Web of Science.

Импакт фактор се однесува на просечниот број на цитирања на трудовите објавени во некое списание во претходните 2 години. Тоа значи дека импакт фактор не се однесува на самиот труд, ниту пак на авторите на трудот, туку на списанието како целина и тоа само за дадена година. На пример, да земеме дека некое научно списание има импакт фактор 1 за 2013 година. Тоа значи дека во просек, секој труд објавен во тоа списание во 2011 и 2012 година бил цитиран по еднаш, во некое индексирано научно списание. Следната година, истото научно списание има различен импакт фактор, а разликата обично е до +-50% .

На почетокот, до пред 10-тина години, импакт факторот како „чесен" и „реален" показател за значењето на некое научно списание беше во голем нагорен тренд. Меѓутоа, во последно време во светските научни кругови се покренува голема дебата за тоа колку навистина импакт фактор на некое списание е показател на квалитетот на поединечни трудови во тоа списание. Воедно, речиси невозможно е да се процени поединечниот „фактор на влијание" или „на значење" на некој истражувач, бидејќи во последните години, трудовите објавени во списанија со висок импакт фактор имаат се поголем број на автори по труд., понекогаш дури и над 10 автори во еден труд. Јасно е дека многу тешко може да се увиди индивидуалниот придонес или импактот на поединецот во таква голема група.

Деновиве една од најчесто застапените теми низ медиумите се импакт факторот (ИФ) или т.н. фактор на влијание и индексирачки бази на научни списанија, цитирања, истражувачи и сл. Во најголем број од медиумите, само површно, паушално, па дури и тенденциозно се изнесуваат наведените поими. Причина за тоа е што со импакт факторот и индекс-базите, наеднаш почнаа да се занимаваат тие што воопшто ниту ги засегало, ниту биле во „допир" или прв пат слушнале за спомената проблематика. Затоа ќе се обидеме многу концизно да појасниме што е импакт фактор и што се научни индекс-бази, меѓу кои и фамозниот Web of Science.

Импакт фактор се однесува на просечниот број на цитирања на трудовите објавени во некое списание во претходните 2 години. Тоа значи дека импакт фактор не се однесува на самиот труд, ниту пак на авторите на трудот, туку на списанието како целина и тоа само за дадена година. На пример, да земеме дека некое научно списание има импакт фактор 1 за 2013 година. Тоа значи дека во просек, секој труд објавен во тоа списание во 2011 и 2012 година бил цитиран по еднаш, во некое индексирано научно списание. Следната година, истото научно списание има различен импакт фактор, а разликата обично е до +-50%.

На почетокот, до пред 10-тина години, импакт факторот како „чесен“ и „реален“ показател за значењето на некое научно списание беше во голем нагорен тренд. Меѓутоа, во последно време во светските научни кругови се покренува голема дебата за тоа колку навистина импакт фактор на некое списание е показател на квалитетот на поединечни трудови во тоа списание. Воедно, речиси невозможно е да се процени поединечниот „фактор на влијание“ или „на значење“ на некој истражувач, бидејќи во последните години, трудовите објавени во списанија со висок импакт фактор имаат се поголем број на автори по труд., понекогаш дури и над 10 автори во еден труд. Јасно е дека многу тешко може да се увиди индивидуалниот придонес или импактот на поединецот во таква голема група.

Импакт факторот треба да се користи исклучиво и многу внимателно за споредување на влијанието на научните списанија во рамките на конкретно научно поле, но не за поединечни трудови во тоа списание, а уште помалку како проценка за „вредност“ или „квалитет“ на авторите (истражувачите) на одреден труд.

Импакт факторот не смее да се (зло)употребува за споредба помеѓу истражувачи од различни дисциплини, бидејќи неспоредливо различен е и бројот на списанија со импакт фактор за нив. Така, во природно-техничките и медицински науки, за одредени дисциплини има преку 100 списанија со ИФ поголем од 5, додека за општествените, посебно за некои историско-демографски, лингвистички, етнолошки и други дисциплини, воопшто нема списанија со импакт фактор поголем од 2-3, а и вкупниот број на такви списанија е помал од 30-50.

Германската и Американската научна фондација, во 2010 година веќе донесоа нови упатства, според кои индивидуалниот придонес на истражувачот за постигнување и напредување во научните позиции, како и за финансирање на проекти, менторства и други научни активности, ќе се оценува според неговата севкупна научна работа и квалитетот на објавените трудови. Се заклучува дека во претходниот период преголемо значење му било посветено на импакт факторот, h-индексот и други библиометриски показатели, кои не укажуваат на севкупниот квалитет на научникот.

Поради преголемото значење кое се дава на импакт факторот во проценка на квалитетот на научниот кадар и бројните злоупотреби поврзани со тоа, во 2012 година во Сан Франциско-САД е донесена Declaration on Research Assessment (DORA) која до сега е потпишана од десетина илјади научници и огромен број врвни научни институции.

Самиот Thomson Reuters во чии рамки е познатата датабаза на индексирани списанија Web of Sciences јасно укажува дека „ниту една метрика не може целосно да ја изрази комплексноста на придонесот на истражувачите во нивните научни дисциплини, па затоа освен импакт факторот и индекс базите, многу други научни активности треба да се земат во предвид при проценка на некој истражувач".

Како заклучок, може да се каже дека цитираноста на некој истражувач, бројот на објавени трудови во списанија со импакт фактор, индексираноста на научните списанија во кои се објавени трудовите и слично, треба секако да се земе во предвид при проценка на научната ефектива и да му се даде соодветно значење, со особена внимателност кон различни дисциплини и научни полиња (во различни држави во светот, на овој сегмент му се даваат од 20-70% од севкупната „научна проценка" на истражувачот, извор). Меѓутоа, за севкупната научна активност, потребно е да се процени севкупната научна и стручна ангажираност, посветеност и влијание во дадено поле или дисциплина: од презентации на меѓународни конференции, експертско учество во научни и апликативни проекти, трудови во национални списанија со посебно значење, научни монографии, промоција и популаризација на науката, изработка на дела со посебна содржина (атласи, збирки, акти и сл.). Тоа е особено значајно за мали држави како што е Македонија, која тежнеејќи кон „глобалниот импакт фактор", може целосно да го загуби „националниот научен интерес" (всушност, ваква опасност увиде и Словенија, па затоа изврши одредена корекција во научната евалуација). За некакво реално квантифицирање на сите наведени сегменти, треба да се изготви и постојано да се ажурира детална датабаза на научниот кадар во Република Македонија, а за секоја активност да се утврди одреден индекс или број на поени. Во оваа постапка, многу важно место, покрај останатите, преку своја програма треба да има Националната Универзитетска Библиотека (НУБ), како што е во речиси сите земји на регионот и пошироко.

Она што е важно, дека во светот, најчесто на универзитетите или научните институции и институти им се остава сами да утврдат критериуми за научен квалитет на своите истражувачи. Тоа е во духот на трката по знаење, квалитет и рејтинг на секој универзитет одделно. Всушност, токму по тоа се препознаваат водечките универзитети и научни центри во некоја држава, регион или пак на глобално ниво.

4.

San Francisco Declaration on Research Assessment

Putting science into the assessment of research

– ◊ –

There is a pressing need to improve the ways in which the output of scientific research is evaluated by funding agencies, academic institutions, and other parties.

To address this issue, a group of editors and publishers of scholarly journals met during the Annual Meeting of The American Society for Cell Biology (ASCB) in San Francisco, CA, on December 16, 2012. The group developed a set of recommendations, referred to as the San Francisco Declaration on Research Assessment. We invite interested parties across all scientific disciplines to indicate their support by adding their names to this Declaration.

The outputs from scientific research are many and varied, including: research articles reporting new knowledge, data, reagents, and software; intellectual property; and highly trained young scientists. Funding agencies, institutions that employ scientists, and scientists themselves, all have a desire, and need, to assess the quality and impact of scientific outputs. It is thus imperative that scientific output is measured accurately and evaluated wisely.

The Journal Impact Factor is frequently used as the primary parameter with which to compare the scientific output of individuals and institutions. The Journal Impact Factor, as calculated by Thomson Reuters, was originally created as a tool to help librarians identify journals to purchase, not as a measure of the scientific quality of research in an article. With that in mind, it is critical to understand that the Journal Impact Factor has a number of well-documented deficiencies as a tool for research assessment. These limitations include: A) citation distributions within journals are highly skewed [1–3]; B) the properties of the Journal Impact Factor are field-specific: it is a composite of multiple, highly diverse article types, including primary research papers and reviews [1, 4]; C) Journal Impact Factors can be manipulated (or "gamed") by editorial policy [5]; and D) data used to calculate the Journal Impact Factors are neither transparent nor openly available to the public [4, 6, 7].

Below we make a number of recommendations for improving the way in which the quality of research output is evaluated. Outputs other than research articles will grow in importance in assessing research effectiveness in the future, but the peer-reviewed research paper will remain a central research output that informs research assessment. Our recommendations therefore focus primarily on practices relating to research articles published in peer-reviewed journals but can and should be extended by recognizing additional products, such as datasets, as important research outputs. These recommendations are aimed at funding agencies, academic institutions, journals, organizations that supply metrics, and individual researchers.

A number of themes run through these recommendations:

– the need to eliminate the use of journal-based metrics, such as Journal Impact Factors, in funding, appointment, and promotion considerations;

– the need to assess research on its own merits rather than on the basis of the journal in which the research is published; and

– the need to capitalize on the opportunities provided by online publication (such as relaxing unnecessary limits on the number of words, figures, and references in articles, and exploring new indicators of significance and impact).

We recognize that many funding agencies, institutions, publishers, and researchers are already encouraging improved practices in research assessment. Such steps are beginning to increase the momentum toward more sophisticated and meaningful approaches to research evaluation that can now be built upon and adopted by all of the key constituencies involved.

The signatories of the San Francisco Declaration on Research Assessment support the adoption of the following practices in research assessment.

General Recommendation

1. Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions.

For funding agencies

2. Be explicit about the criteria used in evaluating the scientific productivity of grant applicants and clearly highlight, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.

3. For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.

For institutions

4. Be explicit about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published.

5. For the purposes of research assessment, consider the value and impact of all research outputs (including datasets and software) in addition to research publications, and consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice.

For publishers

6. Greatly reduce emphasis on the journal impact factor as a promotional tool, ideally by ceasing to promote the impact fac-

tor or by presenting the metric in the context of a variety of journal-based metrics (e.g., 5-year impact factor, EigenFactor [8], SCImago [9], h-index, editorial and publication times, etc.) that provide a richer view of journal performance.

7. Make available a range of article-level metrics to encourage a shift toward assessment based on the scientific content of an article rather than publication metrics of the journal in which it was published.

8. Encourage responsible authorship practices and the provision of information about the specific contributions of each author.

9. Whether a journal is open-access or subscription-based, remove all reuse limitations on reference lists in research articles and make them available under the Creative Commons Public Domain Dedication [10].

10. Remove or reduce the constraints on the number of references in research articles, and, where appropriate, mandate the citation of primary literature in favor of reviews in order to give credit to the group(s) who first reported a finding.

For organizations that supply metrics

11. Be open and transparent by providing data and methods used to calculate all metrics.

12. Provide the data under a licence that allows unrestricted reuse, and provide computational access to data, where possible.

13. Be clear that inappropriate manipulation of metrics will not be tolerated; be explicit about what constitutes inappropriate manipulation and what measures will be taken to combat this.

14. Account for the variation in article types (e.g., reviews versus research articles), and in different subject areas when metrics are used, aggregated, or compared.

For researchers

15. When involved in committees making decisions about funding, hiring, tenure, or promotion, make assessments based on scientific content rather than publication metrics.

16. Wherever appropriate, cite primary literature in which observations are first reported rather than reviews in order to give credit where credit is due.

17. Use a range of article metrics and indicators on personal/supporting statements, as evidence of the impact of individual published articles and other research outputs [11].

18. Challenge research assessment practices that rely inappropriately on Journal Impact Factors and promote and teach best practice that focuses on the value and influence of specific research outputs.

References:

Adler, R., Ewing, J., and Taylor, P. (2008) Citation statistics. A report from the International Mathematical Union.

www.mathunion.org/publications/report/citationstatistics0

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

Editorial (2005). Not so deep impact. Nature 435, 1003–1004.

Vanclay, J.K. (2012) Impact Factor: Outdated artefact or stepping-stone to journal certification. Scientometric 92, 211–238.

The PLoS Medicine Editors (2006). The impact factor game. PLoS Med 3(6): e291 doi:10.1371/journal.pmed.0030291.

Rossner, M., Van Epps, H., Hill, E. (2007). Show me the data. J. Cell Biol. 179, 1091–1092.

Rossner M., Van Epps H., and Hill E. (2008). Irreproducible results: A response to Thomson Scientific. J. Cell Biol. 180, 254–255.

http://www.eigenfactor.org/

http://www.scimagojr.com/

http://opencitations.wordpress.com/2013/01/03/open-letter-to-publishers

http://altmetrics.org/tools/

5.

Time to discard the metric that decides how science is rated

David Drubin
The Conversation, 11 June 2014

– ◊ –

Scientists, like other professionals, need ways to evaluate themselves and their colleagues. These evaluations are necessary for better everyday management: hiring, promotions, awarding grants and so on. One evaluation metric has dominated these decisions, and that is doing more harm than good.

This metric, called the journal impact factor or just impact factor, and released annually, counts the average number of times a particular journal's articles are cited by other scientists in subsequent publications over a certain period of time. The upshot is that it creates a hierarchy among journals, and scientists vie to get their research published in a journal with a higher impact factor, in the hope of advancing their careers.

The trouble is that impact factor of journals where researchers publish their work is a poor surrogate to measure an individual researcher's accomplishments. Because the range of citations to articles in a journal is so large, the impact factor of a journal is not really a good predictor of the number of citations to any individual article. The flaws in this metric have been acknowledged widely – it lacks transparency and, most of all, it has unintended effects on how science gets done.

A recent study that attempted to quantify the extent to which publication in high-impact-factor journals correlates with academic career progression highlights just how embedded the impact factor is. While other variables also correlate with the likelihood of getting to the top of the academic ladder, the study shows that impact factors and academic pedigree are rewarded over and above the quality of publications. The study also finds evidence of gender bias against women in career progression and emphasises the urgent need for reform in research assessment.

Judging scientists by their ability to publish in the journals with the highest impact factors means that scientists waste valuable time and are encouraged to hype up their work, or worse, only in an effort to secure a space in these prized journals. They also get no credit for sharing data, software and resources, which are vital to progress in science.

This is why, since its release a year ago, more than 10,000 individuals across the scholarly community have signed the San Francisco Declaration on Research Assessment (DORA), which aims to free science from the obsession with the impact factor. The hope is to promote the use of alternative and better methods of research assessment, which will benefit not just the scientific community but society as a whole.

The DORA signatories originate from across the world, and represent just about all constituencies that have a stake in science's complex ecosystem – including funders, research institutions, publishers, policymakers, professional organisations, technologists and, of course, individual researchers. DORA is an attempt to turn these expressions of criticism into real reform of research assessment, so that hiring, promotion and funding decisions are conducted rigorously and based on scientific judgements.

We can also take heart from real progress in several areas. One of the most influential organisations that is making positive steps towards improved assessment practices is the US National Institutes of Health. The specific changes that have come into play at the NIH concern the format of the CV or "biosketch" in grant applications. To discourage the grant reviewers focusing on the journal in which previous research was published, NIH decided to help reviewers by inserting a short section into the biosketch where the applicant concisely describes their most significant scientific accomplishments.

At the other end of the spectrum, it is just as important to find individuals who are adopting new tools and approaches in how they show their own contributions to science. One such example is Steven Pettifer, a computer scientist at University of Manchester, who gathers metrics and indicators, combining citations in scholarly journals with coverage in social media about his individual articles to provide a richer picture of the reach and influence of his work.

Another example, as reported in the journal Science, comes from one of the DORA authors, Sandra Schmid at the University of Texas Southwestern Medical Center. She conducted a search for new faculty positions in the department that she leads by asking applicants to submit responses to a set of questions about their key contributions at the different stages in their career, rather than submitting a traditional CV with a list of publications. A similar approach was also taken for the selection of the recipients for a prestigious prize recognising graduate student research, the Kaluza Prize.

These examples highlight that reform of research assessment is possible right now by anyone or any organisation with a stake in the progress of science.

One common feature among funding agencies with newer approaches to research assessment is that applicants are often asked to restrict the evidence that supports their application to a limited number of research contributions. This emphasises quality over quantity. With fewer research papers to consider, there is greater chance that the evaluators can focus on the science, rather than the journal in which it is published. It would be encouraging if more of these policies also explicitly considered outputs beyond publications and included resources such as major datasets, resources and software, a move made by the US National Science Foundation in January 2013. After all, the accomplishments of scientists cannot be measured in journal articles alone.

There have been at least two initiatives that focus on metrics and indicators at the article level, from US standards' agency NISOand UK's higher education body HEFCE. Although moves towards a major reliance on such metrics and indicators in research assessment are premature, and the notion of an "article impact factor" is fraught with difficulty, with the development of standards, transparency and improved understanding of these metrics, they will become valuable sources of evidence of the reach of individual research outputs, as well as tools to support new ways to navigate the literature.

As more and more examples appear of practices that don't rely on impact factors and journal names, scientists will realise that they might not be as trapped by a single metric as they think. Reform will help researchers by enabling them to focus on their research and help society by improving the return on the public investment in science.

6.

DEAR DORA

HOWY JACOBS

EMBO REPORTS(2013)14,947
DOI 10.1038/EMBOR.2013.157 |PUBLISHED ONLINE 01.11.2013

– ◊ –

Six months have passed since the publication of the San Francisco Declaration on Research Assessment (DORA). DORA recommended that the evaluation of research portfolios of individuals, institutions or even entire countries should be based on the real content of research, rather than the arbitrary use of impact factor (IF) points. The original signatories included the editors of some of the most prestigious academic journals, directors of top research institutes, as well as many Nobel laureates. Thousands of scientists have subsequently added their voices. DORA can no longer be portrayed as a statement of righteous indignation from an elite pressure group, but has become an expression of the collective view of the worlds' scientists.

Writing as an editor, one of the most astonishing things about the way a journal's IF is perceived is that most scientists consider it essentially immutable. I recently asked the members of my own laboratory to write down what they thought the IF is, without consulting any source. Only about half of them were aware that it had anything to do with citations, and less than a handful were close to the truth. This may account for its widespread misuse. Authors assume that the IF of a journal automatically guarantees their paper a specific level of exposure and recognition, as if it were a badge of quality delivered by an august body of heavenly sages.

The number of citations of individual papers published in any given journal typically follows a highly skewed distribution. A few papers attract many citations, whilst the vast majority are cited infrequently. Citation distributions also vary between journals, between subject areas and over time. If IF were calculated as a median it would look quite different. For assessing individual works, a journal's IF is simply meaningless.

DORA set out to eradicate the tyranny of the impact factor, enjoining funding agencies and institutions to ignore it as a mark of quality, and instead focus on the actual scientific value of published work. The problem is then how to recognize and measure it.

One unfortunate trend that I have noticed is that scientists and institutions increasingly decorate their CVs with actual numbers of citations of each published paper, along with aggregate metrics such as the *h*-index. But we need to ask whether personalized publication metrics are any more reliable an indicator of scientific quality than the cruder system they replace. This was certainly not the intention of the authors of DORA.

Where an article is published is still a major factor influencing how often it is cited. Authors still prefer to cite papers in 'good journals', even if very similar findings were reported elsewhere at the same time. Individual citation metrics are therefore distorted in similar ways as journal-based metrics. But let's imagine that we abolished all journals and instead operated a system whereby academic papers that passed peer-review were published in a single public repository. Would individual citations then be a reliable indicator of quality? A paper can be cited for a variety of reasons. It could report a trail-blazing discovery, or a useful new technique. But it might also just arouse controversy and subsequently prove erroneous. The reputation of the senior author remains important. The field trusts work reported by a distinguished investigator with a stellar track record more than that of a newbie in a far-away country of which little is known. A paper that is frequently cited may indicate originality, but it may also point to the opposite, that is, that the author has just jumped on a bandwagon. It may reflect an easy choice of topic, where much can be learned by obvious and straightforward methods, rather than a painstaking analysis using techniques few can understand. Or it may demonstrate careful attention to serendipitous findings, leading to a new paradigm. And then we have the unsolved problem of the number of co-authors and their order. Is 7th authorship on a paper cited 1,000 times worth more than first authorship on a paper cited 70 times?

Rather than design ever more elaborate ways of objectifying such criteria, we need to follow the spirit of DORA and go back to more traditional means of peer-review. Assessment exercises should be based on a detailed reading of the papers of the person or unit being evaluated, perhaps accompanied by a written or oral justification of their main points and significance.

Even if we excluded all numerical measures of scientific output, we would still face the problem of divergent opinions. An evaluation panel split down the middle on a given portfolio will tend to trash it, even if half the panel thought it is the best science they ever read. Reviewers from Britain, where 80% is a top mark in an exam, may under-rate the best projects. The significance of research in niche areas might be invisible to most outsiders. Conversely, assessors drawn from too narrow a research area may fail to see the broader relevance.

At some point we are going to need a daughter of DORA; an agreed but flexible set of guidelines whereby scientific output can be more meaningfully judged, even if such assessment is laborious and imperfect. In the end, combining many subjective opinions is likely to prove safer and fairer than any supposedly objectified system.

7.

Beyond CVs and Impact Factors: An Employer's Manifesto

Sandra L. Schmid
Department of Cell Biology at the University of Texas Southwestern Medical Center

Science, September 03, 2013

– ◊ –

More than 9000 individuals and 360 institutions have signed the San Francisco Declaration on Research Assessment (DORA), which calls for scientists to reject journal impact factors as a criterion for assessing scientific accomplishment. DORA recommends that institutions "be explicit about the criteria used to reach hiring, tenure, and promotion decisions, clearly highlighting, especially for early-stage investigators, that the scientific content of a paper is much more important than publication metrics or the identity of the journal in which it was published."

It's a nice step—but our signatures are meaningless unless we change our hiring practices. As the new chair of the Department of Cell Biology at the University of Texas Southwestern Medical Center (UT Southwestern), in Dallas, I have the opportunity, over the next few years, to recruit several assistant professors to join us in our recently completed, state-of-the-art lab space. In response to our first call for candidates last year, we received almost 300 applications. Given time and cost constraints, only a few candidates can be chosen to interview. These choices are often based on generic and frequently inadequate information, or on whether the candidate is already known to, or has been directly referred to, the search committee.

This situation repeats itself every year in institutions across the United States, and it makes me wonder how many talented scientists we miss who might have been a perfect fit for the unique research environment provided by our departments.

Let's first look at how applicants for faculty positions are selected currently, at UT Southwestern and elsewhere. Typically, each candidate provides a brief cover letter—but there are few guidelines as to what its content should be. As a result, the utility of the cover letter varies. Statements of past, present, and future research interests, which also vary in length, form, and content, accompany the cover letters.

Often, the only formulaic and succinct document in the application package is the CV. Because CV's can be scanned quickly and compared directly, they frequently become a 'filter' through which an application must pass before the more extensive package is read carefully.

But as a filter the CV is inherently flawed. CVs provide a brief description of past training—including the researcher's pedigree—as well as a list of awards, grants, and publications. A CV provides little insight into attributes that will ensure future success in the right environment. For example, a CV is unlikely to reflect the passion, perseverance, and creativity of individuals who struggled with limited resources and created their own opportunities for compelling research. Nor is a CV likely to identify bold and imaginative risk-takers who might have fallen—for the moment—just short of a major research success. The same is true for those who found, when they realized their goal, that their results exceeded the imaginations of mainstream reviewers and editors, the gatekeepers of high-profile journals. Finally, for junior hires at early stages of their careers, a CV is unlikely to reveal individuals who are adept at recombining knowledge and skills gained from their graduate and postdoctoral studies to carve out new areas of research, or those able to recognize and take advantage of unique opportunities for collaboration in their next position.

Because the information we have access to is limited, we base our decisions on where and for how long the candidate trained, and—too frequently—on the prestige of the journals they published in.

To do a better job of screening applicants—and to avoid inappropriate criteria such as journal impact factors—we need more efficient and direct means to accurately assess the potential of applicants to succeed in our department. Although each department must recruit to match its own vision and scientific priorities, here's how we plan to do it in cell biology at UT Southwestern.

We will be asking applicants to write succinct cover letters describing, separately and briefly, four elements: (1) their most significant scientific accomplishment as a graduate student; (2) their most significant scientific accomplishment as a postdoc; (3) their overall goals/vision for a research program at our institution; and (4) the experience and qualifications that make them particularly well-suited to achieve those goals. Each of the cover letters will be read by several faculty members—all cell biology faculty members will have access to them—and then we will interview, via video conferencing technologies, EVERY candidate whose research backgrounds and future interests are a potential match to our departmental goals.

This approach tests candidates' ability to communicate the significance of their achievements, as well as to articulate a clear vision for and the potential impact of their goals. It also gives candidates a chance to describe the unique experiences in which they have demonstrated the passion, commitment, perseverance, and confidence needed to execute these goals. Evidence for these elements should emerge from the cover letters, and will be explored (with opportunities for synergy with our department) during the video conferencing interviews. The content of the candidates' published papers, and their letters of recommendation, will provide further essential information to help in our selection of the top candidates to visit the department.

Our goal is to identify future colleagues who might otherwise have failed to pass through the singular artificial CV filter of high-impact journals, awards, and pedigree. For example, we encourage applications from candidates who are ready and eager to launch their independent careers, but might feel sidelined because their paper has yet to be, or perhaps won't be, published in a high-impact journal. We believe we can recognize excellence that has been missed by journal editors. By increasing the number of interviewed candidates (and also by not imposing requirements for consensus on a faculty subcommittee) we will increase our chances of identifying individuals who are likely to be the most synergistic, intellectually and personally, with our current and future faculty.

Let's run this experiment!

8.

Halt the avalanche of performance metrics

Colin Macilwain

Nature 500, 255: 14 August 2013 doi:10.1038/500255a

– ◊ –

The leaders of major universities around the world used to maintain a healthy scepticism towards league tables and the metrics that underpin them. But now, officials at institutions that do well in such assessments — partly on merit, and partly because they use the English language or have other historical advantages — are becoming beguiled with quantitative measures to rate the performance of academic staff. People who care about genuine quality in research and teaching need to resist that shift.

Universities evolved as self-governing bodies of academics. Originally, the president or vice-chancellor had a purely housekeeping role, once described in US parlance as assuring parking for the staff, sex for the students and sport for the alumni. But lately — not least in Britain, where schemes such as the Research Assessment Exercise have come to dominate academic life — power has moved from the departments to the vice-chancellor. And university leaders, flanked by research managers and associated flunkies, want to use metrics to shift that balance still further.

Eight leading British universities are now energetically engaged in the joint development of a formidable computer tool that allows them to compare the performances of their researchers and departments against rivals, according to grant income, number of patents applied for, or pretty much any other criteria they choose. The tool is called Snowball (www.snowballmetrics.com) and the institutions signed up to it include the universities of Oxford and Cambridge, Imperial College London and University College London.

Like any metrics system, Snowball can, in theory, be used for good or ill. I suspect that in practice, however, it will end up being used mainly to exercise yet more control over academic staff, with every aspect of their professional lives tracked on the system.

Although Snowball was developed by people of genuine integrity who want to establish a fuller understanding of research performance, it shares a fundamental defect with other quantitative research-assessment tools: it is largely built on sand. It cannot directly measure the quality of research, never mind teaching, so instead it uses weak surrogates, such as the citation indices of individuals.

Citation indices — which rank research in terms of the average number of citations for articles — were robustly challenged earlier this year, when organizations led by the American Society for Cell Biology signed the San Francisco Declaration On Research Assessment (DORA), pledging to take a stand against the ever-expanding reach of journal-based metrics. One of DORA's best ideas is to ask that citation databanks be available openly, for all researchers to use. I wish them luck with that. University managers know that information is power — and they want not just the data, but to dictate how they are manipulated.

A major problem with metrics is the well-charted tendency for people to distort their own behaviour to optimize whatever is being measured (such as publications in highly cited journals) at the expense of what is not (such as careful teaching). Snowball is supposed to get around that by measuring many different things at once. Yet it cannot quantify the attributes that society values most in a university researcher — originality of thinking and the ability to nurture students. Which is not the same as scoring highly in increasingly ubiquitous student questionnaires.

Senior scientists have known for a long time that bogus measures of 'scientific quality' can threaten the peer-review system that has been painstakingly built up, in most leading scientific nations, to distribute funds on the basis of merit. In the United States in 1993, for example, Congress passed the Government Performance and Results Act, which compelled federal agencies to start measuring their results. However, the US scientific establishment was strong and self-assured at that time, and successfully derailed the prospect that agencies such as the National Science Foundation (NSF) would start inventing numbers to 'measure' the work of its grant recipients. Instead, the NSF sticks to measuring things such as time to grant.

Nations with weaker scientific communities are less well-placed to fend off the march of metrics. The hazards are perhaps most immediate in places such as Italy, where peer review for grants has never fully taken hold, and China, where it has rarely even been tried. There is a worrying tendency in developing countries, especially, for research agencies to skip the nuanced business of orchestrating proper peer review, and to move straight to the crude allocation of funds on the basis of measured performance. This bypasses quality and, bluntly, invites corruption.

But I see trouble ahead at the leading universities in the United Kingdom and the United States, too. Their reputations were built by autonomous academics, working patiently with students. If the name of the game becomes strong performance measured in numbers — as the vice-chancellors seem to want — it will kill the goose that laid the golden egg.

Defenders of Snowball say they are baffled that scientists, given what they do for a living, remain sceptical of research-performance metrics. But science seeks to identify and measure good surrogates, to test falsifiable hypotheses. Seen in that light, quantifiable research assessment does not measure up. Nevertheless, the snowball has started rolling down the mountain — and it is hard to see how its momentum will be arrested.

9.

The demise of the Impact Factor:

The strength of the relationship between citation rates and IF is down to levels last seen 40 years ago

George Lozano, The Impact Blog
London School of Economics and Political Science

– ◊ –

Thomson Reuters assigns most journals a yearly Impact Factor (IF), which is defined as the mean citation rate during that year of the papers published in that journal during the previous 2 years. The IF has been repeatedly criticized for many well-known and openly acknowledged reasons. However, editors continue to try to increase their journals' IFs, and researchers continue to try to publish their work in the journals with the highest IF, which creates the perception of a mutually-reinforcing measure of quality. More disturbingly, although it is easy enough to measure the citation rate of any individual author, a journal's IF is often extended to indirectly assess individual researchers. Jobs, grants, prestige, and career advancement are all partially based on an admittedly flawed concept. A recent analysis by myself, Vincent Larivière and Yves Gingras identifies one more, perhaps bigger, problem: since about 1990, the IF has been losing its very meaning.

Impact factors were developed in the early 20th century to help American university libraries with their journal purchasing decisions. As intended, IFs deeply affected the journal circulation and availability. Even by the time the current IF (defined above) was devised, in the 1960s, articles were still physically bound to their respective journals. However, how often these days do you hold in your hands actual issues of printed journals?

Until about 20 years ago, printed, physical journals were the main way in which scientific communication was disseminated. We had personal subscriptions to our favourite journals, and when an issue appeared in our mailboxes (our physical mailboxes), we perused the papers and spent the afternoon avidly reading the most interesting ones. Some of us also had a favourite day of the week in which we went to the library and leafed through the 'current issues' section of a wider set of journals, and perhaps photocopied a few papers for our reprint collection.

Those days are gone. Now we conduct electronic literature searchers on specific subjects, using keywords, author names, and citation trees. As long as the papers are available digitally, they can be downloaded and read individually, regardless of the journal whence they came, or the journal's IF.

This change in our reading patterns whereby papers are no longer bound to their respective journals led us to predict that in the past 20 years the relationship between IF and papers' citation rates had to be weakening.

Using a huge dataset of over 29 million papers and 800 million citations, we showed that from 1902 to 1990 the relationship between IF and paper citations had been getting stronger, but as predicted, since 1991 the opposite is true: the variance of papers' citation rates around their respective journals' IF has been steadily increasing. Currently, the strength of the relationship between IF and paper citation rate is down to the levels last seen around 1970.

Furthermore, we found that until 1990, of all papers, the proportion of top (i.e., most cited) papers published in the top (i.e., highest IF) journals had been increasing. So, the top journals were becoming the exclusive depositories of the most cited research. However, since 1991 the pattern has been the exact opposite. Among top papers, the proportion NOT published in top journals was decreasing, but now it is increasing. Hence, the best (i.e., most cited) work now comes from increasingly diverse sources, irrespective of the journals' IFs.

If the pattern continues, the usefulness of the IF will continue to decline, which will have profound implications for science and science publishing. For instance, in their effort to attract high-quality papers, journals might have to shift their attention away from their IFs and instead focus on other issues, such as increasing online availability, decreasing publication costs while improving post-acceptance production assistance, and ensuring a fast, fair and professional review process.

At some institutions researchers receive a cash reward for publishing a paper in journals with a high IF, usually Nature and Science. These rewards can be significant, amounting to up to $3K USD inSouth Korea and up to $50K USD inChina. InPakistan a $20K reward is possible for cumulative yearly totals. In Europe andNorth America the relationship between publishing in high IF journals and financial rewards is not as explicitly defined, but it is still present. Job offers, research grants and career advancement are partially based on not only the number of publications, but on the perceived prestige of the respective journals, with journal "prestige" usually meaning IF.

I am personally in favour of rewarding good work, but the reward ought to be based on something more tangible than the journal's IF. There is no need to use the IF; it is easy enough to obtain the impact of individual papers, if you are willing to wait a few years. For people who still want to use the IF, the delay would even make it possible to apply a correction for the fact that, independently of paper quality, papers in high IF journals just get cited more often. So, of two equally cited papers, the one published in a low IF journal ought to be considered "better" than the one published in an elite journal. Imagine receiving a $50K reward for a Nature paper that never gets cited! As the relation between IF and paper quality continues to weaken, such simplistic cash-per-paper practices bases on journal IFs will likely be abandoned.

Finally, knowing that their papers will stand on their own might also encourage researchers to abandon their fixation on high IF journals. Journals with established reputations might remain preferable for a while, but in general, the incentive to publish exclusively in high IF journals will diminish. Science will become more democratic; a larger number of editors and reviewers will decide what gets published, and the scientific community at large will decide which papers get cited, independently of journal IFs.

10.

Sick of Impact Factors

by Stephen of Reciprocal Space
Brought to you by Occam's Typewriter

– ◊ –

I am sick of impact factors and so is science. The impact factor might have started out as a good idea, but its time has come and gone. Conceived by Eugene Garfield in the 1970s as a useful tool for research libraries to judge the relative merits of journals when allocating their subscription budgets, the impact factor is calculated annually as the mean number of citations to articles published in any given journal in the two preceding years.

By the early 1990s it was clear that the use of the arithmetic mean in this calculation is problematic because the pattern of citation distribution is so skewed. Analysis by Per Seglen in 1992 showed that typically only 15% of the papers in a journal account for half the total citations. Therefore only this minority of the articles has more than the average number of citations denoted by the journal impact factor. Take a moment to think about what that means: the vast majority of the journal's papers — fully 85% — have fewer citations than the average. The impact factor is a statistically indefensible indicator of journal performance; it flatters to deceive, distributing credit that has been earned by only a small fraction of its published papers.

But the real problem started when impact factors began to be applied to papers and to people, a development that Garfield never anticipated. I can't trace the precise origin of the growth but it has become a cancer that can no longer be ignored. The malady seems to particularly afflict researchers in science, technology and medicine who, astonishingly for a group that prizes its intelligence, have acquired a dependency on a valuation system that is grounded in falsity. We spend our lives fretting about how high an impact factor we can attach to our published research because it has become such an important determinant in the award of the grants and promotions needed to advance a career. We submit to time-wasting and demoralising rounds of manuscript rejection, retarding the progress of science in the chase for a false measure of prestige.

Twenty years on from Seglen's analysis a new paper by Jerome Vanclay from Southern Cross University in Australia has reiterated the statistical ineptitude of using arithmetic means to rank journals and highlighted other problems with the impact factor calculation. Vanclay points out that it fails to take proper account of data entry errors in the titles or dates of papers, or of the deficient and opaque sampling methods used by Thomson Reuters in its calculation. Nor, he observes, does the two-year time limit placed on the impact factor calculation accommodate variations in the temporal citation patterns between different fields and journals; peak citations to Nature papers occurs 2-3 years following publication whereas citations of papers in Ecology take much more time to accrue and are maximal only after 7-8 years). Whichever way you look, the impact factor is a mis-measure.

Vanclay's paper is a worthy addition to the critical literature on the impact factor. Its defects and perverse effects are well known and have been dissected by David Colquhoun, Michael Eisen andPeter Lawrence, among others. Even Philip Campbell, editor-in-chief of Nature, which has one of the highest impact factors in the business, has recognised that we need to escape its dispiriting hold over the lives of researchers.

Writing in 2008, Campbell (albeit somewhat uncertainly) saw a possible solution to the impact factor conundrum in the rise of mega-journals like PLoS ONE, which publish exclusively online and judge papers only on their novelty and technical competence, and in the potential of article-level metrics to assess the scientific worth of papers and their authors. In the end, however, he couldn't shake the editorial habit of selection, writing of the contents of archives and mega-journals: "nobody wants to have to wade through a morass of papers of hugely mixed quality, so how will the more interesting papers […] get noticed as such?" Four years later such views are being buffeted by the rising tides of open access and social media. It might sound paradoxical but nobody should have to wade through the entire literature because everybody could be involved in the sifting.

The trick will be to crowd-source the task. Now I am not suggesting we abandon peer-review; I retain my faith in the quality control provided by expert assessment of manuscripts before publication, but this should simply be a technical check on the work, not an arbiter of its value. The long tails of barely referenced papers in the citation distributions of all journals — even those of high rank — are evidence enough that pre-publication peer review is an unreliable determinant of ultimate worth. Instead we need to find ways to attach to each piece of work the value that the scientific community places on it though use and citation. The rate of accrual of citations remains rather sluggish, even in today's wired world, so attempts are being made to capture the internet buzz that greets each new publication; there are interesting innovations in this regard from the likes of PLOS, Mendeley and altmetrics.org.

The old guard may be shaking their heads and murmuring darkly about gaming of any system that tries to capture the web-chatter sparked by new research. But they shouldn't be so concerned. Any working scientist will have experienced the thrill of hearing exciting new findings reported at a conference where results do not need to be wrapped between the covers of a particular journal for their significance to be appreciated. All it takes is for people to gather together in the coffee break and talk. The web allows something akin that process to be energised on a daily basis; if we tap in online as the community of scientists downloads and flags up the papers of greatest interest to them, we could recover a sense of the worth of the scientific literature (and the efforts behind it) that is factual rather than fictional.

These developments go hand in hand with the rise of open access (OA) publishing. Though primarily motivated by the research and societal benefits that will accrue from freeing the dissemination of the research literature, open access is also needed to optimise crowd-sifting of the literature by making it accessible to everyone. But the growth of open access is also being held back by the leaden hand of the impact factor. This year has seen several significant policy developments in the US, EU and UK, but we still have a considerable way to go. In the long term open access can only work by moving to a gold 'author pays' model that has to be funded by monies released from subscription cancellations, but while we continue to place false value in impact factors, the publishers of high ranking journals can claim that the cost of sifting and rejecting scores of manuscripts must be borne by the system and therefore warrants exorbitant charges for gold OA.

It doesn't have to be this way. We can avoid high cost gold OA and achieve a system of valuation that works by ridding ourselves of the impact factor.

I don't wish to under-estimate the difficulties. I am well aware of the risks involved, particularly to young researchers trying to forge a career in a culture that is so inured to the impact factor. It will take a determined and concerted effort from those in a position of influence, not least by senior researchers, funders and university administrators. It won't be easy and it won't be quick. Two decades of criticism have done little to break the addiction to a measure of worth that is statistically worthless.

But every little helps, so, taking my cue from society's assault on another disease-laden dependency, it is time to stigmatise impact factors the way that cigarettes have been. It is time to start a smear campaign so that nobody will look at them without thinking of their ill effects, so that nobody will mention them uncritically without feeling a prick of shame.

So consider all that we know of impact factors and think on this: if you use impact factors you are statistically illiterate.

If you include journal impact factors in the list of publications in your cv, you are statistically illiterate.

If you are judging grant or promotion applications and find yourself scanning the applicant's publications, checking off the impact factors, you are statistically illiterate.

If you publish a journal that trumpets its impact factor in adverts or emails, you are statistically illiterate. (If you trumpet that impact factor to three decimal places, there is little hope for you.)

If you see someone else using impact factors and make no attempt at correction, you connive at statistical illiteracy.

The stupid, it burns. Do you feel the heat?

11.

WHY YOU SHOULD NOT USE THE JOURNAL IMPACT FACTOR TO EVALUATE RESEARCH

EDITAGE INSIGHTS, RESOURCES FOR AUTHORS AND JOURNALS

– ◊ –

Eugene Garfield,1 the founder of the Journal Impact Factor (JIF), had originally designed it as a means to help choose journals. Unfortunately, the JIF is now often used inappropriately, for example, to evaluate the influence of individual pieces of research or even the prestige of researchers. This metric has recently come under considerable criticism owing to its inherent limitations and misuse.

The impact factor of a journal is a simple average obtained by considering the number of citations that articles in the journal have received within a specific time frame. A previous article "The impact factor and other measures of journal prestige" touched upon its calculation and features. This article delves a little deeper into the fallacies of the impact factor and points that you should consider when using it.

Below are listed some of the features and shortcomings of the JIF that should be well understood in order to prevent misuse of this metric.

The JIF is a measure of journal quality, not article quality. The JIF measures the number of citations accrued to all the articles in a journal, not to individual articles. Following the well-known 80-20 rule, the top 20% articles in a journal receive 80% of the journal’s total citations; this holds true even for the most reputed journals like Nature. So, an article published in a journal with a high JIF has not necessarily had high impact: it is very well possible that the article itself has not received any citations. Conversely, a few highly cited papers within a particular year can result in anomalous trends in a journal's impact factor over time.

How the JIF should be used	How the JIF should not be used
As a measure of journal prestige and impact	To evaluate the impact of individual articles and researchers
To compare the influence of journals within a specific subject area	To compare journals from different disciplines
By librarians, to manage institutional subscriptions	By funding agencies, as a basis for grant allocation
By researchers, to identify prestigious field-specific journals to follow and possibly submit to	By authors, as a singular criterion of consideration for journal selection
By journals, to compare expected and actual citation frequency and compare themselves with other journals within their field	By hiring and promotion committees, as a basis for predicting a researcher's standing
By publishers, to conduct market research	By authors, to compare themselves

Characteristics of the JIF

Only citations within a two-year time frame are considered. The JIF is calculated considering only those citations that a particular journal has received within 2 years prior. However, different fields exhibit variable citation patterns. While some fields such as health sciences receive most of their citations soon after publication, others such as social sciences garner most citations outside the two-year window. Thus, the true impact of papers cited later than the two-year window goes unnoticed.

The nature of the citation is ignored. As long as a paper in a journal has been cited, the citation contributes to the journal's impact factor, regardless of whether the cited paper is being credited or criticized. This means that papers being refuted or exemplified as weak studies can also augment a journal's impact factor. In fact, even papers that have been retracted can increase the impact factor because, unfortunately, citations to these papers cannot be retracted.

Only journals indexed in the source database are ranked. Thomson Reuters' Web of Science®, the source database for the calculation of the JIF, contains more than 12,000 titles. Although this figure is reasonably large and is updated annually, several journals, especially those not published in English, are left out. Thus, journals not indexed in Web of Science don't have an impact factor and cannot be compared with indexed journals.

The JIF varies depending on the article types within a journal. Review articles are generally cited more often than other types of articles because the former present a compilation of all earlier research. Thus, journals that publish review articles tend to have a higher impact factor.

The JIF is discipline dependent. The JIF should only be used to compare journals within a discipline, not across disciplines, as citation patterns vary widely across disciplines. For example, even the best journals in mathematics tend to have low IFs, whereas molecular biology journals have high IFs.

The data used for JIF calculations are not publicly available. The JIF is a product of Thomson Reuters®, a private company that is not obliged to disclose the underlying data and analytical methods. In general, other groups have not been able to predict or replicate the impact factor reports released by Thomson Reuters.8

The JIF can be manipulated. Editors can manipulate their journals' impact factor in various ways. To increase their JIF, they may publish more review articles, which attract a large number of citations, and stop publishing case reports, which are infrequently cited. Worse still, cases have come to light wherein journal editors have returned papers to authors, asking that more citations to articles within their journal—referred to as self-citations—be added.

These are some of the reasons you should not look at the JIF as a measure of research quality. It is important to explore other more relevant indicators for this purpose, possibly even in combination. If the JIF is used by a grant-funding body or your university, it might be a good idea to list your h index and citation counts for individual articles, in addition to the impact factors of journals in which you have published. This will help strengthen your argument on the quality and impact of your papers, regardless of the prestige of the journals you have published in.

Concluding remarks

Finally, remember that the nature of research is such that its impact may not be immediately apparent to the scientific community. Some of the most noteworthy scientific discoveries in history were recognized years later, sometimes even after the lifetime of the contributing researchers. No numerical metric can substitute actually reading a paper and/or trying to replicate an experiment to determine its true worth.

Bibliography

Garfield E (2006). The history and meaning of the journal impact factor, The Journal of the American Medical Association, 295: 90-93.

Brumback RA (2009). Impact factor wars: episode V–The empire strikes back. Journal of Child Neurology, 24: 260-262.

Brischoux F and Cook T (2009). Juniors seek an end to the impact factor race. Bioscience, 59: 238-239.

Rossner M, Epps HV, and Hill E (2007). Show me the data. The Journal of Cell Biology, 179: 1091-1092.

Adler R, Ewing J, Taylor P (2008). Citation Statistics. Joint Committee on Quantitative Assessment of Research, International Mathematical Union. [http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf]

Garfield E (2005). The agony and the ecstasy—the history and meaning of the journal impact factor. Presented at the International Congress on Peer Review and Biomedical Publication. [http://garfield.library.upenn.edu/papers/jifchicago2005.pdf]

Saha S, Saint S, and Christakis DA (2003). Impact factor: a valid measure of journal quality? Journal of the Medical Library Association, 91: 42-46.

Neylon C and Wu S (2009). Article-level metrics and the evolution of scientific impact. PLoS Biology, 7: 1-6.

Craig I (2007). Why are some journal impact factors anomalous? Publishing News. Wiley-Blackwell.

[http://blogs.wiley.com/publishingnews/2007/02/27/why-are-some-journal-impact-factors-anomalous/]

EASE. EASE statement on inappropriate use of impact factors. [http://www.ease.org.uk/publications/impact-factor-statement]

West R and Stenius K. To cite or not to cite? Use and abuse of citations. In: Publishing Addiction Science: A Guide for the Perplexed. Babor TF, Stenius K, and Savva S (eds). International Society of Addiction Journal Editors. [http://www.who.int/substance_abuse/publications/publishing_addiction_science_chapter4.pdf]

Katchburian E (2008). Publish or perish: a provocation. Sao Paulo Medical Journal, 202-203.

The PLoS Medicine Editors (2006). The impact factor game. PLoS Medicine 3(6): e291.

Smith L (1981). Citation analysis. Library Trends, 30: 83-106.

Sevinc A (2004). Manipulating impact factor: An unethical issue or an Editor's choice? Swiss Medical Weekly, 134:410.

12.

Why are some Journal Impact Factors Anomalous?

Iain Craig, Wiley Exchanges

- ◊ -

One of the consequences of the pareto-like distribution of citations amongst articles is that a small number of highly cited articles can exert an overwhelming effect on the Impact Factor. Large fluctuations in Impact Factor from year to year can often be explained by the inclusion of a small number of papers. However, what goes up, must come down, and once these exceptionally highly cited papers move out of the citation window, the Impact Factor often falls in an alarming fashion.

Evidence of this effect was described in an article published in Astronomische Nachrichten (H.A Abt, An anomalous journal impact factor, Astronomische Nachrichten 327 (7): 737-738 2006.http://dx.doi.org/10.1002/asna.200510622). This examined the change in Impact Factor of the journal Astrophysical Journal Supplement Series, which had jumped from a value of 6.247 in JCR 2003 to 15.231 in JCR 2004. The author investigated whether this was a computing error on behalf of Thomson Scientific who compile the Impact Factors, or whether there was some particularly unusual series of papers which have been exceptional highly cited.

It turned out that the massive increase in Impact Factor had been due to the inclusion of a set of 13 articles from a highly topical special issue dealing with the results from the first year's use of the Wilkinson Microwave Anisotropy Probe. These articles had an average of 28 times as many citations as other articles published in the journal.

As predicted by the author, the Impact Factor fell back somewhat in JCR 2005 (to 14.428), due to the fact that each year's Impact Factor is an average based on citations to two years of publications, and current indications suggest that the JCR 2006 value will be around 8.

Highly topical special issues can indeed have an exceptional effect on raising the Impact Factor, particularly in those areas where absolute number of citations and/or publications is low. However, this effect is likely to be short lived, unless real efforts are made to capitalize on the momentum that this spike in interest in the journal can bring.

13.

UNDERSTANDING JOURNAL IMPACT FACTORS

IAIN CRAIG, WILEY EXCHANGES

– ◊ –

It is an unfortunate truth that the Impact Factor is a measurement which can exert an undue influence on journal policies. Irrespective of the well documented limitations of the Impact Factor in establishing 'quality', a journal generally seeks to increase its score. When attempting to systematically improve the Impact Factor, it is sensible to first determine the current relative contribution of the articles published in the journal. Where possible, a comparison with a number of key journals in the field will assist in providing an indication as to the potential for an increase in impact. With that information in hand, more informed decisions can be made as to the direction of the future development of the journal.

Different article types, reviews, full-length articles, and short communications have different citation profiles, with differing proportions of total citations being received in the Impact Factor citation window. Determining an average number of citations per article type gives an indication as to which article types may lead to the most rapid increase in impact.

In addition to looking at the relative proportions of each of these articles types, it is also sensible to break down articles by subject area, and determine the relative impact of these different subject areas. This comparison is particularly useful when made for a number of journals in the field, and can identify strengths and weaknesses in current journal make-up.

Beyond looking at which article types or subject areas are generating the most citations, it is often instructive to look too at which areas are generating the fewest citations, and to identify the reasons behind this. Reducing the number of seldom cited articles that a journal publishes is a sure way of raising impact.

Finally, not withstanding the comments above, increasing the Impact Factor of a journal is not an exact science – the distribution of citations across articles is highly skewed, so caution is recommended when interpreting average citation scores, particularly where the number of articles and citations involved is small – and results from this systematic approach are likely to take several years to propagate through the measurement cycle. A common sense approach to development should be taken, combining analytical data as determined above, with the qualitative 'instinct' that only an expert in the subject area can possess.

14.

Do publication delays lead to artificially inflated Impact Factors?

Iain Craig, Wiley Exchanges

– ◊ –

Among the scores of papers published each month concerned with Impact Factors there are occasionally some interesting ones which pose new questions. Such a paper was published in PLOS ONE at the end of December 2012, on Hogmanay in fact: Rising Publication Delays Inflate Journal Impact Factors. The basic premise of the article is that articles which are published online prior to their compilation and publication in a print issue get a head start in accumulating citations. As the Impact Factor window is based on year of issue, publication articles/journals with a long delay between the initial article-based publication and issue publication are at a citation advantage compared with those which publish straight into issues, or which have a shorter lag between initial and issue publication. For a sample of 61 neuroscience journals the authors calculated a publication delay then re-calculated an Impact Factor-like value to take account of this, effectively removing the differences in publication delay. They observed that the values and rankings in their sample did change:

"We have demonstrated that the longer a paper is available as in press, the higher are its chances of getting cited in the 2-year window after it is published in print (presumably due to the fact that papers available online can be read and used by scientists more promptly and are thus cited earlier after publication), leading to an increase in the impact factor of journals with long online-to-print lags. Thus, it is at least feasible that increasing online-to-print lags might represent an active editorial policy to try to raise impact factors in some cases."

The question of whether there is a beneficial effect of initial article-based publication, what Wiley terms EarlyView, is one that has generated much discussion among Editors over many years. The PLOS ONE article provides some evidence that it is indeed beneficial, albeit in a simulated scenario. The observation that Editors may deliberately delay publication to build a buffer of articles outside of issues to allow them to mature in order to maximize their citation potential is concerning. While this could lead to increased Impact Factors, it is ethically dubious. As publishers our aim is rapid turnaround from EarlyView article to issue; anything else does the authors of that article a disservice.

Moving beyond the headlines, there are some open questions regarding the authors' calculation of the online to print delay used in their model, which will affect the reported magnitude of the effect. According to the Methods section, two dates were obtained from PubMed for each article and the difference between them determined. A schematic in their article describes this as the difference between Online Appearance and Publication in Print, and is what drives their calculations. I believe that there is a confusion between four different dates: the date of first online publication (such as in EarlyView), the date of online issue publication, the date of print issue publication, and the cover date of the issue. The authors' estimation of the publication lag between online and print is likely to be incorrect. It will sometimes measure the lag between online issue publication and cover date, and at other times, the time between EarlyView publication and cover date. The cover date is rarely a true date, rather it is an approximate date based on a desire to e.g. publish 12 issues throughout the year at regular intervals and to give them a notional month of publication. This, combined with the sometimes capricious nature of journal publication schedules, will lead to a calculated delay which is sometimes too high, sometimes too low. This throws into question the magnitude of the reported effects, although not the theory that there should be such an effect. Also, while the journal may benefit from extra citations in its JCR metrics, the probability/risk of a citation not being allocated to the individual article is higher for a citation to an EarlyView article than one to an article in an issue, due to a greater degree of ambiguity in the former. Citing authors are less certain about the citation form to use when citing EarlyView than when citing an article with a volume, issue, and page number. Citation matching by DOI is not the dominant mechanism by which citations are allocated, and any citations without a year, or which say 'In Press' are discarded from JCR metrics.

A second discussion point in the article was concerned with changing the methodology of the calculation of JCR metrics, from one based on print issue publication to one based on online article publication. The purpose of this is to remove any advantage that can be gained by having a long publication delay. Currently the year that counts for JCR metrics is the cover year of the issue, which is the same irrespective of whether the issue is in print or online (see end of post for more details). The authors' proposal is unclear as to whether by online publication they mean the actual date the article first appeared online (in any form), or the actual date of online issue publication. Assuming it is the former, this would necessitate a major re-work of the data collection approach for JCR metrics, and would be a fundamental departure, almost a philosophical one, for Thomson Reuters. Other newer metrics such as SNIP and SJR do match at an article-to-article level, but JCR metrics are based on allocation at a journal level.

In summary, is it a good idea to publish articles as EarlyView? Undeniably yes. The sooner the research is out there the better. There may be a benefit to the Impact Factor, but the jury is still out on the magnitude of this effect, although we can be reasonable confident that it will be greatest in areas where citations build quickly and where there are delays in EarlyView conversion. However, for an individual article, the longer the article is kept in EarlyView, the greater the risk that citations are not allocated to that article due to citation form ambiguity.

15.

Just say no to impact factors

Ismael Rafols and James Wilsdon

The Guardian, 17 May 2013

– ◊ –

Campaigners against the use of journal impact factors as a proxy for research excellence received a shot in the arm last night with the launch of the San Francisco Declaration on Research Assessment (DORA). With an impressive line-up of founding signatories, including individual scientists, research funders and journal editors, DORA states in no uncertain terms that journal impact factors (which rank journals by the average number of citations their articles receive over a given period) should not be used "as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contribution, or in hiring, promotion or funding decisions."

In an accompanying editorial in Science, Bruce Alberts, its editor-in-chief, condemns the misuse of journal impact factors as "highly destructive". He argues that they encourage gaming that can bias journals against publishing certain types of papers; waste time by overloading some journals with inappropriate submissions; and encourage "me-too" science in favour of more risky, potentially groundbreaking work.

This is by no means the first time that these arguments have been made. A number of UK-based scientists, including Stephen Curry, Dorothy Bishop andAthene Donald, have been prominent among those calling for an end to the crude application of impact factors. Similar arguments have long been made by scientometricians (who generate the metrics) and science policy researchers.

And yet, the rise of the impact factor has continued, propelled by a seemingly unstoppable managerial logic. Across UK universities, anyone involved in preparations for the forthcoming Research Excellence Framework (REF) cannot fail to be concerned by the fierce pressures that researchers now face to publish in particular disciplinary journals, and the way impact factors are being applied uncritically to determine who will and won't be submitted to the exercise (which in turn carries paramount weight in recruitment and promotion decisions).Claims by ministers and others that this isn't the case just don't stack up against the reality of what those of us in universities are now experiencing. DORA has come too late to halt the REF juggernaut for 2014. But it should prompt policymakers and the funding councils to pause and take stock before the next assessment cycle gets underway.

More broadly, the practice of journal-based evaluation has become deeply institutionalised, and it remains to be seen whether a declaration like this will be enough to change behaviour. Reflecting on his blog yesterday, Stephen Curry expressed the hope that DORA becomes a "landmark document". We share this hope, but like any well-meaning statement of intent, it's unclear how institutions, funders or individuals that continue to use metrics in assessment will be corrected or disciplined. Reputation is a crucial reward system in science and fraud is severely castigated: will it now become shameful to boast of the impact factors of one's publications? How will the organisations that supply metrics respond to the declaration, as they are torn between clients' demands for journal-based assessment and the rigorous handling of scientometric data?

DORA argues that articles and researchers should be judged on "their own merits" and emphasises that the "value of all research outputs should be considered", not just publications. One way to achieve this may be through greater use of altmetrics, which offer new insights into the impact of research. But even here we need to be conscious of the dangers of gaming and the difficulties of capturing some channels of impact.

For science to thrive and respond to societal challenges, diversity is key. Current practices based on journal-based metrics are a serious threat to this. One example of this loss of diversity, which alarms us as science policy researchers, is the suppression of interdisciplinarity. But there are many others, such as a lack of consideration being paid to certain research topics. This is why DORA is important and why it deserves widespread support.

16.

Scholars Seek Better Ways to Track Impact Online

Jennifer Howard

The Chronicle of Higher Education, January 29, 2012

– ◊ –

In academe, the game of how to win friends and influence people is serious business. Administrators and grant makers want proof that a researcher's work has life beyond the library or the lab.

But the current system of measuring scholarly influence doesn't reflect the way many researchers work in an environment driven more and more by the social Web. Research that used to take months or years to reach readers can now find them almost instantly via blogs and Twitter.

That kind of activity escapes traditional metrics like the impact factor, which indicates how often a journal is cited, not how its articles are really being consumed by readers.

An approach called altmetrics—short for alternative metrics—aims to measure Web-driven scholarly interactions, such as how often research is tweeted, blogged about, or bookmarked. "There's a gold mine of data that hasn't been harnessed yet about impact outside the traditional citation-based impact," says Dario Taraborelli, a senior research analyst with the Strategy Team at the Wikimedia Foundation and a proponent of the idea.

Interest in altmetrics is on the rise, but it's not quite right to call it a movement. The approach could better be described as a sprawling constellation of projects and like-minded people working at research institutions, libraries, and publishers.

They've been talking on Twitter (marking their messages with the #altmetrics hashtag), sharing resources and tools online, and developing ideas at occasional workshops and symposia. They're united by the idea that "metrics based on a diverse set of social sources could yield broader, richer, and timelier assessments of current and potential scholarly impact," as a call for contributions to a forthcoming altmetrics essay collection puts it.

Jason Priem, a third-year graduate student at the School of Information and Library Science at the University of North Carolina at Chapel Hill, is a leader in this push to track impact via the social Web. Scholarly workflows are moving online, leaving traces that can be documented—not just in articles but on social networks and reference sites such as Mendeley and Zotero, where researchers store and annotate scholarship of interest. "It's like we have a fresh snowfall across this docu-plain, and we have fresh footprints everywhere," he says. "That has the potential to really revolutionize how we measure impact."

Mr. Priem helped write a manifesto, posted on the Web site altmetrics.org, which articulates the problems with traditional evaluation schemes. "As the volume of academic literature explodes, scholars rely on filters to select the most relevant and significant sources from the rest," the manifesto argues. "Unfortunately, scholarship's three main filters for importance are failing."

Peer review "has served scholarship well" but has become slow and unwieldy and rewards conventional thinking. Citation-counting measures such as the h-index take too long to accumulate. And the impact factor of journals gets misapplied as a way to assess an individual researcher's performance, which it wasn't designed to do.

"I'm not down on citations," Mr. Priem says. "I'm just saying it's only part of the story. It's become the only part of the story we care about."

That's where altmetrics comes in. It's a way to measure the "downstream use" of research, says Cameron Neylon, a senior scientist at Britain's Science and Technology Facilities Council, and another contributor to the manifesto. Any system that turns out to be a useful way to measure influence will tempt the unscrupulous to try and game it, though. One concern is that someone could build a program, for instance, that would keep tweeting links to an article and inflate its altmetrics numbers.

Devising a Method

So how do you reliably measure fluid, fast-paced, Web-based, nonhierarchical reactions to scholarly work? That problem has been keeping Mr. Priem busy. He's part of the team that designed an altmetrics project called Total-Impact.

Researchers can go to the site and enter many forms of research, including blog posts, articles, data sets, and software they've written. Then the Total-Impact application will search the Internet for downloads, Twitter links, mentions in open-source software libraries, and other indicators that the work is being noticed. "We go out on the Web and find every sort of impact and present them to the user," Mr. Priem explains. When possible, they gather data directly from services' open-application programming interfaces, or API's.

These are very early days for Total-Impact, and there's a lot of information it doesn't gather yet. For instance, right now it only searches blogs indexed by the site Research Blogging. That "amounts to a very small subset of science blogs," according to Mr. Priem, who adds that most of the other metrics are more robust.

"Although it's still in alpha and has plenty of bugs, if you upload identifiers, you can and do get all sorts of impact information back," he says. "We've gotten many reports of people using the application, although certainly not in vast numbers" yet. "We've also gotten many requests from academic publishers and creators of scholarly Web applications to embed TI data into their pages" using Total-Impact's open API, he says.

He doesn't know yet how significant Total-Impact will prove to be. Will scholars take to it? Will tenure-and-promotion gatekeepers be willing to add altmetrics to the evaluation mix any time soon? Those are big unknowns right now. The long-term goal is "to completely change the way scholars and administrators think about academic impact" and get them to move away from what Mr. Priem calls "a citation-fetishizing article monoculture." But he's realistic. "Clearly, that's going to take some time," he says.

The Total-Impact site features several cautions about how it should and should not be used. It may help a researcher ascertain the "minimum impact" his or her work has made on the scholarly community; it can provide a sense of who's bookmarking or responding to that work. But it's not yet an indicator of comprehensive impact. "Take it all with a grain of salt," a warning on the site advises. "The meaning of these metrics are not yet well understood."

One of Mr. Priem's Total-Impact partners is Heather A. Piwowar. As a postdoctoral researcher at DataOne, affiliated with the National Evolutionary Synthesis Center and the Dryad digital repository, she studies patterns in how researchers share and reuse data. She and Mr. Priem have been building Total-Impact in their spare time. "Our day jobs are being a grad student and a postdoc," she says, but "we just couldn't stop ourselves. It seemed to have such profound possibilities."

The main difficulty they've encountered, she says, is finding sources of open data. Every blog post has a URL, and "you can search Twitter and other places for that URL," she says. But the Total-Impact algorithms can't just rely on Google searches, because those "aren't open and free data," she says. There's a lot of information behind the results of a Google search that Total-Impact can't really get to yet.

Another technical challenge for altmetrics is what to do about multiple digital "addresses" for a specific article online. Someone who tweets about a paper will probably link to a URL but not include the digital object identifier, or DOI, that makes the paper more permanently findable online, even if the URL changes. "So it's been more of a challenge than we expected to gather all of the synonym identifiers for an object and then search for all of them" in all the places where people might leave evidence of use, Ms. Piwowar says.

Right now, the Total-Impact group has to go ask Mendeley for an article's permanent Mendeley address, or "identifier," Pub-

Med for its identifier, and so on. "Having one place where a lot of these identifiers are aggregated would be very helpful," she says.

Software and data can be especially tricky to track. A piece of code may be hosted by an open repository like GitHub but not cited in ways that are easily recognized.

And scholarly culture doesn't always encourage openness. "There's a lack of reward for sharing data," Ms. Piwowar says.

Altmetrics' emphasis on openness aligns it with the open-access movement, whose goal is to make published research freely available online. "Once you see the potential for using the Web for research communication," says Britain's Mr. Neylon, it's hard to look at the traditional model of scholarly communication "without a growing sense of horror."

Altmetrics has made some inroads in the publishing world. For instance, one open-access publisher, the Public Library of Science, or PLoS, has been experimenting seriously with article-level metrics, a fresh way to measure who's using PLoS articles and how.

Unlike PLoS, however, many publishers are not keen to share usage statistics with the world. Neither are some institutional repositories.

Ms. Piwowar says that proprietary attitude is the wrong approach for publishers to take. Altmetrics "is a call to people who host research projects to make information about their impact openly accessible," she says.

Gaming the System

As its proponents themselves acknowledge, the altmetrics approach has vulnerabilities that go beyond how much data can be had for free. Just because an idea gets buzz online doesn't always mean it has genuine intellectual value, as anyone who follows social media knows. And what about gaming the system?

"Can Tweets Predict Citations?" asked a paper published last year in the Journal of Medical Internet Research by Gunther Eysenbach, a senior scientist and professor of health policy at the University of Toronto. Based on a survey he conducted, Dr. Eysenbach concluded that the answer is yes; tweets often do flag papers that turn out to be important.

But measures of influence on Twitter "should be primarily seen as metrics for social impact (buzz, attentiveness, or popularity) and as a tool for researchers, journal editors, journalists, and the general public to filter and identify hot topics," the researcher wrote. He cautioned that significant research in many fields wasn't necessarily going to get picked up by people who are on Twitter.

But traditional citations too have limitations, Dr. Eysenbach pointed out; social-media-based metrics should be considered complementary to citations rather than alternatives to them.

The key question might be how vulnerable altmetrics, or any metrics, is to being gamed. Traditional measures of influence aren't immune to corruption; journals have been known to drive up their impact factors by self-citing.

Mr. Taraborelli of the Wikimedia Foundation says "we should expect major attempts at gaming the system" if and when altmetrics really catches on. "My expectation is it will be an arms race," he says. But there are ways to build in safeguards against gaming, he says, much as people keep creating better spam filters.

The inclusive, diffuse approach that drives altmetrics may actually help protect it. A Godzilla-like monster ranking "is the best way to manipulate the system, to make it dependent on curation strategies that may end up invalidating the metric itself," Mr. Taraborelli says. "The last thing we want is a system that's dominated by a monolithic ranker for all the scholarly literature."

Researchers' behavior on the social Web works against the idea that one number should rule them all, Mr. Taraborelli says: *"I think we're moving to a system where, regardless of the benefits of single, monopolistic metrics, people will be able to set their own filters" to locate the research they're most interested in, wherever it lives.*

17.
ARE SCIENTISTS NEARSIGHTED GAMBLERS? THE MISLEADING NATURE OF IMPACT FACTORS

JULIEN MAYOR
BASQUE CENTER ON COGNITION, BRAIN AND LANGUAGE
FRONT. PSYCHOL., 06 DECEMBER 2010

– ◊ –

Despite a "Cambrian" explosion in the number of citation metrics used (Van Noorden, 2010), the impact factor (IF) of a journal remains a decisive factor of choice when publishing your ultimate research results and evaluating research productivity. Most other citation metrics correlate with the IF and there is little doubt that they reflect the overall impact of different journals. However, there is good reason to be more cautious about IF judgments.

First, the distribution of the number of citations per paper (NCPP) within a journal is heavily skewed. A few highly cited papers often account for a significant amount of the total citation count of a journal (25% of the papers in Nature account for 89% of the IF "Not-so-deep impact," 2005) and a recent report highlighted that even a single article can dramatically bias the IF of a small journal (Dimitrov et al., 2010). The mean NCPP, as captured with the IF, should therefore never be used. A more appropriate measure is the median NCPP. Figure 1 (left) plots the median of the total NCPP against the mean, for three potential publication outlets for psychologists; Psychological Review (IF2009 = 9.1), Nature (IF2009 = 34.5), and Psychological Science (IF2009 = 5.1), for different years (data compiled from ISI Web of Knowledge). Nature follows a distinct trend when compared to specialist journals: the median seems independent from its mean. This apparent dissociation results from the skew of the citation distribution observed in Nature, in which up to 35–40% of the published articles are never cited. Moreover, when using a robust metric for skewed distributions, Psychological Science's median is about seven times higher than Nature's even though its IF is about seven times lower than Nature's. This discrepancy even holds for specialized journals with very low impact factors; despite possessing an IF nearly 35 times lower than Nature's, the Journal of Child Language's median is higher than Nature's (data not shown).

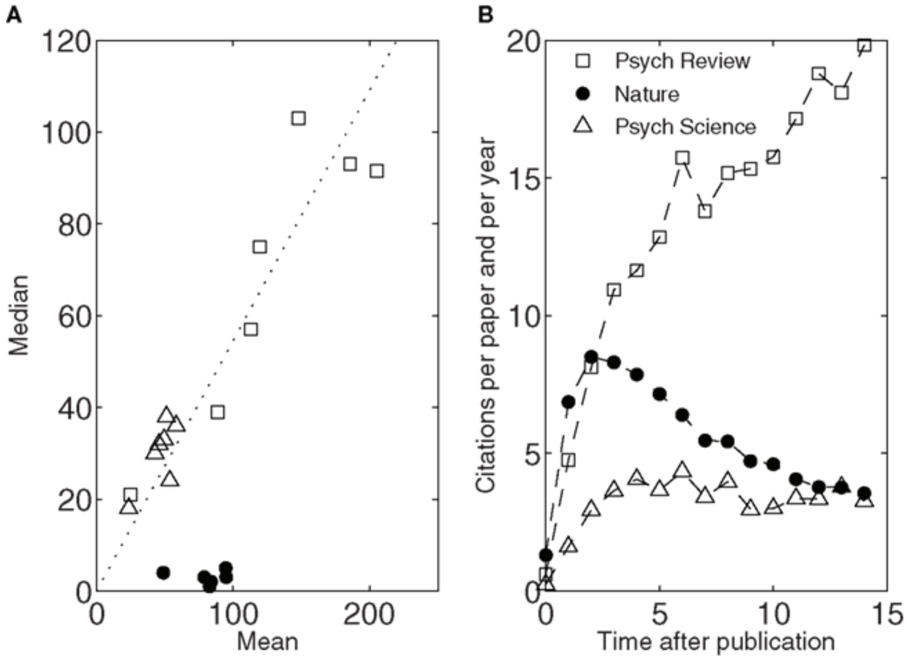


Figure 1. (A) Median number of citations per article as a function of the mean, for different years. (B) Time course of citations for articles published in 1995.

A second interesting property of the IF is that it focuses on citations of recently published articles only. For example, the 2009 IF of a journal considers the number of citations in 2009 to articles published in 2007 and 2008 only. However, the citations' time course differs dramatically from one journal to another. Figure 1 reports the mean NCPP per year, as a function of the number of years since publication. Once again, Nature possesses a distinct trend in its citation profile – the number of citations peaks 2 years after publication – whereas specialist journals have a steady increase in the NCPP. An article in Psychological Review will see its influence grow with age and would be outperformed by Nature when the IF monitors only short-lived citation patterns.

References

Dimitrov, J., Kaveri, S., and Bayry, J. (2010). Metrics: journal's impact factor skewed by a single paper. Nature 466, 179.

Not-so-deep impact. (2005). Nature 435, 1003–1004.

Van Noorden, R. (2010). Metrics: a profusion of measures. Nature 465, 864.

18.

Taking the Impact Factor seriously is similar to taking creationism, homeopathy or divining seriously

Bjoern Brembs
The Impact Blog, London School of Economics and Political Science

– ◊ –

I wasn't planning to write anything on Stephen Curry's latest piece on the negotiated, irreproducible and mathematically unsound Impact Factor sold by Thomson Reuters to gullible university administrators. I agree with most of what he writes there and, as he correctly cites a 20 year-old paper, all of it has been known for a decade or more. So why now pick it up anyway? First, apparently a lot of people are recommending the article, raising the suspicion that there may be some last refuges of scientists out there who are isolated from common knowledge. Second, he mentions a smear campaign against the IF, or rather shaming the use of IF. Of course, everybody who is using the IF immediately disqualifies themselves and by now everybody knows that. In fact, every scientist who is not aware of the unscientific nature of the IF should ask themselves if they are in the right profession. Taking the IF seriously is similar to taking creationism, homeopathy, or divining seriously. Stephen writes:

If you publish a journal that trumpets its impact factor in adverts or emails, you are statistically illiterate. (If you trumpet that impact factor to three decimal places, there is little hope for you.)

Which I thought was worth pointing out in the light of Nature Publishing Group's aggressive spamcampaign earlier this year touting their impact factors to the third decimal. NPG really is beyond hope, it seems.

The other reason I thought I should comment was a post on DrugMonkey's blog, where he writes that:

This notion that we need help "sifting" through the vast literature and that that help is to be provided by professional editors at Science and Nature who tell us what we need to pay attention to is nonsense. And acutely detrimental to the progress of science. I mean really. You are going to take a handful of journals and let them tell you (and your several hundred closest sub-field peers) what to work on? What is most important to pursue? Really? That isn't science. That's sheep herding. And guess what scientists? You are the sheep in this scenario.

I couldn't agree more. There is no evidence in the published literature that journal rank has any persuasive predictive property for any measure of scientific quality, be it expert review, citations, sound methodology, anything. On the contrary, there is solid evidence that journal rank is predicting article unreliability. Thus, the point needs to be made that there is no need for a replacement of IFs – we need to get rid of journals altogether as the concept of a journal is outdated and can be shown to be detrimental for science. This is an old idea and new data has only supported this insight. What we do need is a scientific metrics system that assists us in discovering, sorting and filtering the fraction of articles from the roughly 1.5 million published articles every year that are relevant to our research. There are currently many article-based metrics being developed which already today, with existing technology, can easily replace any journal-based metric for this task.

19.

SHOW ME THE DATA

MIKE ROSSNER, THE ROCKEFELLER UNIVERSITY PRESS
HEATHER VAN EPPS, THE *JOURNAL OF EXPERIMENTAL MEDICINE*
EMMA HILL, EXECUTIVE EDITOR, THE *JOURNAL OF CELL BIOLOGY*

– ◊ –

The integrity of data, and transparency about their acquisition, are vital to science. The impact factor data that are gathered and sold by Thomson Scientific (formerly the Institute of Scientific Information, or ISI) have a strong influence on the scientific community, affecting decisions on where to publish, whom to promote or hire (1), the success of grant applications (2), and even salary bonuses (3). Yet, members of the community seem to have little understanding of how impact factors are determined, and, to our knowledge, no one has independently audited the underlying data to validate their reliability.

Calculations and negotiations

The impact factor for a journal in a particular year is declared to be a measure of the average number of times a paper published in the previous two years was cited during the year in question. For example, the 2006 impact factor is the average number of times a paper published in 2004 or 2005 was cited in 2006. There are, however, some quirks about impact factor calculations that have been pointed out by others (e.g., 1, 4, 5), but which we think are worth reiterating here:

1. The numerator of the impact factor contains every detectable citation to a journal's content from the previous two years, regardless of the article type (6). For example, the 2006 impact factor numerator contains all citations to all content published in 2004 and 2005. The denominator of the impact factor,

however, contains only those articles designated by Thomson Scientific as primary research articles or review articles. Journal "front matter", such as Nature "News and Views" is not counted (4). Thus, the impact factor calculation contains citation values in the numerator for which there is no corresponding value in the denominator.

2. Articles are designated as primary, review, or "front matter" by hand by Thomson Scientific employees examining journals (6) using various bibliographic criteria, such as keywords and number of references (7).

3. Some publishers negotiate with Thomson Scientific to change these designations in their favor (5). The specifics of these negotiations are not available to the public, but one can't help but wonder what has occurred when a journal experiences a sudden jump in impact factor. For example, Current Biology had an impact factor of 7.00 in 2002 and 11.91 in 2003. The denominator somehow dropped from 1032 in 2002 to 634 in 2003, even though the overall number of articles published in the journal increased (see ISI Web of Science: http://portal.isiknowledge.com/, subscription required).

4. Citations to retracted articles are counted in the impact factor calculation (8). In a particularly egregious example, Woo Suk Hwang's stem cell papers in Science from 2004 and 2005, both subsequently retracted, have been cited a total of 419 times (as of November 20, 2007). We won't cite them again here to prevent the creation of even more citations to this work.

5. Because the impact factor calculation is a mean, it can be badly skewed by a "blockbuster" paper. For example, the initial human genome paper in Nature (9) has been cited a total of 5,904 times (as of November 20, 2007). In a self-analysis of their 2005 impact factor, Nature noted that 89% of their citations came from only 25% of the papers published (4).

When we asked Thomson Scientific if they would consider providing a median calculation in addition to the mean they already publish, they replied, "It's an interesting suggestion…The median…would typically be much lower than the mean. There are other statistical measures to describe the nature of the citation frequency distribution skewness, but the median is probably not the right choice." Perhaps so, but it can't hurt to provide the community with measures other than the mean, which, by Thomson Scientific's own admission, is a poor reflection of the average number of citations gleaned by most papers.

6. There are ways of playing the impact factor game, known very well by all journal editors, but played by only some of them. For example, review articles typically garner many citations, as do genome or other "data-heavy" articles (see example above). When asked if they would be willing to provide a calculation for primary research papers only, Thomson Scientific did not respond.

Integrity

As journal editors, data integrity means that data presented to the public accurately reflect what was actually observed. To help ensure this, The Rockefeller University Press instituted a policy of scrutinizing image data in accepted manuscripts for evidence of manipulation. We realize that image data is only one type of data we publish, but it is a type that can be easily examined for integrity. If a question is raised about the data in a figure, we ask the authors to submit the original data for examination by the editors. We consider it our obligation to protect the published record in this way.

Thomson Scientific makes its data for individual journals available for purchase. With the aim of dissecting the data to determine which topics were being highly cited and which were not, we decided to buy the data for our three journals (The Journal of Experimental Medicine, The Journal of Cell Biology, and The Journal of General Physiology) and for some of our direct competitor journals. Our intention was not to question the integrity of their data.

When we examined the data in the Thomson Scientific database, two things quickly became evident: first, there were numerous incorrect article-type designations. Many articles that we consider "front matter" were included in the denominator. This was true for all the journals we examined. Second, the numbers did not add up. The total number of citations for each journal was substantially fewer than the number published on the Thomson Scientific, Journal Citation Reports (JCR) website (http://portal.isiknowledge.com, subscription required). The difference in citation numbers was as high as 19% for a given journal, and the impact factor rankings of several journals were affected when the calculation was done using the purchased data (data not shown due to restrictions of the license agreement with Thomson Scientific).

Your database or mine?

When queried about the discrepancy, Thomson Scientific explained that they have two separate databases—one for their "Research Group" and one used for the published impact factors (the JCR). We had been sold the database from the "Research Group", which has fewer citations in it because the data have been vetted for erroneous records. "The JCR staff matches citations to journal titles, whereas the Research Services Group matches citations to individual articles", explained a Thomson Scientific representative. "Because some cited references are in error in terms of volume or page number, name of first author, and other data, these are missed by the Research Services Group."

When we requested the database used to calculate the published impact factors (i.e., including the erroneous records), Thomson Scientific sent us a second database. But these data still did not match the published impact factor data. This database appeared to have been assembled in an ad hoc manner to create a facsimile of the published data that might appease us. It did not.

Opaque data

It became clear that Thomson Scientific could not or (for some as yet unexplained reason) would not sell us the data used to calculate their published impact factor. If an author is unable to produce original data to verify a figure in one of our papers, we revoke the acceptance of the paper. We hope this account will convince some scientists and funding organizations to revoke their acceptance of impact factors as an accurate representation of the quality—or impact—of a paper published in a given journal.

Just as scientists would not accept the findings in a scientific paper without seeing the primary data, so should they not rely on Thomson Scientific's impact factor, which is based on hidden data. As more publication and citation data become available to the public through services like PubMed, PubMed Central, and Google Scholar®, we hope that people will begin to develop their own metrics for assessing scientific quality rather than rely on an ill-defined and manifestly unscientific number.

References

1. Monastersky, R. 2005. The number that's devouring science. The impact factor, once a simple way to rank scientific journals, has become an unyielding yardstick for hiring, tenure, and grants. Chron. High. Educ.52:A12.

2. Wells, W.A. 2007. The returning tide: how China, the world's most populous country, is building a competitive research base. J. Cell Biol.176:376–401 doi:10.1083/jcb.200701150.

3. Editorial. 2006. Cash-per-publication is an idea best avoided. Nature. 441:786 doi:10.1038/441786b.

4. Editorial. 2005. Not-so-deep impact. Research assessment rests too heavily on the inflated status of the impact factor. Nature.435:1003–1004 doi:10.1038/4351003b.

5. The PLoS Medicine Editors. 2006. The impact factor game. It is time to find a better way to assess the scientific literature.

PLoS Med.3:e291 doi:10.1371/journal.pmed.0030291.

6. Garfield, E. 1999. Journal impact factor: a brief review. Can. Med. Assoc. J.161:979–980.

7. The Thomson Scientific Impact Factor.

http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/(accessed November 29, 2007).

8. Liu, S.V. 2007. Hwang's retracted publication still contributes to Science's impact factor. Sci. Ethics.2:44–45.

9. Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al. 2001. Initial sequencing and analysis of the human genome. Nature.409:860–921.

20.

The Impact Factor Game

The *PLoS Medicine* Editors

PLoS Med. Jun 2006; 3(6): e291.

– ◊ –

We would be lying if we said that our journal's impending first impact factor is not of interest to us. What-PLoS Medicine's impact factor might be is certainly one of the questions that crops up most regularly in discussions with authors, and because our authors' opinions matter to us, we are obliged to take it seriously. However, for a number that is so widely used and abused, it is surprising how few people understand how a journal's impact factor is calculated, and, more importantly, just how limited it is a means of assessing the true impact of an individual publication in that journal.

A journal's impact factor is calculated from this equation:

Journal X's 2005 impact factor =

Citations in 2005 (in journals indexed by Thomson Scientific [formerly known as Thomson ISI]) to all articles published by Journal X in 2003–2004

divided by

Number of articles deemed to be "citable" by Thomson Scientific that were published in Journal X in 2003–2004

What is obvious from this equation is that the impact factor *depends crucially on which article types Thomson Scientific deems as "citable"*—the fewer, the better (i.e., the lower the denominator, the higher the impact factor).

Because a journal's impact factor is derived from citations to all articles in a journal, this number cannot tell us anything about the quality of any specific research article in that journal, nor of the quality of the work of any specific author. These points become particularly evident by understanding that a journal's impact factor can be substantially affected by the publication of review articles (which usually acquire more citations than research articles) or the publication of just a few very highly cited research papers.

Moreover, a journal's impact factor says nothing at all about how well read and discussed the journal is outside the core scientific community or whether it influences health policy. For a journal such as PLoS Medicine, which strives to make our open-access content reach the widest possible audience—such as patients, health policy makers, non-governmental organizations, and school teachers—impact factor is a poor measure of overall impact.

Despite these evident limitations, the impact factors of journals that authors publish in are very influential. Although even Thomson Scientific acknowledges that the impact factor has grown beyond its control and is being used in many inappropriate ways, the impact factors of journals have been used to decide whether or not authors get promoted, are given tenure or are offered a position in a department, or are awarded a grant. In some countries, government funding of entire institutions is dependent on the number of publications in journals with high impact factors.

Small wonder, then, that authors care so much about journals' impact factors and take them into consideration when submitting papers. Should we, as the editors of PLoS Medicine, also care about our impact factor and do all we can to increase it? This is not a theoretical question; it is well known that editors at many journals plan and implement strategies to massage their impact factors. Such strategies include attempting to increase the numerator in the above equation by encouraging authors to cite articles published in the journal or by publishing reviews that will garner large numbers of citations. Alternatively, editors may decrease the denominator by attempting to have whole article types removed from it (by making such articles superficially less substantial, such as by forcing authors to cut down on the number of references or removing abstracts) or by decreasing the number of research articles published. These are just a few of the many ways of "playing the impact factor game."

One problem with this game, leaving aside the ethics of it, is that the rules are unclear—editors can, for example, try to persuade Thomson Scientific to reduce the denominator, but the company refuses to make public its process for choosing "citable" article types. Thomson Scientific, the sole arbiter of the impact factor game, is part of The Thomson Corporation, a for-profit organization that is responsible primarily to its shareholders. It has no obligation to be accountable to any of the stakeholders who care most about the impact factor—the authors and readers of scientific research. Although we have not attempted to play this game, we did, because of the value that authors place on it, attempt to understand the rules. During discussions with Thomson Scientific over which article types in PLoS Medicine the company deems as "citable," it became clear that the process of determining a journal's impact factor is unscientific and arbitrary. After one in-person meeting, a telephone conversation, and a flurry of e-mail exchanges, we came to realize that Thomson Scientific has no explicit process for deciding which articles other than original research articles it deems as citable. We conclude that science is currently rated by a process that is itself unscientific, subjective, and secretive.

During the course of our discussions with Thompson Scientific, PLoS Medicine's potential impact factor—based on the same articles published in the same year—seesawed between as much as 11 (when only research articles are entered into the denominator) to less than 3 (when almost all article types in the magazine section are included, as Thomson Scientific had initially done—wrongly, we argued, when comparing such article types with comparable ones published by other medical journals). At the time of writing this editorial, we do not know exactly where our 2005 impact factor has settled. But whatever it turns out to be, as you might guess from this editorial, we feel the time has come for the process of "deciding" a journal's impact factor to be debated openly. Something that affects so many people's careers and the future of departments and institutions cannot be kept a secret any longer.

Even more importantly, it is time to reconsider the whole process of accurately assessing an individual paper's worth not only to scientists, but also to the wider community of readers. First, although any measure of impact will remain flawed in some way, when assessing the impact of individual articles or of the papers of individuals or groups of scientists, it surely makes more sense to measure the citations specifically to those individual articles (or to papers by individuals or groups of scientists) rather than using a journal's impact factor as a proxy measure. However, it is not clear whether Thomson Scientific could measure such individual article citations accurately. Second, we urge the company to take its responsibility seriously and increase transparency and accountability. Third, we suggest that the company's staff engage in the ongoing debate among other shareholders of scientific publishing and recognize that, there are—finally—other ways of measuring impact and visibility of scholarly articles. Thomson Scientific now faces competition from organizations that have developed online tools for citation counting, such as Google Scholar and CrossRef, and this competition may help to bring about overdue change. Other measures of scientific impact may also become widely adopted, such as the usage factor, which is being promoted by the United Kingdom Serials Group (http://www.uksg.org/rfp.pdf), or the Y factor, a combination of both the impact factor and the weighted page rank, developed by Google (http://www.soe.ucsc.edu/~okram/papers/journal-status.pdf).

These new measures may go some way to helping assess and perhaps quantify the many roles that medical journals have, in a way that measuring citations only to research articles cannot. Magazine sections, such as those that we and other medical journals publish, not only “add value” to the research articles by interpreting them for a wider audience but have other vital roles: they may help to set agendas—by publishing policy papers or highlighting neglected health issues; give underrepresented groups, such as medical students or patient groups, a voice; or provide educational materials to physicians. Such articles will rarely be cited in indexed journals, but may be influential, for example, in changing health policy, or may be of educational value. For such articles, more relevant measures of impact may be the number of times they are downloaded, or covered in news articles, or referenced in policy documents.

Perhaps even measures such as these will become outmoded as the Internet allows for users to interact more directly with published articles. Journals have taken a step toward such a future with the publication of e-letters, and the physics preprint server arXiv.org has been promoting such interaction for many years. As more and more articles are available in full electronically and as search engines get more sophisticated at mining the Web and assessing usage, such interaction with the literature will become easier and readers will be able to judge papers for themselves rather than relying on outmoded surrogates for quality such as the impact factor. If authors are going to quote the impact factor of a journal, they should understand what it can and cannot measure. The opening up of the literature means that better ways of assessing papers and journals are coming—and we should embrace them.

21.

Not-so-deep impact

Editorial, Nature 435, 1003-1004 (23 June 2005)
doi:10.1038/4351003b; Published online 22 June 2005

– ◊ –

Every year at the end of June, scientific publishers' eyes turn to Philadelphia, where the Institute for Scientific Information (ISI) releases a snippet of data that they crave: the impact factor of each journal. In due course, bureaucrats in research agencies will roll the impact figures into their performance indicators, and those scientists who worry about such things will quietly note which journal's number wins them the most brownie points.

Attempts to quantify the quality of science are always fraught with difficulty, and the journal impact factors are among the few numbers to persist. The result is an overemphasis of what is really a limited metric.

To obtain the latest impact factors, which were released last week, the ISI number-crunchers added the total number of citations from all the monitored journals during 2004 to items in the journal of interest that were published in 2002 and 2003. They then divided that total by the number of ‘citable items’ — loosely, papers and review articles — that were published in the journal during those same two years.

The impact factor is taken by some administrators as a measure of the typical citation rate for the journal. But for many journals, it isn't ‘typical’ at all.Nature's latest impact factor is 32.2, an increase on last year and a high number that we're proud of, but it's one that merits a closer look.

For example, we have analysed the citations of individual papers in Nature and found that 89% of last year's figure was generated by just 25% of our papers.

The most cited Nature paper from 2002–03 was the mouse genome, published in December 2002. That paper represents the culmination of a great enterprise, but is inevitably an important point of reference rather than an expression of unusually deep mechanistic insight. So far it has received more than 1,000 citations. Within the measurement year of 2004 alone, it received 522 citations. Our next most cited paper from 2002–03 (concerning the functional organization of the yeast proteome) received 351 citations that year. Only 50 out of the roughly 1,800 citable items published in those two years received more than 100 citations in 2004. The great majority of our papers received fewer than 20 citations.

These figures all reflect just how strongly the impact factor is influenced by a small minority of papers — no doubt to a lesser extent in more specialized journals, but significantly nevertheless. However, we are just as satisfied with the value of our papers in the 'long tail' as with that of the more highly cited work.

The citation rate of our papers also varies sharply between disciplines. Many of Nature's papers in immunology published in 2003 have since received between 50 and 200 citations. Significant proportions of those in cancer and molecular and cell biology have been in the 50–150 range. But papers in physics, palaeontology and climatology typically achieved fewer than 50 citations. Clearly, these reflect differences in disciplinary dynamics, not in quality.

The impact factor also mixes citations to diverse types of content: unsurprisingly, review articles are typically the most highly cited, but citations of our Commentaries, News Features and News & Views articles also contribute in a minor way to the numerator (although these items are not counted in the denominator).

The net result of all these variables is a conclusion that impact factors don't tell us as much as some people may think about the respective quality of the science that journals are publishing.

Neither do most scientists judge journals using such statistics; they rely instead on their own assessment of what they actually read.

None of this would really matter very much, were it not for the unhealthy reliance on impact factors by administrators and researchers' employers worldwide to assess the scientific quality of nations and institutions, and often even to judge individuals. There is no doubt that impact factors are here to stay. But these figures illustrate why they should be handled with caution.

22.

Journal accused of manipulating impact factor

Richard Smith, BMJ

BMJ Volume 314 15 February 1997

– ◊ –

The journal Leukemia, which is owned by Macmillan magazines, has been accused of trying to manipulate its impact factor, the measure used to rank the importance of scientific journals. The accusation comes from Terry Hamblin, consultant haematologist at the Royal Bournemouth Hospital and editor of Leukemia Research, a rival to Leukemia.

Dr Hamblin has sent the BMJ a copy of a letter received by authors who had submitted a paper to Leukemia in October 1996 asking them to increase the number of references to papers published in Leukemia. This would increase the journal's impact factor, which is calculated by dividing the number of citations of papers in the journal by the number of papers that could be cited. The impact factor has become much more important in recent years because many countries consider the impact factors of the journals in which researchers publish when judging the researchers and making decisions about future funding (p 498).

The letter from Leukemia said: "Manuscripts that have been published in Leukemia are too frequently ignored in the reference list of newly submitted manuscripts, even though they may be extremely relevant. As we all know, the scientific community can suffer from selective memory when giving credit to colleagues. While we have little power over other journals, we can at least start by giving you and others proper credit in Leukemia. We have noticed that you cite Leukemia [once in 42 references]. Consequently, we kindly ask you to add references of articles published in Leukemia to your present article."

"This is a blatant attempt to increase the journal's impact factor," said Dr Hamblin. "I accept that authors sometimes do not cite relevant papers, but I have never encountered a journal that specifically requested an increase in the number of times that journal is cited in the bibliography."

Dr Nicole Muller-Bérat, the editor of Leukemia, denies that the journal is trying to manipulate its impact factor. "We introduced the policy of asking people to cite Leukemia for two main reasons. Firstly, we have received, and published, letters from authors saying that papers we have published have neglected to cite important papers published in Leukemia. Secondly, our reviewers remember important papers published in major journals like Blood, Cell, and the British Journal of Haematology, but they forget about important papers published in Leukemia."

Dr Muller-Bérat believes that Dr Hamblin is motivated to make his accusation by professional jealousy. She founded Leukemia Research, the journal she edits, with her husband, but he became the editor in 1986. Since then the impact factor has fallen from 2.7 to 1.179. She and her husband also founded Leukemia in 1987, and by 1991 it had an impact factor of 3.059. Following changes in the editorial team, the impact factor fell to 1.7 but has now risen to 2.35.

David Pendlebury, an analyst at the Institute of Scientific Information in Philadelphia, which calculates journals' impact factors, said: "We have never heard of a case like this before. It is a distortion of the scientific process." Richard Horton, editor of the Lancet, said: "Given the importance attached to impact factors this manipulation seems an appalling lapse of editorial judgment."

23.

The Impact Factor and Using It Correctly

Eugene Garfield

Der Unfallchirurg, 48(2) p.413, June 1998.

– ◊ –

I would like to comment on the statement by Prof. H-J. Oestern and Prof. J. Probst of the German Trauma Society which appeared in the October 1997 issues of your journal as well as the Anaesthesist. The authors assert that the work of German specialists in these fields is published "primarily" in German language journals. And without any supporting data they assert that the impact factor is not appropriate for judging scientific achievements in trauma surgery and most important that "its use leads to an unjustified disadvantage in comparison with other fields."

Where has it been written that such comparisons should be made? In a series of unrelated assertions, none of which are supported by data, claims of bias are made repeatedly. Would these authors assert that German scientists, even in trauma surgery, do not publish in the international journals? In 1997, scientists from Germany published over 77,000 papers in Science Citation Index covered journals – almost 7.8% of the total ISI database. About 12,000 of those articles were published in German.

A detailed analysis and definition of the field of trauma surgery is necessary in order to determine the percentage of papers from each country and their relative performance. An agreed upon list of cohort source journals is essential.

The authors talk about self-citations, but there is no evidence that this changes impact factor. In fact, for the smaller journals, self-citations may be the major source of citations, especially when they are new.

The authors allege that there is discrimination against other unspecified journals. Are they aware that all references, regardless of the journals in which they were published, are included in the SCI data?

Unfallchirurg and other Springer journals are included in SCI. Which are the journals that are missing whose citations might improve their impact? In fact, as we add more low-impact journals we find that they cite the high-impact journals heavily and thereby increase their already high impact. But these generalizations need to be supported by specific studies in each field.

The idea that non-English language journals do not have a chance to be cited is untrue. The SCI processes all references regardless of the journal cited, with the exception of those in exotic alphabets such as Chinese and Japanese. German scientists who publish in English-language journals are never forbidden to cite relevant work in German. One might argue that English-speaking readers may not read the original German journals, but today the use of English titles and abstracts means that few important articles are missed. But even that in itself does not guarantee citation. Authors cite one another because they become familiar with their work through many channels including international meetings and educational exchanges. And hopefully good refereeing keeps most references relevant.

It is absurd to make invidious comparisons between specialist journals and multi-disciplinary general journals like Nature and NEJM. To compare journals you should stick to a particular category as is explained very carefully in the Guide to Journal Citation Reports. Incidentally, the anomaly of the old journal Clinical Research is explained in a footnote on page 7 of the JCR printed guide. That journal primarily contains meeting abstracts and its title has been changed to Journal of Investigative Medicine.

The source of much anxiety about Journal Impact Factors comes from their misuse in evaluating individuals, e.g. during the Habilitation process. In many countries in Europe, I have found that in order to shortcut the work of looking up actual (real) citation counts for investigators the journal impact factor is used as a surrogate to estimate the count. I have always warned against this use. There is wide variation from article to article within a single journal as has been widely documented by Per O. Seglen of Norway and others.

These questions will be discussed in Oslo in April by myself and Dr. Seglen and then in May at the Conference of Biology Editors in Salt Lake City. All editors are all welcome to attend.

For detailed information on over 8,000 journals, ISIs Journal Performance Indicators can be a valuable source of data for comparing journals. These ISI databases are available through David Pendlebury at ISI (215-386-0100, x1411). Long-term impact factors rather than current impact may be more appropriate in certain clinical fields.

Citation data and analysis should always be used in combination with other indicators when evaluating departments or individuals. For nation by nation comparisons, there is very little controversy about the use of citation indicators. Further, they have been used in the USA to evaluate 5,000 departments at the leading universities. Similar research assessment exercises are performed in the UK.

To test the validity of the ISI data you should identify a cohort of experts in trauma surgery and see how their citation records compare. This would augment an article-by-article citation audit of articles published in your journal. This can be done by contacting Pendlebury or it can be done by use of the various public online vendors of citation indexes or the Web of Science.

PS The following reference is of possible interest since it was written by a German scientist and demonstrates that even journals not included in SCI can have their impacts calculated.

24.

Impact factors—a critique

J. J. Ramsden

Journal of Biological Physics and Chemistry 9 (2009) 139-140

– ◊ –

The so-called impact factor (IF) started as a rather esoteric scientometric tool, itself derived from the science citation index (SCI), which originated with Eugene Garfield in the 1950s. The IF of a journal for a particular year is defined as the quotient of the number of citations received in that year by papers published in that journal in the two preceding years and the number of "citable" papers published in those two years.

Had the IF remained, as might have been reasonable to expect, a specialist statistic mainly of interest to information professionals as represented by librarians and others there would be no need to expend energy on criticizing it. Unfortunately its use—or rather misuse—has vastly expanded in recent years, which makes criticism very necessary.

Following the failure to secure public funding for the establishment of the SCI, it was launched as a commercial venture in 1961 as the Institute for Scientific Information (ISI). In 1992 the ISI was sold to the Thomson Corporation, well known as the publisher of newspapers such as the Sunday Times. Thomson Corporation merged with Reuters in 2008 to form Thomson Reuters, which now claims to be "the world's leading source of intelligent information (sic) for businesses and professionals".

The IF has a number of technical defects. Some of the more obvious ones are:

(i) the IF of journals covering a broad area of science with a rapidly growing but ephemeral literature that tends to cite many articles will inevitably be higher than the IF of more specialist journals whose articles may reach peak citation many years after their publication;

(ii) review articles tend to be much more heavily cited than primary research articles and hence journals carrying some review articles, and especially journals exclusively devoted to them, will have high IF—it seems absurd to include them both in the same statistic;

(iii) as a cursory check will quickly verify, the distribution of citations received by individual articles in a journal is so broad as to make the mean almost meaningless. Technical reports describing a new methodology and "data-rich" papers such as the articles reporting the human genome sequence tend to be very heavily cited. Indeed, it has been admitted by Thomson that the median is very different from the mean [1] without, however, this admission having been followed by revealing the actual distributions. In one field at least, high energy physics, analysis of citations has revealed a power law distribution with at least two exponents [2], hence an ill-defined variance.

Some of the less obvious technical defects are:

(i)inconsistency with respect to papers being cited and receiving citations. The citable papers in the denominator contain only papers categorized as primary research articles or review articles; "news and views" articles are excluded [1]. On the other hand, the numerator contains all citations, even when they are made by such news and views articles. Furthermore, categorization of article type is made by Thomson Reuters staff according to a procedure not publicly available and that moreover appears to be subject to commercial negotiation [1];

(ii) only journals included in the Thomson Reuters database are considered. The database is dominated by North American publications [3]. This not only means that many important journals not in the database do not even receive impact factors, but also that citations in journals outside the database, and in other sources such as books, are not counted at all. Furthermore, the database changes from year to year and the criterion or criteria for including a journal are not publicly available; all one knows is that inclusion depends on some subjective factors enshrined within the company's policy, such as a desire to promote a particular field, much as a newspaper proprietor typically imposes a certain party line on editorial policy.

Because of the extreme tediousness of counting citations, it has been practically impossible to check the accuracy of the published citation figures. In the very rare cases where this has been done, alarming discrepancies have been uncovered [1]. Perhaps the recent general availability of powerful data searching services such as Google Scholar, which make it practicable for individual scientometricians and even scientists interested in the matter to develop and calculate their own indices, will enable more reliable citation metrics based on open and transparent procedures to be developed. It should, however, be appreciated that this is a tricky matter indeed. Some of the alternatives that have already been proposed, such as the h- and m-indices (Hirsch), the gindex (Egghe) and the w-index (Woeginger) have also been subject to severe criticism [4].

There are also defects of a semantic nature. Garfield himself wrote "the citation index...may help a historian to measure the influence of the article—that is, its 'impact factor'." [5]. This vagueness regarding the link between citations (i.e., the IF) and actual impact has not been made more precise by subsequent pronouncements (see [4]). No attempt appears to be made to distinguish between positive and negative influence. Another defect of a semantic nature is the astonishing fact that citations to retracted articles are counted [1]! This is all the more regrettable because the scientist but not the bureaucrat knows that retraction is almost always a consequence of the perpetration of fraud having been uncovered.

It would have been more accurate to have called the IF "popularity factor". If one accepts the so-called IF as a measure of quality, one must also accept that, in the UK, the Sun is a higher-quality daily newspaper than the Financial Times, or (in Switzerland) that Blick is of a higher quality than the Neue Zürcher Zeitung.

It should also be recalled that different fields of science have very different citation rates. The mean IF of mathematics journals in the Thomson Reuters database is less than one and of physics journals is less than two; chemistry fares somewhat better with a mean IF approaching three and biology almost attains a value of five. This of course makes comparisons between different fields meaningless. Popular magazines like Nature and Science that combine the functions of a weekly news bulletin and a scientific journal covering many fields have much higher IF (around 30), presumably because they are widely read and the scholarly practice of only citing what one has read still largely holds sway.

However nebulous the meaning of impact in the term "impact factor", undoubtedly the IF has acquired a great impact, especially in connexion with the bureaucratic controls and financial restrictions imposed on scientists. In these contexts the IF is recklessly used by officials not in a position to appreciate those defects that make it quite unsuitable for the purpose of assessing research quality. As the IMU report eloquently states, "Much of modern bibliometrics seems to rely on experience and intuition about the interpretation and validity of citation statistics." [4]. One can indeed admire Garfield's entrepreneurial success in having turned an obscure, specialist and flawed metric into a highly profitable business. Given that Thomson Reuters is a commercial organization with the goal of maximizing profits it should not be expected that it should follow standards of scholarly rigour. What is surprising is that the majority of scientists themselves appear to have acquiesced in the general misuse of the IF [6].

For the reasons discussed above, most of which have already been amply pointed out in the literature, the IF is incompatible with the scientific method. Acceptance of the IF by a scientist must therefore cast doubt on his or her professional integrity. Long ago the then Director of Research of the Institute for Scientific Information pointed out that "the SCI would work perfectly if every author meticulously cited only the earlier work related to his theme; if it covered every scientific journal published anywhere in the world; and if it were free from economic constraints." [7].

It was already realized then that these conditions were not fulfilled—hence the SCI on which the IF is based is flawed even at the most basic level of the scientist choosing which papers to cite.

References

1. Rossner, M., Van Epps, H. and Hill, E. Show me the data. J. Cell Biol. 179 (2007) 1091–1092.
2. Lehmann, S., Lautrup, B. and Jackson, A.D. Citation networks in high energy physics. Phys. Rev. E 68 (2003) 026113.
3. Seglen, P.O. Why the impact factor of journals should not be used for evaluating research. Br. Med. J. 314 (1997) 497.
4. Adler, R., Ewing, J. and Taylor, P. Citation Statistics. Washington: International Mathematical Union (IMU) (Joint Committee on Quantitative Assessment of Research) (2008).
5. Garfield, E. Citation indexes for science: A new dimension in documentation through association of ideas. Science 122 (1955) 108–111.
6. Not only merely passively, but sometimes to the extent of going to great lengths to manipulate citations—see Z. Corbyn, Times Higher Education, 7 February 2008, p. 6.
7. Cawkell, A.E. Science perceived through the Science Citation Index. Endeavour (new ser.) 1 (1977) 57–62.

25.

Does Criticism Overcome the Praises of Journal Impact Factor?

Masood Fooladi et al.

Asian Social Science; Vol. 9, No. 5; 2013

– ◊ –

Citation is the reference text and recognition of article information. Generally, each article has a reference section at the end including all references cited in that article. Each reference is called a citation. The frequency of citations to an especial article by other articles means citation count. In addition, citation index as a type of bibliographic database traces the references in a published article. Citation index shows that how many times an article has been cited with other articles. In other words, one can find that which later article cites which earlier articles. Citations are applied to measure the importance of information contained in an article and the effect of a journal in related area of research activity and publications.

The first citation index for published articles in scientific journals is Science Citation Index (SCI), which is founded by Eugene Garfield's Institute for Scientific Information (ISI, previously known as Eugene Garfield Associates Inc.) in 1960 (Nigam & Nigam, 2012.). Then, it was developed to produce the Arts and Humanities Citation Index (AHCI) and the Social Sciences Citation Index (SSCI). The SSCI is one of the first databases developed on the Dialog system in 1972. Small (1973) develops his efforts on Co-Citation analysis as a self-organizing classification mechanism namely "Research Reviews". Garner, Lunin, and Baker (1967) describes that the worldwide citation network has graphical nature. Giles, Bollacker, and Lawrence (1998) introduce the autonomous citation indexing, which can provide the automatic algorithmic extraction and classify the citations for any digital scientific and academic articles.

Subsequently, several citation indexing services are created to automate citation indexing including Google Scholar, EBSCOhost, Institute for Scientific Information (ISI) and Elsevier (Scopus).

Among all citation indexing services, ISI, has dominated the citation indexing career. ISI is a part of Thomson Reuters, which publishes the citation indexes in form of compact discs and print format. One can access to the products of ISI through the website with the name 'Web of Science' (WOS). It is possible to access seven databases in WOS including: Social Sciences Citation Index (SSCI), Index Chemicus, Science Citation Index (SCI), Arts & Humanities Citation Index (A&HCI), Current Chemical Reactions, Conference Proceedings Citation Index: Science and Conference Proceedings Citation Index: Social Science and Humanities.

2. Impact Factor

One of the most important product of ISI is the journal impact factor (IF), which is first designed by Eugene Garfield. IF is published by ISI as a standardized measure to reveal the mean number of citations to the published articles in particular journal. Journal Citation Reports (JCR) defines the IF as the frequency of citation to the "average papers" published in a journal in an especial year. In effect, IF is almost used by research institutions, universities and governmental agencies to measure the influence and impact of a professional journal comparing with other journals in the same area of research (Kuo & Rupe, 2007). IF is also applied to evaluate the individual profession of scientists and researchers. Therefore, editors have a high tendency to increase the IF of their journals.

In order to calculate the IF of a journal, the number of citations to the published papers in a journal, which is reported in the JCR year, is divided by the total number of papers published in that journal during two previous years. Citation may be comes from articles published in the same journal or different proceedings, journals or books indexed by WOS. As an example, an IF of 3.0 reveals that, on average, the papers published in a particular journal within two years ago, have been cited three times. Although the IF calculation for journals is based on articles published and cited in the previous two years, one can calculate the average citation using longer time periods.

The 5-year journal IF is the number of citations to a particular journal divided by the total number of papers published in the journal since five years ago. In certain fields of study, the 5-year journal IF is more appropriate than 2-years because the body of citations in 5-years basis might be large enough to make a reasonable comparison. In some area of research, it takes a long time, more than two years, to publish a work and response to previous articles.

The aggregate IF is applied for a subject category rather than a journal only (Khan & Hegde, 2009). The calculation is the same way as the IF for a journal, but it considers the total number of citations to all journals in a particular subject group and the number of papers published in these cited journals. The median IF refers to the median value of IF for all journals in the subject category. The number of citations to a journal divided by the total papers published in the journal in current year is known as immediacy index. The journal's half-life in each year refers to the median age of the article cited in JCR. As an example, if the value of journal's half-life equals to six in 2007, it means that the citations from 2002 until 2007 to the journal are 50% of all the citations to the journal in 2007. The journal ranking table displays the ranking of a journal in related subject category based on the journal IF (Khan & Hegde, 2009).

3. Praises for Impact Factor

IF as a measure of journals' influence has several advantages and benefits for researchers as well as librarians, knowledge managers and information professionals. Some of the advantages are listed below:

1) The primary advantage of impact factor is that it is very easy to measure.

2) Another advantage of IF is to mitigate the absolute citation frequencies. Since large journals provide a considerable body of citable literature which cites to the small journals, IF discounts the benefits of large journals in favour of small journals. In addition, there is a tendency to discount the benefits of high frequently published journals in favour of less frequently published journals and also matured journals over the newly established journals. In other words, journal IF is a remarkable measure for evaluating the journals since it offsets the benefits of age and size among the journals.

3) IF is a useful tool to compare different research groups and journals. It is commonly applied to administer scientific library collections. Librarians, knowledge managers and Information professionals encounter with limited budgets when they aim to select journals for their institutions and departments. ISI provide them with the IF as a measure to choose most frequently cited journals (Khan & Hegde, 2009). About 9000 social science and science journals from 60 countries are indexed by Web of Knowledge.

4) IF for indexed journals are greatly available and easy to understand and use. IF is more acceptable and popular than other alternative measures (Khan & Hegde, 2009).

4. Criticisms for Impact Factor

However, IF has several inherent shortcomings overcoming the advantages of IF. The following problems refer to the calculation method of IF:

1) It should be noted that Thomson Scientific is a commercial institution, which sells its products and evaluations to research institutions and publishers. Data and measures of IF are not publicly established and it is not easy for scientists to access these data and measures. Therefore, it is not subject to a peer review process. Researchers demand a fully transparency over how Thomson Scientific collect data and calculates citation metrics. Thomson Scientific is unable to clarify the used data to support its

published IF and hence its IF is not reliable. Generally, findings of a scientific article could not be accepted by scientists when the primary data is not available.

2) The coverage of database is not complete since Thomson Scientific excludes some specific types of sources from the denominator. As an example, books (as a source for citations) are excluded from the database. Citations from journals which are not indexed in the ISI are not considered in IF calculation (Khan & Hegde, 2009). Falagas, Kouranos, Arencibia-Jorge, and Karageorgopoulos (2008) assert that one of the major shortcomings of journal IF is considering only “citable” papers mainly original papers and reviews.

3) IF could not be an appropriate measure because it is an arithmetic mean of number of citations to each paper and the arithmetic mean is not an appropriate measure (Joint Committee on Quantitative Assessment of Research 2008). IF does not consider the actual quality of research articles, their impressiveness, or the long-term impact of journals. It should be noted that IF and citation indexes indicate a specific, and relatively uncertain, type of performance. They do not provide a direct measure of quality (Moed, 2005). Therefore, Using the IF to derive the quality of individual articles or their authors seems to be idle (Baum, 2011). In addition, the validity and appropriate use of the IF as a measure of journal importance is a subject to controversy and another independent examination might reproduce a different IF (Rossner, Van Epps, & Hill, 2007). IF make more controversy when it is used to assess the articles published in the journals. IF appraises the reputation of the publishing journal not the quality of the content in the single papers. In addition, Falagas et al. (2008) assert that IF takes into account received citations only in a quantitative manner. It is conceivable that the published papers in a journal have a greater influence on science if these papers are cited by articles with a higher scientific quality.

4) The two-year or five-year window for measuring the IF might be logical for some fields of studies with a fast moving process of research, while it is not reasonable for some fields of study, which requires a long period for research or empirical validation

and also takes a long time for review. These kinds of research might take a time longer than two years to be completed and then published. Therefore, citation to the original papers will not be considered in the IF of publishing journal. In addition, articles, which are published in many years ago and are still cited, have a significant impact on research area, but unfortunately, citation to these articles will not be considered due to their old age.

5) Distributing the same value of IF to each article published in a same journal leads to excessive variability in article citation, and provides the majority of journals and articles with the opportunity to free ride on a few number of highly cited journals and articles. Only some articles are cited anywhere near the mean journal IF. In effect, IF highly overstates the impact of most articles, while the impact of rest articles is greatly understated because of the few highly cited articles in the journal (Baum, 2011). Therefore, journal IF could not portray the individual impact of single articles and it is not much related to the citedness of individual articles in publishing journal (Yu & Wang, 2007). Seglan (1997) states that IF is wrongly used to estimate the importance of a single article based on where it is published. Since IF averages over all published papers, it underestimates the importance of the highly cited papers whereas overestimating the citations to the average publications.

6) Although, it is feasible to evaluate the IF of the journals in which an individual has published papers, the IF is not applicable to individual scientists or articles. ISI does not distribute a JCR for the humanities. The relevant number of citations received by a single paper is a better measure of importance and influence of individual papers as well as its authors (Khan & Hegde, 2009). Garfield (1998) caution regarding the "misuse in evaluating individuals" by IF due to a wide variation from paper to paper within an individual journal.

7) The IF is highly depended on different disciplines such as physical sciences and mathematical because of the speed with which articles are cited in a research area (Van Nierop, 2009). It is not logical to use IF as a measure to compare journals across different disciplines. The absolute value of IF does not make any

sense. For example, a journal with the value of IF equal to two could not have much impact in fields and disciplines such as Microbiology, while it would be impressive in Oceanography. Therefore, IF as a measure of comparison between different fields and disciplines could be considered as invalid. In addition, this comparison has been widely made not only between the journals, but also between scientists or university departments, while, absolute IF could not estimate the scientific level of department. In an evaluation program, especially for doctoral degree granting institutions, reviewers consider the IF of journals in examining the scholarly outputs. It makes a bias in which some types of researches are undervalued and falsifies the total contribution of each scientist (Khan & Hegde, 2009).

8) IF is not relevant in some disciplines such as engineering, where the main scientific outputs are technical reports, conference proceedings and patents. Regarding the literature, IF is not a relevant measure because books constitute the most important publications in literature, which cite other books.

9) The number of citations from journals in less universal languages and less-developed countries is understated, since only ISI database are utilized to define the IF. It is argued that the simple methodology applied in the calculation of IF provides editors with the opportunity to use different practices in order to inflate the impact factor of their journals (Garfield, 1996; Hemmingsson, Skjennald, & Edgren, 2002; The PLoS Medicine Editors, 2006). Reviewers may request from the authors to expand the article's citation before publishing the articles. This is usual in peer review process in order to improve the quality of article. On the other hand, IF indicates the importance of a journal in related subject category. Therefore, editors have a high tendency to increase their journal IF (Yu, Yang, & He, 2011).

There are several methods for a journal to increase its IF. One policy is to manipulate the IF (Dong, Loh, & Mondry, 2005). References to a journal can be managed in order to increase the number of citations to a journal and hence raise the IF (Smith, 1997). Therefore, these journals seem to have additional impact and their quality and quantity will be increased. However, there is no actual improvement in the journal impact and influence. This manipulation of IF will only result in artificial evaluation.

1) Self-citation is a common way to manipulate and increase the IF of journals (Miguel & Marti-Bonmati, 2002; Fassoulaki, Papilas, Paraskeva, & Patris, 2002; Falagas & Kavvadia, 2006; Falagas & Alexiou, 2007; Yu et al., 2011). Editorial board may force authors to expand the article's citations not necessarily to improve the quality of scientific articles but they aim to inflate the IF of their journal in order to artificially raising the journal's scientific reputation (Wilhite & Fong, 2012). This is a business policy, which is called coercive citation. In effect, editors require authors to cite papers published in the same journal (self-citation) before the journal accepts to publish the article even if these citations are not relevant to the research. In their study, (Wilhite & Fong, 2012) find that 20% of researchers in psychology, sociology, economics and multiple business disciplines has experienced coercive citation. They find that journals with a lower IF have a high tendency for coercive citation in order to inflate their IF. Another study documents that coercive citation is common in other scientific disciplines.

2) Another policy to manipulate the IF is to focus on publishing review articles rather than research articles because review articles have a large body of literature, which are highly citable (Andersen, Belmont, & Cho, 2006; Falagas et al., 2008; Arnold & Fowler, 2011). Therefore, review articles provide the higher IF for publishing journals and these journals will have the highest IF in their respective research area. Journals have a tendency to publish a large number of their articles, which are likely to receive a high citation in near future or deny accepting papers such as case report in medical journals, which are not expected to be cited.

Considering the above drawbacks of IF, it is suggested that IF as a measure of journal evaluation should have certain characteristics (Offutt, 2008). It should be independent from the length and number of the papers published in the journal. In addition, it should consider essential advances in the area over short term or incremental contributions. IF should have relative stability in a year to year basis. The most important is that IF should evaluate the impact of a journal during a long period since some articles are still cited after 10, 20, or even 50 years from the date of publication. In addition, we should consider the warnings by Garfield against using to compare different IF and scientific field. Beside the IF we can consider other internet databases to calculate the IF of journals or articles which is collaborating with Elsevier.

The journal IF is an indicator of journal performance, and as such many of the suggestions for performance indicators could equally be used to the application of journal IF (Greenwood, 2007). In November 2007 the European Association of Science Editors (EASE) published an official statement suggesting that journal IF can be applied only and cautiously for assessing and comparing the importance and impact of journals, but not for evaluating the single articles, and of course not for appraising the research programmes or researchers (EASE, 2007). The International Council for Science (ICSU) Committee on Freedom and Responsibility in the conduct of Science (CFRS) issued a "Statement on publication practices and indices and the role of peer review in research assessment" in July 2008, which suggest to consider only a limited number of articles published per year by each scientist or even give a penalty to scientists due to excessive number of publications in each year (Smith, 1997). The Deutsche Forschungsgemeinschaft (2010), German Foundation for Science, issued new strategies to assess only papers and no bibliometric information on candidates in decision making process regarding:

"...performance-based funding allocations, postdoctoral qualifications, appointments, or reviewing funding proposals, [where] increasing importance has been given to numerical indicators such as the h-index and the impact factor".

The citation style of manipulated journals, especially the number of self-citations in the last two year appears to be unusual. It can be seen that the self-cited rate of manipulated journals is higher than that of the normal journal. Therefore, self-citation rate can be used as a measure to specify the manipulated journals and calculate the real IF by deducting the self-citations from the reported IF. However, it is a difficult task to specify manually each journal in the JCR database. As discussed, journal might publish a larger number of review articles in order to inflate their IF. Therefore, the Thomson Scientific website can provide directions for removing review articles from the calculation (Khan & Hegde, 2009).

5. Conclusion

Although IF is a popular measure of quality for journals, we conclude that journal IF has its own restrictions. We believe that there should be more examination of whether and how IF evaluates journal quality before it is broadly accepted as a measure of journal quality. The IF measurement would certainly never assess the peer review process of journals (Offutt 2008). It means that IF is not an appropriate measure to assess the quality of published articles. One important factor in assessing a progress case is to consider the number of published articles but it is an imperfect method because it considers the number of articles only instead of the quality of publications. However, it is very hard to assign professionals to read articles and give an independent opinion about the impact of particular scientist on his or her research area.

The implication for researchers and journals is that they should not rely only on this indicator. If we do not consider the above limitations associated with IF, decisions made based on this measure is potentially misleading. Generally, a measure will fall into disuse and disrepute among the scientific community members, if it is found to be invalid, unreliable, or ill-conceived. Although IF has many limitations as discussed in this paper, it does not lose its reputation and application by the scientific society. Indeed, IF attracts more attention and being used more frequently by scientists and librarians, knowledge managers and information professionals. Critically, extensive use of IF may result in destroying editorial and researchers' behaviour, which could compromise the quality of scientific articles. Calculation of IF and policies to increase the IF by journals, may push researchers to consider publication as a business rather than contribution to the area of research. It is not fair that we should rely on such a non-scientific method which IF appraises the quality of our efforts. It is the time of the timeliness and importance of a new invention of journal ranking techniques beyond the journal impact factor.

There should be a new research trend with the aim of developing journal rankings that consider not only the raw number of citations received by published papers, but also the influence or importance of documents which issue these citations (Palacios-Huerta & Volij, 2004; Bollen, Rodriguez, & van de Sompel, 2006; Bergstrom, 2007; Ma, Guan, & Zhao, 2008). The new measure should represent scientific impact as a function not of just the quantity of citations received but of a combination of the quality and the quantity.

References

Andersen, J., Belmont, J., & Cho, C. T. (2006). Journal impact factor in the era of expanding literature. *Journal of Microbiology, Immunology and Infection, 39,* 436-443.

Arnold, D. N., & Fowler, K. K. (2011). Nefarious numbers. *Notices of the American Mathematical Society, 58*(3), 434-437.

Baum, J. A. C. (2011). Free-riding on power laws: Questioning the validity of the impact factor as a measure of research quality in organization studies. *Organization, 18*(4), 449-466.

http://dx.doi.org/10.1177/1350508411403531

Bergstrom, C. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals, *College & Research Libraries News, 68*(5), 314-316.

Bollen, J., Rodriguez, M. A., & Sompel, H. (2006). Journal status, *Scientometrics, 69*(3), 669-687. http://dx.doi.org/10.1007/s11192-006-0176-z

Dong, P., Loh, M., & Mondry, A. (2005). The impact factor revisited. *Biomedical Digital Libraries, 2.* European Association of Science Editors (EASE). (2010). *Statement on inappropriate use of impact factors*. Retrieved from http://www.ease.org.uk/publications/impact-factor-statement

Falagas, M. E., & Alexiou, V. G. (2007). Editors may inappropriately influence authors' decisions regarding selection of references in scientific articles*. International Journal of Impotence Research, 19*, 443-445. http://dx.doi.org/10.1038/sj.ijir.3901583

Falagas, M. E., & Kavvadia, P. (2006). Eigenlob: Self-citation in biomedical journals. *Federation of American Societies for Experimental Biology Journal, 20*, 1039-1042. http://dx.doi.org/10.1096/fj.06-0603ufm

Falagas, M. E., Kouranos, V. D., Arencibia-Jorge, R., & Karageorgopoulos, D. E. (2008). Comparison of SCImago journal rank indicator with journal impact factor. *The Federation of American Societies for Experimental Biology Journal, 22*(8), 2623-2628.

Fassoulaki, A., Papilas, K., Paraskeva, A., & Patris, K. (2002). Impact factor bias and proposed adjustments for its determination. *Acta Anaesthesiologica Scandinavica, 46*, 902-905.

http://dx.doi.org/10.1034/j.1399-6576.2002.460723.x

Garfield, E. (1996). How can impact factors be improved? *British Medical Journal, 313*, 411-413. http://dx.doi.org/10.1136/bmj.313.7054.411

Garfield, E. (1998). Der impact faktor und seine richtige anwendung. *Der Unfallchirurg, 101*(6), 413-414.

Garner, R., Lunin, L., & Baker, L. (1967). *Three Drexel information science research studies*. Madison: Drexel Press.

Giles, C. L., Bollacker, K. D., & Lawrence, S. (1998). An automatic citation indexing system, Digital Libraries 98- Third ACM Conference on Digital Libraries, New York, 89-98.

Greenwood, D. C. (2007). Reliability of journal impact factor rankings. *BMC Medical Research Methodology, 7*(48), 1-6.

Hemmingsson, A., Mygind, T., Skjennald, A., & Edgren, J. (2002). Manipulation of impact factors by editors of scientific journals. *American Journal of Roentgenology, 178*(3), 767-767.

http://dx.doi.org/10.2214/ajr.178.3.1780767

Joint Committee on Quantitative Assessment of Research. (2008, June 12). Citation Statistics, 1-26.

Khan, & Hegde, (2009). Is impact factor true evaluation for ranking quality measure? *Journal of Library & Information Technology, 29*(3), 55-58.

Kuo, W. & Rupe, J. (2007). R-impact: reliability-based citation impact factor. *IEEE Transactions on Reliability 56*(3), 366-367.

http://dx.doi.org/10.1109/TR.2007.902789

Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing and Management, 44*(2), 800-810. http://dx.doi.org/10.1016/j.ipm.2007.06.006

Miguel, A., & Marti-Bonmati, L. (2002). Self-citation: comparison between Radiologia, European radiology and radiology for 1997–1998. *European Radiology, 12*(1), 248-252.

http://dx.doi.org/10.1007/s003300100894

Moed, H. F. (2005). *Citation analysis in research evaluation. Information Science and Knowledge Management.* Springer.

Nigam, A., & Nigam, P. K. (2012). Citation index and impact factor. *Indian Journal of Dermatology, Venereology, and Leprology, 78*(4), 511-516. http://dx.doi.org/10.4103/0378-6323.98093

Offutt, J. (2008). The journal impact factor. *Journal of software testing verification and reliability, 18*(1).

Palacios-Huerta, I., & Volij, O. (2004). The measurement of intellectual influence. *Econometrica, 72*(3), 963-977. http://dx.doi.org/10.1111/j.1468-0262.2004.00519.x

Rossner, M., Van Epps, H., & Hill, E. (2007). Show me the data. *Journal of Cell Biology, 179*(6), 1091-1092. http://dx.doi.org/10.1083/jcb.200711140

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal, 314*(7079), 498-502. http://dx.doi.org/10.1136/bmj.314.7079.497

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269.

http://dx.doi.org/10.1002/asi.4630240406

Smith, R. (1997). Journal accused of manipulating impact factor. *British Medical Journal, 314*(15), 461. http://dx.doi.org/10.1136/bmj.314.7079.461d

The PLoS Medicine Editors. (2006). The impact factor game: It is time to find a better way to assess the scientific literature. *PLoS Med., 3,* 291. http://dx.doi.org/10.1371/journal.pmed.0030291

Van Nierop, E. (2009). Why do statistics journals have low impact factors? *Statistica Neerlandica, 63*(1), 52-62. http://dx.doi.org/10.1111/j.1467-9574.2008.00408.x

Wilhite, A. W., & Fong, E. A. (2012). Coercive citation in academic publishing. *Science, 335*(6068), 542-543.

http://dx.doi.org/10.1126/science.1212540

Yu, G., & Wang, L. (2007). The self-cited rate of scientific journals and the manipulation of their impact factors. *Scientometrics, 73*(3), 356-366. http://dx.doi.org/10.1007/s11192-007-1779-8

Yu, G., Yang, D. H., & He, H. X. (2011). An automatic recognition method of journal impact factor manipulation. *Journal of Information Science, 37*(3), 235-245.

http://dx.doi.org/10.1177/0165551511400954

26.

Study reveals declining influence of high impact factor journals

William Raillant-Clark
Université de Montréal. Nouvelles, 07.11.2012

– ◊ –

The most prestigious peer-reviewed journals in the world, such as Cell, Nature, Science, and the Journal of the American Medical Association (JAMA), have less and less influence amongst scientists, according to a paper co-authored by Vincent Larivière, a professor at the University of Montreal's School of Library and Information Sciences. He questions the relationship between journal "impact factor" and number of citations subsequently received by papers.

"In 1990, 45% of the top 5% most cited articles were published in the top 5% highest impact factor journals. In 2009, this rate was only 36%," Larivière said. "This means that the most cited articles are published less exclusively in high impact factor journals." The proportion of these articles published in major scholarly journals has sharply declined over the last twenty years. His study was based on a sample of more than 820 million citations and 25 million articles published between 1902 and 2009. The findings were published in the Journal of the American Society for Information Science and Technology.

For each year analysed in the study, Larivière evaluated the strength of the relationship between article citations in the two years following publication against the journal impact factor. Then, he compared the proportion of the most cited articles published in the highest impact factor journals. "Using various measures, the goal was to see whether the ‘predictive' power of impact factor on citations received by articles has changed over the years," Larivière said. “From 1902 to 1990, major findings were reported in the most prominent journals," notes Larivière. But this relationship is less true today.”

Larivière and his colleagues George Lozano and Yves Gingras of UQAM's Observatoire des sciences et des technologies also found that the decline in high impact factor journals began in the early 90s, when the Internet experienced rapid growth within the scientific community. “Digital technology has changed the way researchers are informed about scientific texts. Historically, we all subscribed to paper journals. Periodicals were the main source for articles, and we didn't have to look outside the major journals,” Larivière noted. “Since the advent of Google Scholar, for example, the process of searching information has completely changed. Search engines provide access to all articles, whether or not they are published in prestigious journals.”

Impact factor as a measure of a journal's influence was developed in the 1960s by Eugene Garfield, one of the founders of bibliometrics. "It is basically the average number of times a journal's articles are cited over a two-year period," Larivière explained. “Initially, this indicator was used to help libraries decide which journals to subscribe to. But over time, it began to be used to evaluate researchers and determine the value of their publications.” The importance of impact factor is so ingrained in academia's collective consciousness that researchers themselves use impact factor to decide which journals they will submit their articles to.

Various experts in bibliometrics have criticized the use of impact factor as a measure of an academic journal's visibility. A common criticism is that the indicator contains a calculation error. “Citations from all types of documents published by journal are counted,” Larivière said, “but they are divided only by the number of articles and research notes. Impact factor is thus overestimated for journals that publish a good deal of editorials, letters to the editor, and science news, such as Science and Nature.”

Another criticism is that the time frame in which citations are counted in calculating impact factor is too short. “There are research areas in which knowledge dissemination is faster than it is in others,” Larivière said. “We cannot, for example, expect to get the same kind of impact factor in engineering and biomedical sciences.” Yet journal impact factor is established in the two-year period following publication of articles regardless of the discipline.

The research results reveal some interesting points. On the one hand, journals are increasingly poor predictors of the number of citations an article can expect to receive. “Not only has the predictive power of impact factor declined, but also, impact factor is no longer suitable for evaluating research,” Larivière argued. In his opinion, if we want to evaluate researchers and their work, it is best to use citations, which are a true measure of an article's impact. “This indicator is more accurate. It is not an estimation based on the hierarchy of journals.” On the other hand, his work confirms that the dynamics of scholarly journals is changing, due especially to the open access of knowledge made possible by the Internet. “What then is the present function of scholarly journals?" Larivière asked. "One remains: peer review."

27. SCIENCE MAGAZINE: COERCIVE CITATION IN ACADEMIC PUBLISHING

EDAWAX - EUROPEAN DATA WATCH EXTENDED

– ◊ –

As many of you know, the impact factor is a major vehicle for measuring the quality of a scholarly journal. Despite there is a lot of criticism on impact factors, for researchers as well as for journals a high impact factor is as attractive as honey is for the bears.

One side effect of impact factors is that they're creating incentives for editors to coerce authors to add citations to their journal – indicating that more citations are inflating the journal's impact factor.

At the beginning of February, the Science magazine published a remarkable article that deals with forced citations in scholarly journals.

The Authors, Allen W. Wilhite and Eric A. Fong analyzed 6672 responses from a survey they sent to researchers of different scientific disciplines as economics, sociology, psychology, marketing, management, finance, information systems, and accounting. They accompanied the data of 832 journals in those same disciplines in the survey.

The result is disturbing: Wilhite and Fong found out that coercion is "uncomfortably common and appears to be practiced opportunistically." Although 86% of the respondents view coercion as inappropriate, 57 % would add superfluous citations before submitting to a journal known to coerce.

Although most of the interviewed person dispraise coercion, less than 7% thought that an author would refuse to add self-citations if coerced to do so.

"Coercive self-citation is more common in the business disciplines than in economics, sociology, and psychology" the authors stated – but this these disciplines are not immune: Wilhite and Fong mentioned that in every discipline instances of coercion were reported.

Senior faculty members seemed to be more resistant against coercive demands than junior faculty members – that's not surprising when taking into account that the senior members mostly have fixed positions at universities and often are professors while junior researchers are competing hard to get one of these positions. And we all know the dictum: "Publish or perish!"

Well hidden at the end of the supplementary materials the authors are presenting an explosive table – the list of journals identified as coercers by survey respondents. It is a pleasure for me to present you the ten "top scorers" – half of them published by Elsevier:

Journal of Business Research

Journal of Retailing

Marketing Science

Journal of Banking and Finance

Information and Management

Applied Economics

Academy of Management Journal

Group and Organization Management

Journal of Consumer Psychology

Psychology and Marketing

28.

Das schmutzige Geschäft mit Zwangszitaten

Olaf Storbeck

– ◊ –

Anfang Februar ist in der Fachzeitschrift "Science" eine bemerkenswerte Studie erschienen: Die beiden US-Ökonomen Allen W. Wilhite und Eric A. Fong leuchten aus, wie sehr wissenschaftliche Fachzeitschriften versuchen, ihre Reputation zu manipulieren.

Im Kern geht es um die folgende Frage: Wie sehr werden Wissenschaftler, die eine Arbeit bei einer Fachzeitschrift einreichen, von den Herausgebern dazu genötigt, Arbeiten zu zitieren, die sie eigentlich gar nicht zitieren wollen?

Wilhite und Fong haben für ihre Studie mit dem Namen "Coercive Citation in Academic Publishing" fast 6700 Wissenschaftler nach ihren Erfahrungen damit befragt.

Das heißeste Eisen der Studie ist ganz am Ende des Online-Anhangs versteckt: Die Schwarze Liste mit den Namen der Fachzeitschriften, bei denen das nach Angaben der Befragten besonders häufig vorkommt.

Journal of Business Research – 49 Nennungen

Journal of Retailing – 43 Nennungen

Marketing Science – 29 Nennungen

Journal of Banking and Finance- 24 Nennungen

Information and Management – 19 Nennungen

Applied Economics – 18 Nennungen

Academy of Management Journal – 14 Nennungen

Group and Organization Management – 13 Nennungen

Journal of Consumer Psychology – 9 Nennungen

Psychology and Marketing – 8 Nennungen

Fachzeitschriften und auch Einzelforscher können mit erzwungenen Zitaten ihre wissenschaftliche Reputation schönen – denn der Einfluss und die Qualität von Forschung wird heute stark daran gemessen, wie oft andere Forscher die Arbeiten aufgreifen.

Es gilt die Faustregel: Je häufiger eine Studie zitiert wird, desto einflussreicher ist sie. Und je häufiger Artikel aus einer bestimmten Fachzeitschrift zitiert werden, desto stärker wird sie beachtet.

(Auch die VWL- und BWL-Forscherrankings des Handelsblatts funktionieren im Kern nach diesem Prinzip – die Impact Faktoren der Journals bestimmen das Gewicht, dass ein Aufsatz aus dieser Zeitschrift im Ranking erhält.)

Das gibt den Herausgebern der Fachzeitschriften Macht gegenüber Wissenschaftlern, die darin veröffentlichen wollen. Sie können Aufsätze ohne große Begründung ablehnen – oder den Autoren einen Kuhhandel anbieten. Nach dem Motto: Wenn du ein, zwei Arbeiten, die in unserem Journal erschienen sind, erwähnst, dann würden wir den Artikel drucken.

Die Forscher schildern zwei Beispiele:

“Recently a friend had an article accepted for publication. Two weeks after that acceptance the editor sent another letter asking the author to add citations from his journal. Specifically the editor wrote, “you only use one (name of my journal) source which is unacceptable. Please add at least five more relevant-(name of my journal) sources.”

Another colleague had a similar experience at the submission stage; the editor asked her for three more citations to his journal before he would send her manuscript out for review. Note, these citation requests did not mention omitted content or shortcomings in the manuscript’s analysis; they simply asked authors to cite related articles in the editor’s journal.”

Wilhite und Fong stellen fest, dass es sich dabei längst nicht nur um Einzelfälle handelt. Ihr Fazit lautet:

"We find that coercion is uncomfortably common and appears to be practiced opportunistically. As editors game the system and authors acquiesce, the integrity of academic publications suffers. (…) Additionally, some editors seem to target specific articles and authors. (…) Without action, the situation is likely to deteriorate,"

Besonders stark verbreitet ist diese Praxis offenbar in der Betriebswirtschaftslehre – dort kommt es deutlich häufiger vor als in der VWL, der Soziologie und der Psychologie.

Die ganz große Mehrheit der befragten Wissenschaftler (86%) hält die Zitate-Nötigung für unangebracht, und 81 % sind überzeugt: Darunter leidet auf Dauer das Prestige einer Fachzeitschrift.

Dennoch würden 57% sich auf das Spielchen einlassen und im Fall der Fälle eigentlich überflüssige Zitate einbauen. Und nur 7% erwartet, dass sich die Kollegen einem Nötigungsversuch widersetzen würden.

Die beiden Forscher empfehlen, dass Wissenschaftsverbände wie die "American Economic Association" einschreiten sollten:

"Academic associations could help by officially condemning the practice. Their action would raise the cost of coercion to editors and might help persuade organizations that promote impact factors to remove self-citations from those calculations, which would eliminate the coercive motive."

Alles in allem sind die Ergebnisse der Studie wirklich beklemmend – sie stützen eindeutig die These des IfW-Chefs Dennis Snower, der jüngst in einem Interview mit mir harsche Kritik am derzeitigen Publikationsprozess übte:

"Das traditionelle Verfahren gibt den Herausgebern der Journale und den Gutachtern zu viel Macht – diese Leute können Gott spielen. Die ethischen Werte, die für Wissenschaftler zentral sein sollten, sind uns abhanden gekommen. (…)

Im traditionellen Gutachterverfahren fehlt die Transparenz. Herausgeber, die eine Studie aus welchen Gründen auch immer persönlich nicht mögen, können sie ganz bewusst an Gutachter schicken, von denen sie sehr genau wissen, dass sie die

Arbeit ablehnen werden. Viele Gutachter arbeiten auch unglaublich langsam und schreiben unfaire Reports. Das hat vermutlich jeder Ökonom schon mehrfach erlebt. Neue Ideen haben es dadurch sehr schwer."

29.

In medio stat virtus: some thoughts about journal Impact Factor

J. Landeira-Fernandez; D. F. Ventura; A. Pedro de Mello Cruz

Psychol. Neurosci. vol.6 no. 1

– ◊ –

Journal editors are constantly seeking to improve the excellence of their journals. One of the main endeavors of an editor is to attract the attention of the scientific community and thus increase the number of manuscript submissions and quality of papers. A journal's Impact Factor is currently the most widely employed tool to express a journal's quality. Problems in the calculation of this index and its misuse, however, have caused several academic consequences. A group of high-impact journal editors and publishers of scholarly journals convened a discussion group during the annual meeting of the American Society for Cell Biology in December 2012 in San Francisco to address this issue. Further discussions of this group led to the publication of a manifesto in May 2013 known as The San Francisco Declaration on Research Assessment, which recognizes noteworthy limitations of the Impact Factor.

The concept of Impact Factor was introduced by Gross & Gross (1927) who argued that the most frequently cited journals are the most relevant to the field and thus the most valuable journals for a library to purchase. For that reason, they suggested counting references as a measure to rank the use of scientific journals. In 1955, Eugene Garfield first mentioned the idea of an impact factor (Garfield, 1955), although the term "impact factor" itself was introduced later by Garfield & Sher (1963). In 1960, Garfield founded the Institute for Scientific Information (ISI) and initiated the Science Citation Index in 1963. Over time, the ISI absorbed the Social Sciences Citation Index and the Arts & Humanities Citation Index. As a consequence of these bibliographic databases, Journal Citation Reports (JCR) was launched in 1975, which contained the Impact Factor for each journal indexed by the ISI. In 1992, Garfield sold the ISI to Thomson Scientific for a substantial amount of money (USD $210 million). Thomson Scientific currently publishes JCR annually. In 2005, Thomson Scientific began Web of Science, which integrated all of the Thomson Scientific databases. In 2007, Thomson Scientific merged with Reuters to become one of the primary cooperatives for scientific information. In 2009, Thomson Reuters Scientific (TRS) had more than 55,000 employees in more than 100 countries and revenue of ~$13 billion USD annually (Beira, 2010).

The Impact Factor considers only journals indexed by TRS. It is calculated as the ratio between the total number of citations that each paper that was published in that journal received during the preceding 2 years and the total number of papers that the journal published during the same 2-year period. For example, the 2012 Impact Factor reflects the number of citations in 2012 of articles published by the journal in 2010 and 2011, divided by all of the articles that this journal published during the 2010-2011 period. Let us suppose that journal "X" published 100 papers during 2010 and 2011. From all of these paper published during this period, journal "X" was cited 300 times in all of the journals indexed by TRS during 2012. Therefore, the Impact Factor of journal "X" in 2012 is 300/100 = 3.0. Notice that the 2012 Impact Factor of journal "X" can only be published by JCR in 2013 because all of the citations that occurred in 2012 need to be computed.

Several criticisms have been raised about the way in which the Impact Factor is calculated. For example, TRS does not index all scientific journals worldwide. Indeed, TRS is extremely selective about which journals they select. In 2011, TRS covered 16,350 journals. By comparison, Elsevier's Scopus began publishing the

SCImago Journal Rank indicator in 2004, which is a TRS competitor because of its citation tracking capabilities. Scopus covered 26,447 journals during the same period (Nagaraja, & Vasanthakumar, 2011).

One of the main problems of the Impact Factor is that it does not consider the citation performance of each paper published by the journal. Taking our example of journal "X," which had an Impact Factor of 3.0 in 2012, it is possible that only one paper received 300 citations in 2012, and all of the other 99 papers published during the 2010-2011 period had no citations at all. Accordingly, it is well known that the paper citation distribution of a journal is strongly skewed to the right (i.e., a positive skew), meaning that a relatively low number of papers receives a very high number of citations, thus contributing disproportionately to the Impact Factor (Seglen, 1992). In fact, review papers from well-known and very productive groups of investigators tend to enhance the impact factor of a journal (Ketcham & Crawford, 2007). Therefore, the Impact Factor reveals very little about the citation of a paper published by this journal. Papers published in low-Impact Factor journals can receive a considerable number of citations. Conversely, other papers published in high-Impact Factor journals might receive very few citations. Therefore, the Impact Factor is not a comprehensive bibliometric indicator for evaluating the research impact of a particular author. Other parameters such as the H-index (introduced by Hirsch, 2005) may better reflect an author's scientific productivity because it takes into account the scholar's total number of published works and how many times they have been cited.

An often-mentioned criticism of the Impact Factor concerns journal self-citation in which an article in journal "X" cites another article published in the same journal "X." This artificially inflates the Impact Factor, especially if the journal editor encourages their authors to cite papers published by its own journal in the previous 2 years. In some cases, the editor may even force authors who submit papers to journal "X" to add these citations to their article before the journal will agree to publish their work. To correct this coercive citation practice, JCR began to publish an adjusted Impact Factor that excludes journal self-citations. In extreme cases when the journal presents extreme outlier self-citation behavior, TRS may exclude the journal from the database for 2 consecutive years (the journal's inclusion can be reevaluated in the third year).

Another malpractice that can be employed to artificially increase the Impact Factor is that a group of editors can team up for mutual benefit by forming a kind of citation cartel (Franck, 1999). Known as "citation stacking," the editor of journal "X" encourages their authors to cite papers published by journal "Y." In turn, the editor of journal "Y" encourages their authors to cite papers published by journal "X." Although this is a refined coercive citation practice, it can be detected by observing anomalous patterns of citations between journals in which journal "X" has an excessive concentration of citations of journal "Y" and vice versa. In this case, both the donor and recipient journals can be excluded from the TRS database for 1 year and reevaluated the following year.

Despite these problems and potential misuse, the advent of the Impact Factor represented a landmark in the field of scientific publication. As we discussed in our last editorial (Landeira-Fernandez, Ventura, & Cruz, 2012), journal evaluation is a complex task that involves both quantitative and qualitative assessment methods. One of the main elements of this process is the placement of a journal within its field. Publication and citation norms are specific to different areas of knowledge (Amin & Mabe, 2000; Cole, 1983). At the level of scientific journals, the Impact Factor can provide information about a journal's influence and impact within a given field that may help authors decide where they want to publish their manuscripts. Accordingly, a high correlation exists between Impact Factor and the quality ratings of journals by investigators (Saha, Saint, & Christakis, 2003). The Impact Factor can also help journal editors and publishers to track the efficacy and efficiency of editorial policies and the objectives of their journals in the long run. Government and public organizations worldwide can also employ this citation index as one indicator to evaluate journal success across a wide range of scientific fields. Therefore, the Impact Factor also has valuable informative advantages that obviously do not justify its indiscriminate use. For that reason, the Latin expression "in medio stat virtus" (i.e. virtue stands in the middle) expresses the concept that the Impact Factor ought to be used in moderation in conjunction with other journal indicators that consider the place where a journal is published and the field of knowledge that the journal covers, so that different journals worldwide can be evaluated and appropriately ranked.

References

Amin, M., & Mabe, M. (2000). Impact factors: Use and abuse. Perspectives in Publishing, 1, 1-6.

Beira, E. (2010). Eugene Garfield, from ISI to Thomson Reuters: A timeline. Mercados e Empresas: Dinâmicas e estratégias, 104, 1-17.

Cole, S. (1983). The hierarchy of the sciences? American Journal of Sociology, 89, 111-139.

Franck, G. (1999). Scientific communication: A vanity fair? Science 286(5437), 53.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. Science, 122(3159), 108-111.

Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing.American Documentation, 14(3), 195-201

Gross, P. L., & Gross, E. M. (1927). College libraries and chemical education. Science, 66(1713), 385-399.

Hirsch, J. E. (2005) An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102, 16569-16572.

Ketcham, C. M., & Crawford, J. M. (2007). The impact of review articles. Laboratory Investigation, 87(12), 1174-1185.

Landeira-Fernandez, J., Cruz, A. P. M., & Ventura, D. F. (2012). Psychology & Neuroscience is well-ranked by the Brazilian Qualis Psychology Committee (editorial). Psychology & Neuroscience, 5, 1-2.

Nagaraja, A., & Vasanthakumar, M. (2011). Comparison of Web of Science and Scopus impact factors of Indian journals. Library Philosophy and Practice, Paper 596.

Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor: A valid measure of journal quality? Journal of Medical Library Association, 91(1), 42-46.

Seglen, P. (1992). The skewness of science. Journal of American Society for Information Science, 43, 628-638.

30.

A Game without Rules: the Impact Factor System

Fortschr Röntgenstr 2013; 185(8): 697-698

– ◊ –

We actually planned – as is our standard routine – to report to you here on RöFo's 2012 Journal Impact Factor. Routinely – because up to now we have participated in the Impact Factor game [1] and subjugated ourselves to the Impact Factor system. But actually – as some of you have already noted with astonishment, there is no 2012 Journal Impact Factor, at least not for RöFo.

Thomson Reuters pulled RöFo from its Journal Citation Report® in June 2013 without advance notice [2]. Thus, a for-profit corporation in the Impact Factor's country of origin made an isolated decision that will have an impact on radiological research, science and industry in the country where radiology was born.

As the leading radiology journal in the German-speaking world and as a publication that reflects the status of radiology across the entire region, RöFo would have had a 2011 Impact Factor of 3.075, if Thomson Reuters had counted all articles for the entire year consistently and accurately. Instead, in mid-2012 Thomson Reuters published RöFo's 2011 Impact Factor as 2.758, complaining in mid-2013 – a full year later – that the self-citation rate in 2011 had been too high. This is particularly surprising, because citing all relevant articles is one of the key quality criteria for a scientific journal, and because, naturally, many of those citations can be to the leading journal of a given language itself. Even without a single instance of self-citation, RöFo's 2011 Impact Factor would have been a respectable 1.057, better than 36 of 116 journals in a ranking eliminating self-citation. Without any consideration of this data and according to a set of undefined rules, the Impact Factor for RöFo has now been reset to 0.000 for the next two years. Who plays with which data and according to which rules?

This development seemed as good a reason as any to once again ask the question: according to which rules is the Impact Factor determined and used? It's these rules that are currently questioned more and more from a wide range of perspectives [3].

Today, the Impact Factor is used in ways that go far beyond the original intention of its inventor, Dr. Eugene Garfield. Scientists and research organizations warned against the institutional misuse of the Journal Impact Factor in the San Francisco Declaration on Research Assessment (DORA) [4]: "Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions." And yet, the reality at our universities and scientific institutions is something quite different. We must ask ourselves why we use seemingly objective methods to assess academic quality and performance.

Even though the Impact Factor is based on a relatively simple formula, it is virtually impossible to reconstruct how exactly it is calculated. Although in its response [5] to DORA Thomson Reuters stressed that DORA's criticism was aimed at how the Impact Factor is used and not at its calculation, it's obvious that the calculation yields inconsistent results, mainly due to the fact that the data available aren't sufficiently transparent [4] [6]. The situation is exacerbated by human data collection and processing errors. Absurd examples, such as Thomson Reuters' coding of a product advertisement as "editorial material" and including the translation of the product slogan as the headline of the editorial [7], do nothing to alleviate doubts about Impact Factor data quality. The incorrectly calculated Impact Factor for 2011 is yet another result of poor data collection. The current rules determining which journals get included or excluded are as obscure as those that will impact future decisions. In view of this, the demand “show me the data” should be supplemented by the equally important request “show me the rules.”

RöFo suddenly became invisible in the Journal Citation Report® 2012. The reality, however, is the opposite: RöFo has had an enormous impact over the last 116 years and will continue to do so. It is and will remain the leading publication of the German-speaking radiology community.

The Americans’ decision, “we will not add Rofo to the current edition of the JCR” changes nothing. RöFo is still the journal it has always been, a publication that combines academic excellence and high-quality continuing education. All of us – readers, authors, reviewers, editors, and the publisher – will have to work even harder than ever at continuing to further develop RöFo as a scientific journal. Continue reading RöFo, visit us online athttp://roefo.thieme.de, submit your articles, and judge RöFo for yourselves.

Our hope remains that the growing criticism will initiate the development of alternative assessment factors using a scientific approach that employs modern tools, methods, and procedures to measure and report the true scientific impact of journals, authors, and articles. Shouldn’t the cradle of radiology and countless other scientific achievements also be capable of making a valuable contribution to setting the rules of the future [8]?

Yours sincerely,

Walter Heindel, Managing Editor

Gerhard Adam, Editor

Bernd Hamm, Editor

Hans H. Schild, Editor

Udo Schiller, Managing Director

References

1 The PloS Medicine Editors. The impact factor game. PLoS Med 2006; 3: e291 DOI: 10.1371/journal.pmed.0030291

2 The Thomson Reuters Corporation. Journal Citation Reports® Notices. http://admin-apps.webofknowledge.com/JCR/static_html/notices/notices.htm#editorial_information [Accessed 22 July 2013]

3 Osterloh M, Frey BS. Heißt „gut“ publiziert auch „gute“ Publikation? Über die Rolle des Impact Factors in der Wissenschaft. http://www.forschung-und-lehre.de/wordpress/?p=13772#more-13772 [Accessed 22 July 2013]

4 DORA. San Francisco Declaration on Research Assessment.

http://am.ascb.org/dora [Accessed 22 July 2013]

5 The Thomson Reuters Corporation. Thomson Reuters Statement Regarding the San Francisco Declaration on Research Assessment. Available at

http://researchanalytics.thomsonreuters.com/statement_re_sfdra/ [Accessed 22 July 2013]

6 Rossner M, Van Epps H, Hill E. Show me the data. J Cell Biol 2007; 179: 1091-1092 ; http://www.jcb.org/cgi/doi/10.1083/jcb.200711140 [Accessed 22 July 2013]

7 Web of Science.

Accession Number: WOS:000319104700001.

http://apps.webofknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=4&SID=W1DoKfP7bDKLBeKMIbL&page=1&doc=1 [Accessed 22 July 2013]

8 Marx W, Bornmann L. Journal Impact Factor: “the poor man’s citation analysis” and alternative approaches.

http://www.lutz-bornmann.de/icons/declaration.pdf [Accessed 22 July 2013]

31.

Impact Factors: A Broken System

Carly Strasser

DATAPUB, the California Digital Library, May 22 2013

– ◊ –

If you are a researcher, you are very familiar with the concept of a journal's Impact Factor (IF). Basically, it's a way to grade journal quality. From Wikipedia:

The impact factor (IF) of an academic journal is a measure reflecting the average number of citations to recent articles published in the journal. It is frequently used as a proxy for the relative importance of a journal within its field, with journals with higher impact factors deemed to be more important than those with lower ones.

The IF was devised in the 1970s as a tool for research libraries to judge the relative merits of journals when allocating their subscription budgets. However it is now being used as a way to evaluate the merits of individual scientists– something for which it was never intended to be used. As Björn Brembs puts it, "…scientific careers are made and broken by the editors at high-ranking journals."

In his great post, "Sick of Impact Factors", Stephen Curry says that the real problem started when impact factors began to be applied to papers and people.

I can't trace the precise origin of the growth but it has become a cancer that can no longer be ignored. The malady seems to particularly afflict researchers in science, technology and medicine who, astonishingly for a group that prizes its intelligence, have acquired a dependency on a valuation system that is grounded in falsity. We spend our lives fretting about how high an impact factor we can attach to our published research because it has become such an important determinant in the award of the grants and promotions needed to advance a career. We submit to time-wasting and demoralising rounds of manuscript rejection, retarding the progress of science in the chase for a false measure of prestige.

Curry isn't alone. Just last week Bruce Alberts, Editor-in-Chief of Science, wrote a compelling editorial about Impact Factor distortions. Alberts' editorial was inspired by the recently released San Francisco Declaration on Research Assessment (DORA). I think this is one of the more important declarations/manifestoes peppering the internet right now, and has the potential to really change the way scholarly publishing is approached by researchers.

DORA was created by a group of editors and publishers who met up at the Annual Meeting of theAmerican Society for Cell Biology (ASCB) in 2012. Basically, it lays out all the problems with impact factors and provides a set of general recommendations for different stakeholders (funders, institutions, publishers, researchers, etc.). The goal of DORA is to improve "the way in which the quality of research output is evaluated". Read more on the DORA website and sign the declaration (I did!).

An alternative to IF?

If most of us can agree that impact factors are not a great way to assess researchers or their work, then what's the alternative? Curry thinks the solution lies in Web 2.0 (quoted from this post):

…we need to find ways to attach to each piece of work the value that the scientific community places on it though use and citation. The rate of accrual of citations remains rather sluggish, even in today's wired world, so attempts are being made to capture the internet buzz that greets each new publication…

That's right, skeptical scientists: he's talking about buzz on the internet as a way to assess impact. Read more about "alternative metrics" in my blog post on the subject: The Future of Metrics in Science. Also check out the list of altmetrics-related tools at altmetrics.org. The great thing about altmetrics is that they don't rely solely on citation counts, plus they are capable of taking other research products into account (like blog posts and datasets).

32.

High impact factors are meant to represent strong citation rates, but these journal impact factors are more effective at predicting a paper's retraction rate

Bjoern Brembs

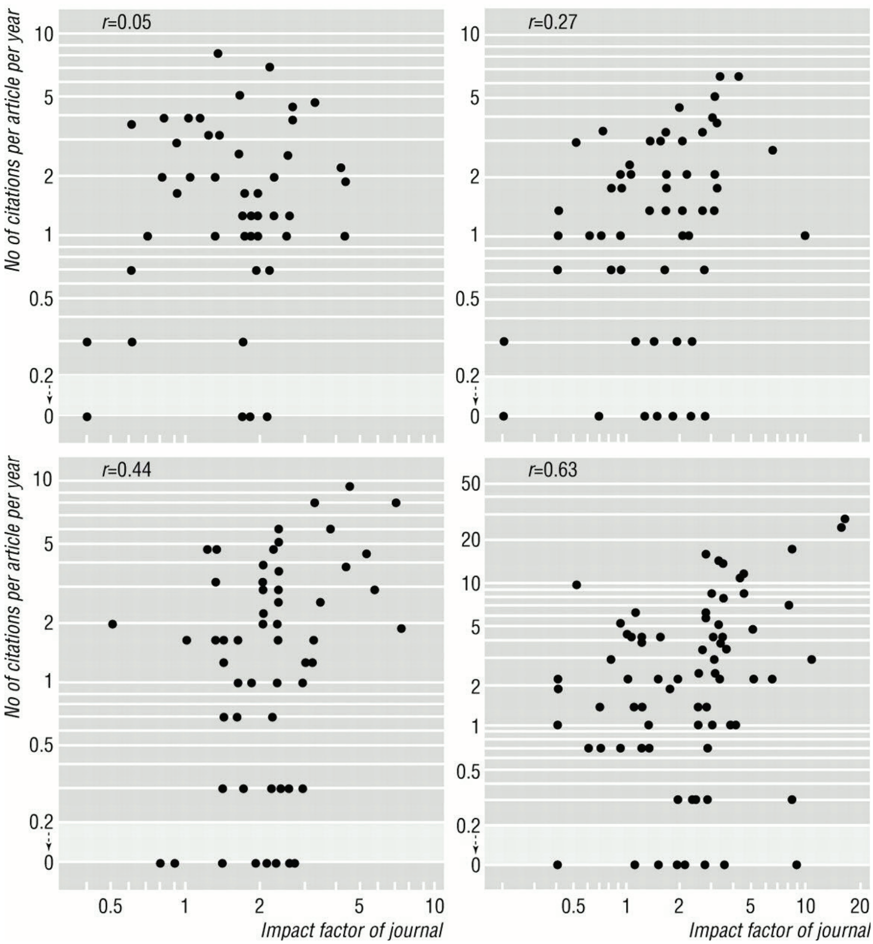
The Impact Blog, London School of Economics and Political Science

– ◊ –

With more than 24,000 scholarly journals in which some piece of relevant research may be published, a ranking scheme seems like a boon: one only needs to read articles from a small, high-ranking subset of journals and safely disregard the low-level chaff. At least this is how one might describe the development of journal ranks in the 1960s and 70s, when scores of new journals began to proliferate.

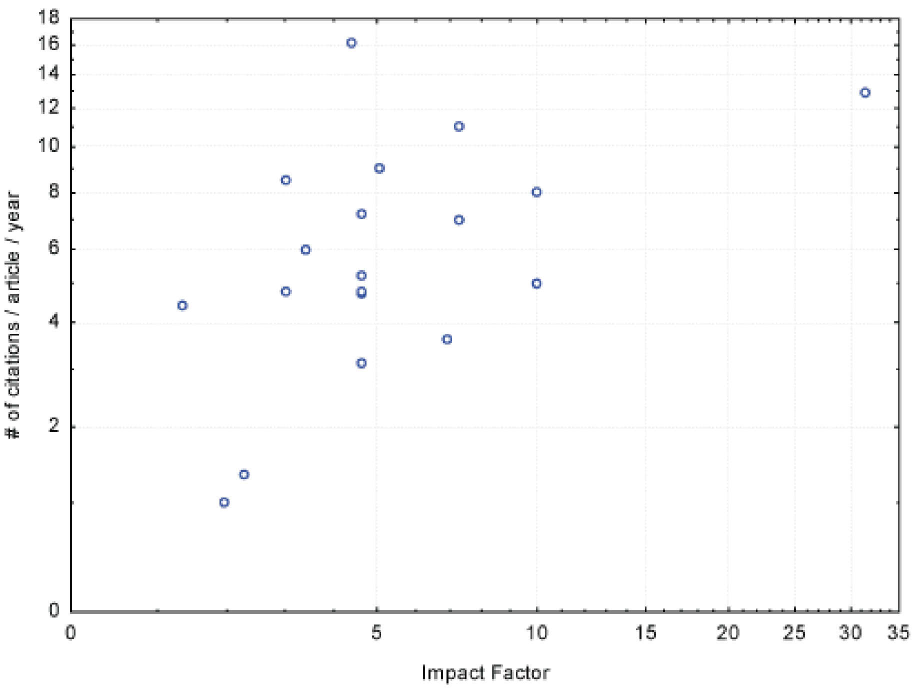
Today, however, journal rank is used for much more than just filtering the paper deluge. Among the half-dozen or so ranking schemes, one de facto monopolist has emerged which dictates journal rank:Thomson Reuters' Impact Factor (IF). At many scientific institutions, funders and governing bodies are using the IF to rank the content of the journals as well: if it has been published in a high-ranking journal, it must be good science, or so the seemingly plausible argument goes. Thus, today, scientific careers are made and broken by the editors at high-ranking journals.

As a scientist today, it is very difficult to find employment if you cannot sport publications in high-ranking journals. In the increasing competition for the coveted spots, it is starting to be difficult to find employment with only few papers in high-ranking journals: a consistent record of 'high-impact' publications is required if you want science to be able to put food on your table. Subjective impressions appear to support this intuitive notion: isn't a lot of great research published in Science and Nature while we so often find horrible work published in little-known journals? Isn't it a good thing that in times of shinking budgets we only allow the very best scientists to continue spending taxpayer funds?

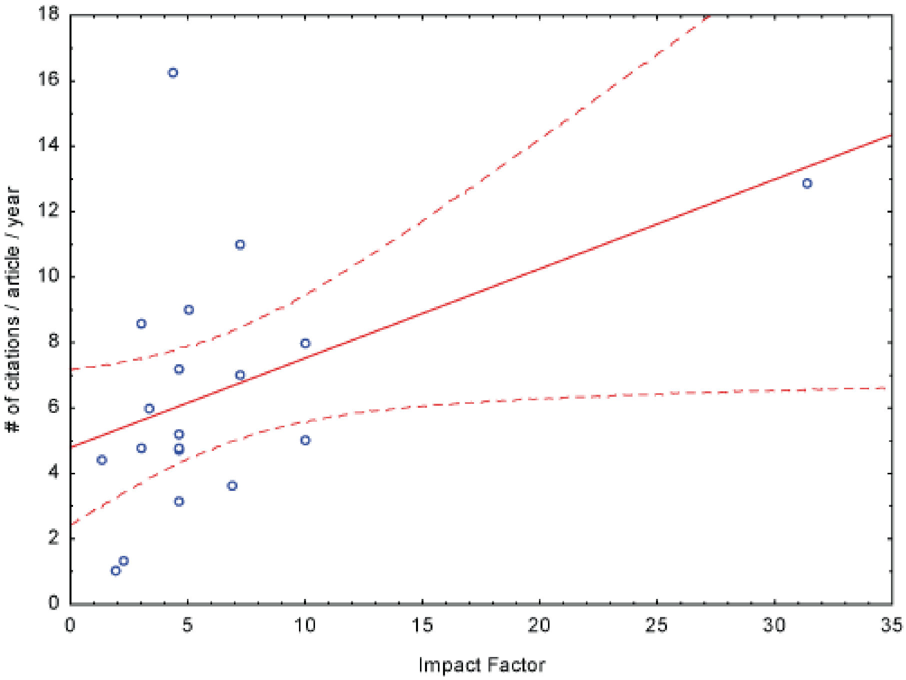


In any area of science, intuition, plausibility and subjective judgment may be grounds for designing experiments, but scientists all the time have to change their subjective judgment or discard their pet hypothesis if the data don't support them. Data, not subjective judgment is the basis for sound science. In this vein, many tests of the predictive power of journal rank have been carried out. One of them was published in 1997 in the British Medical Journal. One figure in this paper shows four examples of researchers whose publications had been plotted with their annual citations against the IF.

As can be seen from the R values at the top left of each graph, the correlation between impact factor and actual citations is not all that great. However, these are only four examples. Maybe this would be different for other researchers? In the absence of any easily available dataset where IFs and citations of individuals are compiled, I just took my own publications (according to Google Citations), looked up the current IFs and plotted them in the same way:

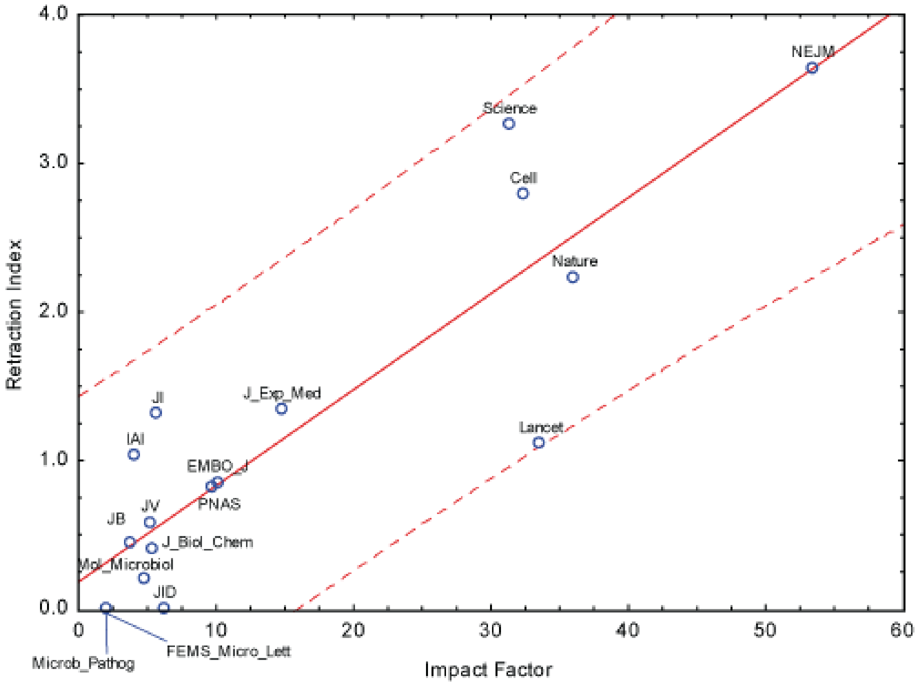


The R value for this correlation is 0.55, so pretty much in the range of the published values. I have also done a linear regression on these data, which provides me with a somewhat more meaningful metric, the "R squared" value, or the Coefficient of Determination. For my data, the adjusted value for this coefficient is less than 0.3, a very weak measure of a correlation, suggesting that the predictive power between IF and citations, at least for my publications, is not very strong, despite it being statistically significant ($p<0.004$). Here is a linear plot of the same data as above:



Our Science paper stands out as an outlier on the far right and often such outliers tend to artificially skew regressions. Confirming the interpretations so far, removing the Science paper from the analysis renders the regression non-significant (adjusted R2: 0.016, p=0.275). Thus, with the available data (to me) so far, there seems to be little reason to expect highly-cited research in high-ranking journals. In fact, our most frequently cited paper is smack in the middle of the IF scale.

More recently, there was another publication assessing the predictive power of journal rank. This time, the authors built on the notion that the pressure ot publish in high-ranking journals. If your livelihood depends on this Science/Nature paper, doesn't the pressure increase to maybe forget this one crucial control experiment, or leave out some data points that don't quite make the story look so nice? After all, you know your results are solid, it's only cosmetics which are required to make it a top-notch publication! Of course, in science there never is certainty, so such behavior will decrease the reliability of the scientific reports being published. And indeed, together with the decrease in tenured positions, the number of retractions has increased at about 400-fold the rate of publication increase. The authors of this study, Fang and Casadevall, were so nice to provide me with access to their data so I could compile the same kind of regression analysis I did for my own publications:



This already looks like a much stronger correlation than the one between IF and citations. How do the critical values measure up? The regression is highly significant at $p<0.000003$, with a coefficient of determination at a whopping 0.77. Thus, at least with the current data, IF indeed seems to be a more reliable predictor of retractions than of actual citations. How can this be, given that the IF is supposed to be a measure of citation rate for each journal? There are many reasons why this argument falls flat, but here are the three most egregious ones:

The IF is negotiable and doesn't reflect actual citation counts;

The IF cannot be reproduced, even if it reflected actual citations;

The IF is not statistically sound, even if it were reproducible and reflected actual citations.

In other words, all the organizations that require scientists to publish in 'high-impact journals' at the same time require them to publish in 'high-retraction journals'. I wonder if requiring publication in high-retraction journals can be good for science?

33.

The widely held notion that high-impact publications determine who gets academic jobs, grants and tenure is wrong. Stop using it as an excuse.

By Michael Eisen | Published: February 4, 2012

– ◊ –

In response to my previous post on boycotting non-OA journals, my friend Gavin Sherlock made the following comment:

I laud what you are doing, and you have changed the world of publishing forever for the better. However, I was specifically told by my chair that I need a Nature or Science paper to make my tenure packet bulletproof, so you shouldn't underestimate the tenure argument.

This comment pretty much sums up why closed access publishing still dominates. Like most scientists, Gavin agrees that the system we have is bad, and that progress towards open access is a good thing. But, in the face of advice that he needs a glamour mag paper for his tenure package his pre-tenure facebook feed was filled with "Just submitted a paper to Nature" and "Just submitted a paper to Science".

I am not here to criticize Gavin. These few indiscretions not withstanding, he has a long and exemplary history of open access publication. Rather I use him as an example of just how powerful and toxic the glamour mag culture in science has become. If it can get to him, it can get to anyone.

I am also not here to dwell on how crude a measure of impact the impact factor is [1], or how the tyranny of the impact factor is destroying science. Peter Lawrence (see also this list put together by Sean Eddy) has written extensively and eloquently on the subject.

Rather I want to challenge the key assumption – made by nearly everyone – that choosing not to publish your work in the highest impact factor journal you can convince to accept it is tantamount to career suicide. It is ubiquitously repeated by everyone from the most successful senior scientists to first year graduate students. And, judging by their publishing practices, most of them must believe it to be true. But I don't think it is.

Before I explain, I should note that my comments will deal exclusively with science in the United States. We have, mercifully, not followed the incredibly misguided policies used in many European and Asian countries which use formula that explicitly include impact factor to allocate jobs and money. The underlying attitude may be as strong here, but at least it is not hard-coded.

I can not deny that there is a very strong correlation between the impact factor of the journals in which someone has published and their success in landing jobs, grants and tenure. The evidence is all around me: 11 of the 15 assistant professors in my department at Berkeley had published at least one Science, Nature or Cell paper as a graduate student or postdoctoral fellow (we'll return to the other four later). And more systematic studies have found a similar correlation [2].

But, as we know, correlation does not imply causation. Even if hiring, grant review and tenure committees completely ignored journal titles and focused exclusively on the quality of the science (as they should), we would still expect there to be a strong correlation between success and impact factor. Scientists are so conditioned to believe that impact factors matter that most design their experiments, write their papers and jostle with editors to get their work into the "best" journal possible. Since the peer reviewers who ultimately make (or at least strongly influence) the publishing decisions are drawn from the same pool of scientists who make hiring, funding and tenure decisions, it is no surprise that the same work is valued in all of these venues. Thus, the idea that impact factors are paramount would be a self-fulfilling prophesy even if it were completely untrue!

Of course it is not completely untrue. I have seen too many colleagues lazily use the presence or absence of SNC publications as the primary factor in screening job applicants, as a reason to or not to fund a grant application, and as a proxy for whether someone should or should not be tenured. It is also undoubtedly true that, all other things being equal, high impact publications can make a difference. However, glamour mag publications are neither necessary (see the 4 assistant professors discussed above), nor sufficient (we routinely pass on candidates with SNC publications).

Encouraging the people we train to focus so exclusively on journal titles as the determinant of their success downplays the many other factors that play into these decisions: letters of recommendation, how effectively they communicate in person, and, most importantly, the inherent quality of their science. Sure, reviewers sometimes take shortcuts, but the quality of the underlying science and candidate matter a lot – and in most cases are paramount.

My own lab provides several examples that demonstrate this reality. My graduate students have gone on to great postdocs and many have landed prestigious fellowships “despite” having only published in open access journals. More curiously, I have had four postdoctoral fellows go out onto the academic job market, who all got great jobs: at Wash U., Wisconsin, Idaho and Harvard Medical School. Not only did none of them have glamour mag publications from my lab. None of them had yet published the work on the basis of which they were hired! They got their interviews on the basis of my letters and their research statements, and got the jobs because they are great scientists who had done outstanding, as of yet unpublished, work. If anything demonstrates the fallacy of the glamour mag or bust mentality this is it.

So, when I suggest that we all refuse to publish in non-open access journals, I am not being cavalier about the career prospects of the next generation. I don't suggest we abandon them to the winds of fate. Rather I believe we can simultaneous do right by science, by the public AND by our trainees by explaining to them what is at stake, pointing out the holes in the prevailing wisdom they hear from all sides, and then explaining and defending their actions to the hilt when we write letters on their behalf.

Scientific publishing is broken, and it's dragging down the field. We can either sit by and do nothing, allowing another generation to be captured by the allure of high impact publications. Or we can show some courage, shake off this silly dogma, and lead the next generation to a place that will be better for all concerned. You know what I choose. Please join me.

UPDATE: I want to reemphasize my central point. Getting jobs, grants and tenure is a competitive process in which the quality of an individual scientists previous work and future plans are evaluated. Getting a paper published a competitive process in which the quality of a piece of work and its potential impact is evaluated. It is no surprise that the results of these two processes are correlated. But it is a logical fallacy of the highest degree to conclude from this correlation that it is the journal in which your work gets published, rather than its inherent merits, that plays the dominant role in determining your success in science.

In spite of this perfectly reasonable (and I believe correct) alternative hypothesis, the scientific establishment, and most scientists, continue to reinforce the idea that one must always grope for the highest impact factor journal. Given that this leads to so many negative consequences for science – encouraging glamour over rigor, slowing scientific progress by delaying publication while papers bounce from journal to journal, and massively inhibiting the much-needed transformation from subscription-based to open access publication – it is absolutely essential that we not only fail to act on its precepts, but that we challenge its underlying assumptions, highlight empirical evidence that counters it, and otherwise do whatever we can to eradicate this deeply cynical and highly destructive mentality from our field.

34.

Do not resuscitate: the journal impact factor declared dead

Brendan Crabb
The Conversation, 21 May 2013

– ◊ –

Science is a highly competitive business so measuring the impact of scientific research, meaningfully and objectively, is essential. The journal impact factor (JIF) has emerged over the past few decades as the most used scientific metric for the assessment of research quality.

As a research scientist, Medical Research Institute Director and former Editor-in-Chief of a scientific journal, I have to confess my own way too common use of the JIF, and to delight when that particular parameter fell in my favour.

But the truth is the JIF has major flaws.

There are better ways to gauge the impact of a piece of research, the quality of an individual researcher and even the quality of a peer-review journal. The San Francisco Declaration on Research Assessment attempts to formally address the deficiencies in the JIF measurement and proposes the adoption of different practises in assessing quality of research publications.

More on the declaration later – but first, some discussion of the JIF. What is it and what's the problem with using it?

Working the JIF

Impact factors were first devised by American scientist Eugene Garfield in 1955. The current JIF system is a measure of how frequently recently published papers from a particular journal are cited (referenced in another body of work).

Hence, it is said to be a measure of the "impact" of the research published in that journal.

Technically, it is the number times in any given year that papers published in the previous two years were cited. For example, a journal with an impact factor of 10 in 2012 means the papers published in that journal in 2010 and 2011 were cited an average of 10 times each in 2012.

On the face of it, this should be a good measure of the scientific quality of a journal, but even in this regard the JIF has only limited value.

The JIF can be greatly skewed by an extraordinarily highly cited individual paper. It also does not take account of the different sizes of particular scientific disciplines nor the fact that review articles tend to get cited more often than primary research articles.

These and other deficiencies mean that the JIF is not only a blunt metric when it comes to assessing the quality of a journal but it is open to manipulation. Journals may decide to publish on certain topics or certain article types (such as reviews) to maximise their JIF.

So what are the options?

While JIF may have some limited value in assessing journal quality, for the individual or individual piece of research the JIF is even less reliable.

In the context of peer-reviewed publications at least, what's important is the frequency with which others cite that individual's publication, preferably in a way that considers the variability of the size of different scientific fields.

One day, it even may prove to be significant to give "bonus marks" for individuals publishing highly cited papers in low impact journals. Such metrics are being developed and are becoming more commonplace.

One example is the h-index, proposed by physicist Jorge E. Hirsch in 2005, which is a metric that links both the number of papers an author has published with the number of times those papers have been cited. While emerging metrics themselves have deficiencies (for example your h-index generally gets better as you get older!) they are part of an important trend.

Looking beyond citations

The Declaration on Research Assessment is signed by an impressive list of influential individual scientists and organisations as well as the editors-in-chief of many major journals, including Science.

To some extent the declaration states what many in research in this country already know and have begun applying to their judgements of individual researchers and of the quality of scientific studies. However, the declaration is probably that clear line in the sand that the scientific world needed in relation to JIFs.

It is the next development that will be the most interesting: a time where we look beyond simple publication metrics to judge more fully the impact of a piece of research. For example:

what resources (online or otherwise) were produced as a result of the work?

how many people accessed these resources?

what impact did the work have on policy and practice?

what was the economic, social or environmental benefit of the work?

Therefore, the significance of the San Francisco Declaration on Research Assessment needs to be seen not simply as announcing the death of the JIF, but also as a step along a pathway to a more enlightened method of assessing the impact of research.

35.

Impact factor: Time to move on

K. Satyanarayana & Anju Sharma

Indian J Med Res 127, January 2008, pp 4-6

– ◊ –

"Not everything that counts can be counted,
and not everything that can be counted counts".
Sign hanging in Einstein's office in Princeton University.

Ever since its discovery in 1960s, impact factor (IF) continues to be in the news, often for the wrong reasons. Both its proponents (we suspect, those mostly outside the business of creating innovative knowledge) and the increasing voices arguing for a more rational use to wholesale rejection (mostly practicing scientists or policy makers, especially from the Continent) are equally vocal. Not many parameters can match IF for its propensity for sustained misuse and/or abuse. Maybe due to the glamour and simplicity of use, IF and other citation-based indices are the only 'numbers' widely used in seemingly unbiased decision making. This despite these data being understood by so few and of dubious value in several contexts of evaluation. We would like to focus only on the IF and argue for the need for a serious rethink toward a better and more credible alternative.

Despite the occasional cautionary word by its 'inventor' Eugene Garfield, citation data continue to be used to assess science whether it is the impact of individual articles, journals, researchers, research departments, and even countries. But to our mind, the most worrying facet is the possible inappropriate application in critical decision making in research policy and support systems, as IF and other citation-based parameters can potentially skew the direction of research. Scientists tend to research into fundable, popular, mainstream areas and not necessarily the research questions that should be addressing. Worse still, researches who have outstanding track records in unfashionable but relevant subjects may not get adequate funding or recognition. It is well known that funding follows what is considered 'significant' in science which is usually science driven by citation hype and publications in high impact journals. Even in India funding allocations follow global (read American) trends, and, to that extent, the research agenda continues to be set on flawed thinking.

Right at the beginning of their careers, researchers get fastened with the citation nose-band with the IF virtually dictating their future career path. Young scientists who drive the research enterprise would like to be 'known', which could happen only if they publish in high impact journals and get cited widely. For the huge number of researchers from countries like India, China, South Korea, etc. which export bulk of scientific workforce to the US, such a paper could well be the visa for entry into a prestigious laboratory. Sadly, this is just the beginning of the unending rat race on the citation treadmill.

Journal publishers and editors are perhaps among the worst culprits. They need to lure the best brains to write for them to sell their product in an increasingly competitive market. There cannot be a cheaper, easier and more effective means of publicity for a journal than the prominent display of its high IF, some highly cited papers with the authors' photographs thrown in. Authors just love impact factors, and adulation. To attract good papers and to retain subscriptions, journal editors also join the bandwagon. As the PLoS Medicine admitted candidly, IF does matter to editors. Even to us at the IJMR. For journals from the developing world, it is doubly hard to stay on the citation treadmill. There is a very stiff entry barrier to the Web of Science (WoS) databases with a huge number of journals trying to join the elite club. Once in, journals try all means at disposal to increase citation counts.

But the dice are heavily loaded against journals from the developing countries. The almost pathological reluctance of Indian scientists to cite their own earlier papers from Indian journals when they publish abroad can only match the near contemptuous indifference of journals with high IF to their poor cousins from developing countries. At the other end, there appears to be a cozy arrangement between the publishers of 'big' journals like Nature, just to name one, and the Thomson ISI. While a huge number of journals are denied entry for years into the WoS, new journals spawn from some publishing groups with IF right from inception!

A major criticism leveled by scientists on citation based indices like IF is that all citations are treated equal, irrespective of the context of citation and the publishing journal. IF therefore cannot distinguish between journals with varying quality of content and the 'importance' of the quote. The argument is that by merely counting the frequency of citations per article and disregarding the prestige of the citing journals, the impact factor becomes merely a metric of 'popularity', not of 'prestige'.

But even from early 1970s there have been serious attempts to overcome the limitations of citation analysis and find appropriate alternatives. As early as 1976, Gabriel Pinski and Francis Narin had proposed a recursive impact factor and tried to compute and analyse citation data to give citations from journals that have high impact greater weight than citations from lowimpact journals. Largely thanks to increasing webbased access and use of scholarly literature, several others have developed and proposed innovative methods and tools to rank scholarly journals and refine the evaluation of both science and scientists both within and outside the citation-based systems. Some of these include Page Rank, weighed Page Rank, h-index, gfactor, Y- factor, EuroFactor and Faculty of 1000, briefly described below.

The h-index quantifies both the actual scientific productivity and the apparent scientific impact of a scientist based on the set of the scientist's most quoted papers and the number of citations that they have received in other people's publications. The index can also be applied to study the productivity and impact of a group of scientists, such as a department, university or country. The g-factor attempts quantifying the scientific productivity of scientists and is calculated based on the distribution of citations received by a given researcher's publications. The EuroFactor or the European Journal Quality Factor was designed following algorithm for analyzing the biometric relation between European journals from the EuroFactor database. This was developed primarily to address the under-representation of European journals in the Thomson-ISI databases.

PageRank™ is a software system for ranking web pages developed by Google and has also been used to rank publications and later in the system of weighted PageRank. The advantage with these tools is that they use a broad range of open data sources from the Google Scholar (GS) etc. which generally results in more comprehensive coverage of cited literature than the WoS, an expensive subscription-based service with about 6000 journals. The PageRank algorithm makes a clear distinction between popularity and expert appreciation or prestige of published research. Popular journals are those that are cited frequently by journals with little prestige. These journals therefore may have a very high ISI IF and a very low weighted PageRank. Prestigious journals, on the other hand are those not frequently cited, but their citations come from highly prestigious journals. These journals have a very low ISI IF and a very high weighted PageRank. Thus the weighted version of the popular PageRank algorithm can be used to obtain a metric that reflects prestige. Analysis of journals according to their ISI IF and their weighted PageRank shows significant overlaps and differences. The Y-factor is a simple combination of both the ISI IF and the weighted PageRank. Significantly, the authors claim that the resulting journal rankings correspond well to a general understanding of journal status. For example, while the IF ranking lists five review journals, the Y-factor column had none. Two primary research journals Cell and the Proceedings of the National Academy of Sciences USA, rated highly by peers figure in the Y-factor list.

Peer ranking of research papers outside the citation number game has also been tried and a prominent one being the Faculty of 1000. Faculty of 1000 Medicine comprises a select set of carefully chosen experts currently estimated to be over 2400 in various areas of medical research and clinical practice. The service is divided into 18 specialties and over 200 sub-specialties. Each

Member is expected to read and comment on one or two interesting articles every month. The final F1000 Factor is consensual incorporating the ratings it receives and the number of times it is selected by different Faculty Members. Outstanding work thus gets its deserved peer recognition irrespective and independent of citation counts.

Despite the criticism, IF is perhaps here to stay. Both the International Society for Scientometrics and Informetrics (1993) and the journal Scientometrics (1978) are going strong. In all the major scientific agencies like the Council of Scientific and Industrial Research (CSIR), Department of Science & Technology (DST), Indian Council of Medical Research (ICMR), etc. IF and citations get weightage in recruitments, promotions, rewards and other recognitions. What is more, even the Office of the Principal Scientific Advisor to the Prime Minister of India sponsors studies on the ‘citation health’ of the Indian science. Happily, so far the Indian Government does not have a cash-for-highlycited-paper system prevalent in Pakistan, China, South Korea, etc.

The increasing application of citation-based data for recruitments and funding decisions are especially a cause of concern. In common with other countries, some experts and academic administrators who take critical decisions in India are remarkably ignorant about citation-based indicators like the IF, especially potential distortions. Admittedly, it is tough to define the quality of an academic publication using only non-quantifiable factors, such as its potential influence on the next generation of scientists. But citation-based data should at best be used to supplement peer judgement as done by the UK Research Assessment Exercises. It is about time that we seriously look for measures beyond the IF. As a starter, studies could be commissioned on the limitations of IF to come out with India-specific measures to evaluate science using open data sources like the Google Scholar. And also look at the current funding patterns vis-a-vis our priorities. The Journal would be pleased to carry this debate forward.

References

1.http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor.

2. Garfield E. The history and meaning of impact factor. JAMA 2006; 295 : 90-3.

3. Bordons M, Fernández T, Gomez I. Advantages and limitations in the use of impact factor measures for the assessment of research performance. Scientometrics 2002; 53 : 195-206.

4. Braun T, Schubert A. Indicators of research output in the sciences from 5 central European countries, 1990-1994. Scientometrics 1996; 38 : 145-65.

5. Anon. The impact factor game. It is time to find better way to assess the scientific literature. PloS Med 2006 June 3 (6) e291.

6. Satyanarayana K, Jain NC. IF is on the rise. Indian J Med Res 2006; 123 : 717-8.

7. Bollen J, Rodrigues MA, Van de Sompel H. Journal Status. Scientometrics December 2006. Available from:

http://arxiv.org/PS_cache/cs/pdf/0601/0601030v1.pdf.

8. Pinski G, Narin F. Citation influence for journal aggregates of scientific publications: theory, with application to the literature of physics. Inf Process Manage 1976; 12 : 297-312.

9. Hirsch JE. An index to quantify an individual's scientific research output. PNAS 2005; 102 : 16569-72.

10. Egghe L. Theory and practice of the g-index. Scientometrics 2006; 69 : 131-52.

11. Euro-Factor. Available from: http://www.univ-lille1.fr/lea/Menu_du_Site/Publications/Acrobat/VICEREUROFACTOR.pdf.

12. Faculty of 1000. http://www.f1000 medicine.com/about/view.

13. Warner J. Citation analysis and Research Assessment in the United Kingdom. Bull Am Soc Infor Science Technol 2003; October/November, p. 26-8.

36.

Nine reasons why Impact Factors fail and using them may harm science

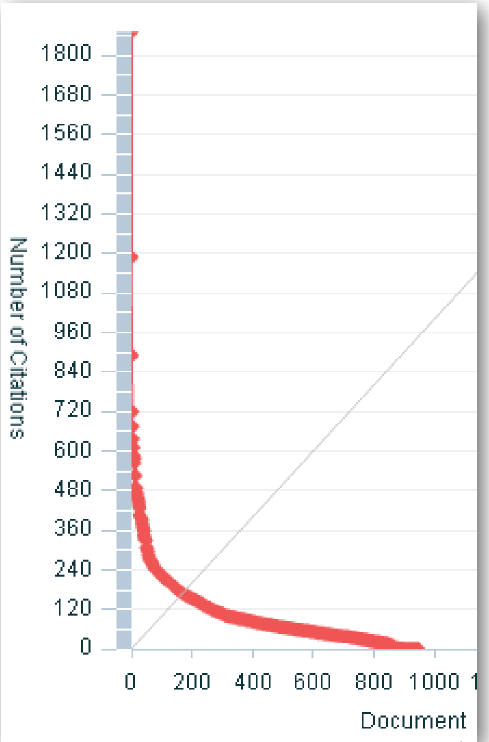
Jeroen Bosman
I&M / I&O 2.0, Universiteitsbibliotheek Utrecht, 03/11/2013

– ◊ –

Over the past few months I have come across many articles and posts highlighting the detrimental effect of the enormous importance scholars attach to Impact Factors. I feel many PhDs and other researchers want to break out of the IF rat race, but obviously without risking their careers. It is good to see that things are changing, slowly but surely. It seems a good moment to succinctly sum up what is actually wrong with Impact Factors.

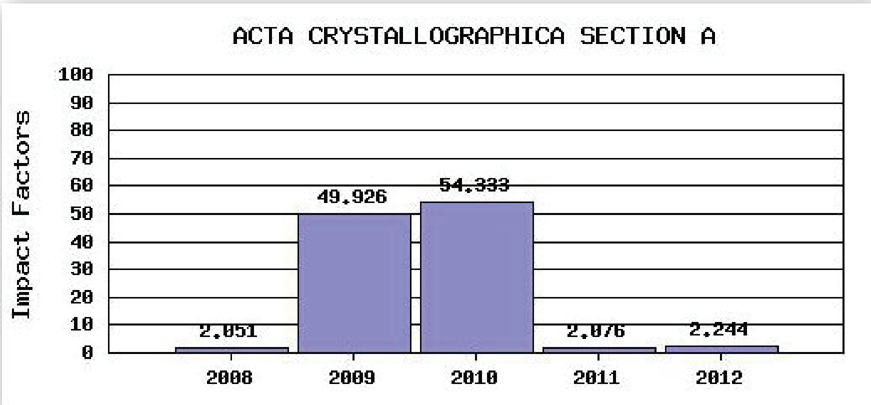
1) Using Impact Factors to judge or show the quality of individual papers or the authors of those papers is a clear ecological fallacy. In bibliometrics normal distributions are very rare and consequently mean values as the IF are weak descriptors. This certainly holds for citation distributions of articles published in a journal. They are often very skewed with a small number of very highly cited articles. The citation distributions of high IF journals are probably even more skewed.

You simply cannot say that this or that paper is better because it has been published in a high IF journal. For decades now the producers of the impact factors and all bibliometricians have warned against using journal averages for judging single papers or authors. In the Journal Citation Reports it reads "You should not depend solely on citation data in your journal evaluations. Citation data are not meant to replace informed peer review. Careful attention should be paid to the many conditions that can influence citation rates such as language, journal history and format, publication schedule, and subject specialty." Judging papers by their IF is actually judging the ability of papers and authors to get accepted by editors and to pass peer review. That ability is not all there is to say about the quality or importance of a paper and it is simply to small a basis for decisions on tenure, promotion, grants etc.. Recently there has been evidence that with the introduction of online journals and search engines the correlation between IF and a paper's citations is weakening [id. at ArXiv].



A typical journal citation distribution: Citations in 2011-2013 to Nature articles published in 2010 (made 2013.11.03 from Scopus data)

[Update 20131106: How outliers may influence impact factors is examplified by the extreme fluctuations of the IF of Acta Crystallographica A over the recent years. The article "A short history of SHELX" was cited 10,ooos of times (perhaps also becuase it explicitly asks to do so in the abstract), thereby skyrocketing the journals IF. Authors of average or mediocre papers in the 2009 and 2010 volumes must have been pleasantly surprised to be able to mention IFs of 49 and 54! In less extreme forms this happens with many journals.]



Citation outliers may cause extreme IF fluctuations

2) The use of impact factors creates and sustains a double Matthew effect. The disproportial attention for high IF journals makes that the same paper accepted by such a journal will get far more eyeballs and will even receive more citations than it would have received in other journals. Related to this is the effect that authors and asessors deem a paper better just because it was published in a journal with high IF or rejection rates and may think that citing a paper from high IF journals is the safe way to go to avoid comments from referees. Thus, part of the citations these papers receive are a free ride based on the visibility and citability of the well known journal brands in which they are published.

3) Impact Factors calculated by Thomson Reuters (formerly ISI) are only available for a minority of all journals. Only 10,853 of the approximately 30,000 scholarly journals are included in the journal citation reports. Arts and humanities are left out as are the majority of non-English journals. It is very difficult for a journal to

join the club. That might be acceptable if it was just based on stringent quality control. But inclusion in the Journal Citation Reports in which Impact Factors are published is also based on the number of times they are cited in journals that are already included. But meeting that criterion is adversely affected by the fact that authors like to cite papers from (high) IF journals. As long as your journal is out it is difficult to get in. By only taking papers seriously if they are published in journals with an Impact Factor we hinder the free development of science.

Related to this is the US/UK bias of inclusion. Historically ISI/Thomson built the Science Citation Index on US learned society journals, followed by UK journals and later also the vast number of English languages journals from commercial publishers, also including Elsevier from the Netherlands and Springer from Germany. But most journals from other areas of the world were left out. In ISI/Thomson terminology, the English language and mostly US/UK based publications are "international journals" while all journals from the rest of the world are termed "regional journals'.

To be fair, it is good to see that recently Thomson has started to internationalize the journal base of their citation indexes, presumably due to competition by the rival Scopus that indexes vastly more journals (21,500). We now also have freely available Scopus-based metrics such as the SJR (status normalized, by SCImago) and the SNIP (field normalized, by CWTS). Although it is nice to have these alternatives for so many more journals, they share most of the issues described here surrounding the Impact Factors.

Of course, giving too much weight to Impact Factors gives an even more limited view of real relevance, importance or promise in fields were large shares of the output are not in journals at all, but in books, reports, code, software and more. For example, at Utrecht University in medicine output in term of numbers of items consisted for 96% of refereed journal articles, but in social science it is only 77% and in the humanities 32%.

4) Impact Factors are irreproducible and not transparent. It is not possible, using the Web of Science data, to reproduce im-

pact factors. What often is also not realized is that there is a difference between Impact Factor levels and average citation levels. That has to do with publication types. In the numerator of the Impact Factor citations to all publications types are counted, but in the denominator only articles and reviews are counted. That means that journals publishing lots of content other than articles and reviews, such as letters, notes etc. receive a free boost to their Impact Factors.

[Update 20131106: General science journals such as Nature and Science typically profit substantially from these free rides, because they publish a lot of material (letters etc.) that gathers citations for their numerator, but that is not counted in their IF denominator]

[Update 20131105: The lack of transparancy is clearly demonstrated by changing numbers of articles taken into account for the same journal in the same year in different JCR editions, as shown by Björn Brembs and by the negotiability and unclear process of assigning document types by Thomson, as shown in this PLoS Medicine editorial.]

5) Related to the previous point, Impact Factors are affected by the mix of publication types in a journal. In particular, the share of review articles positively affects the IF as these articles have very high citation rates. Even when publishing a regular primary research paper in a journal with many review articles, you benefit in a world were Impact Factors are seen as a proxy for your paper's quality or importance.

6) (Coercive) journal self citations are one of the more perverse effects of overly relying on Impact Factors. Journals want to compete and attract attention and citations and these are at least partially based on their IFs. Some journal editors bluntly ask their authors to cite content from the journal they publish in. If very gross and detected, Thomson Reuters will punish the journals by temporarily removing them from the Journal Citation Reports, but one wonders what goes undetected. In the 2012 edition no fewer than 65 journals were suppressed because of these kinds of 'anomalies'.

Another case is that of editorial self citations. These may in some cases account for over 20% of the Impact Factor.

7) There is a correlation, but no clear cut relation between citation numbers and paper quality. All citation metrics are biased by other factors. [Update 20131210: there is even no hard evidence for a correlation, also becuase of the vagueness of the 'quality' concept]. The most important other factor is perhaps visibility. For a paper to get cited it first of all needs to be encountered by potential citing authors. The chance of encounters depends on:

- findability (being indexed in WoS,. Scopus and Google Scholar a.o.)
- attractiveness of the title (e.g. titles mentioning main finding/result)
- language
- whether the journal is licensed (e.g. as part of a library big deal with main publishers)
- whether the journal has Open Access
- whether the journal has many personal subscriptions
- whether the paper has already been cited (mark that the results ranking in Google Scholar takes the number of citations received into account)
- promotion through social media and researcher profiling cites etc.

All these visibility factors play their role in citation chance and thus in the Impact Factor. They all have nothing to do with paper quality or importance.

There are more distortive factors that render the perception that IFs reflect journal quality untenable. We need to menion citation sentiment. Papers may be cited for lots of reasons and not all citations are endorsements.

8) Impact Factors do not facilitate cross-discipline comparison and indirectly may make interdisciplinary work less attractive [id. at ArXiv]. Publication and citation cultures are very diverse across disciplines. The average number of citations per paper and the distributions of citations over time make that average

impact factors of fields vary enormously. While the two year citation window used in the standard Impact Factor may capture the peak or bulk of citations to papers in molecular biology, it certainly does not in mathematics or geology. Using a five year window solves that to a certain extent, but the problem of different average numbers of references per paper in the various fields remains. When researchers use IF in their selection of journals to submit their paper to, it may even prevent them from doing interdisciplinary work with field having lower IFs, the more so because interdisciplinary journals are often younger and still building up their reputation. Using SNIP or Eigenfactor metrics, both field normalized, may redress this particular problem, but not most of the other issues.

9) Impact factors have a long delay. Impacts Factors reflect citation to papers published in print between 1.5 and 4.5 years ago (as the Journal Citation report are published in June and reflect citation made in the previous year to papers published in the two years preceding that year.. The papers themselves may even be much older, because of the time lag between online and print publication and the time that passes between finishing the manuscript and publication. By the way: long time lags between online and print publication distort citation figures. Papers from journals with these long gaps have more time to gather citations. The problem with this delay is that it tells us next to nothing about the attractiveness of more recent papers. Article level metrics such as downloads and take up by other media reflected in Altmetrics do have this capability, but have their own problems of interpretation.

This list of issues shows that although there are some inherent problems in the Impact Factor, most problems are caused by using it to support decisions that it is not intended to support and by the perverse effects.

IFs were made for librarians to support journal subscription decisions [Update 20131106: Actually the first intended use was for ISI itself: to now which journals should be added to the Journal Scietation Index]. Citation indexes were made to discover related literature, follow discussions or trace new research topics.

Not even in his wildest dreams did Eugene Garfield envision researchers quoting IFs on their homepages or in tenure, promotion and grant applications or over a glass of beer or wine in a bar. By the way, even for librarians Impact Factors are becoming less useful, because journals are more and more selected by package/publisher (whatever you may think of that), because many journals lack an Impact Factor, and because many journals are now Open Access. Speaking of librarians: they play an important role as they are often involved in information literacy courses teaching on how to select quality publications and how to evaluate resources. I truly hope they will help lead us away from Impact Factor madness and focus on teaching skills to evaluate content in stead of status and paint the information landscape beyond IF journals.

All this is not new. But in the last decade the focus on Impact Factors and their perverse effects has reached such a degree as to hinder free development of science, causing frustration and costing time and money. But things may change. The Research Excellence Framework (REF) in the UK has already stated that using rankings or metric in evaluation is no longer allowed and the Australian ERA may do the same. Many societies and thousands of scholars have signed the San Francisco declaration on research assessment (DORA), that states: “Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions”. There is increasing evidence of adverse effects of reliance on rankings, but there are no easy solutions. New movements of scholars advocating fundamental reorientation, such as Science in Transition in the Netherlands, are gaining momentum. The intention is not to throw away all that we have but to reorient our tools and habits to make science even more valuable.

37.

Why not use the Journal Impact Factor: a Seven Point Primer

Iddo @ Byte Size Biology, October 11th, 2013

– ◊ –

After a series of tweets and a couple of Facebook posts about the problems of the Journal Impact Factor (JIF), I was approached by a colleague who asked me: “so why are you obsessed with this”? My answer was that it irks me that I have to use the JIF next to my publications in so many different reports (grant reports, university annual activities, proposals, etc.) since it is a bad metric to evaluate the merit of my papers, and as a scientist, I do not like using bad metrics.

I assume that many readers of my blog constitute the proverbial choir on which my preaching would be wasted. Specifically, those who understand what the Thomson-Reuters Journal Impact Factor is, and how became such a poorly-understood and overused and abused metric. However, for those who have no idea what I am talking about, or for those who are thinking “what is wrong with the Impact Factor”? this post would hopefully be informative, if not valuable. It is a brief post. There was a lot written about the JIF, and the plausible alternatives that can be used to assess journal quality and impact, and I provide a list of further reading material sources at the end. Finally, for those who, like me, think that there are many wrong things with this metric and its use, and would like to convey that information, I hopefully provide some basic arguments.

What is the Journal Impact Factor?

The JIF is supposed to be a proxy for a journal's impact: i.e. how much influence the journal has in the scientific community. It is calculated as follows:

A = number of times that articles published in the journal in years X-1 and X-2 from this journal were cited in year X

B = number of citable items in the journal in years X-1 and X-2

JIF(X) = A/B

Seems simple enough. The ratio of the number of citations to the number of publications. The higher the ratio, the more the articles are being cited. Therefore, the journal's impact is higher.

Why is the JIF a bad metric? There are several reasons.

I. The JIF does not measure what it's supposed to measure

1. The distribution of citations over articles in highly biased

An editorial in Nature from 2005 stated that:

For example, we have analysed the citations of individual papers in Nature and found that 89% of last year's figure was generated by just 25% of our papers.

Furthermore:

The most cited Nature paper from 2002–03 was the mouse genome, published in December 2002. That paper represents the culmination of a great enterprise, but is inevitably an important point of reference rather than an expression of unusually deep mechanistic insight. So far it has received more than 1,000 citations. Within the measurement year of 2004 alone, it received 522 citations. Our next most cited paper from 2002–03 (concerning the functional organization of the yeast proteome) received 351 citations that year. Only 50 out of the roughly 1,800 citable items published in those two years received more than 100 citations in 2004. The great majority of our papers received fewer than 20 citations. (Emphases added my me).

So there are few papers with very high impact factor. The distribution of different citations per papers is enormously skewed. Not that one would expect that all research published in a high-impact journal would be the same, but with such a skewed distribution it is evident that only a small minority of papers contribute to the journal's impact factor. So it is not the research that is published on the whole in Nature that makes it so impactful. It is a handful of high impact papers that granted Nature its high impact factor. A recent analysis by Björn Brembs and colleagues shows that this is pretty much the situation all around. Even more so: the higher the overall JIF, the more skewed the citations-per-paper distribution. So each year, the impact factor of high-impact journal is supported by a few "superstar" papers. Using a mean as a measure of central tendency in a skewed distribution is meaningless. (Get it? Meaningless. Haha. Groan.)

2. The JIF is a negotiated, irreproducible metric

Look at the equation describing how JIF is calculated. Note that the definition of B, the denominator, is "citable items". If follows that the lower B is, the higher a journal's overall impact factor. So the value of B is important for the overall impact factor. How, exactly, is the number of citable items in each journal determined? Yeah, good luck with answering that. From the editors of PLoS Medicine:

During discussions with Thomson Scientific over which article types in PLoS Medicine the company deems as "citable," it became clear that the process of determining a journal's impact factor is unscientific and arbitrary. After one in-person meeting, a telephone conversation, and a flurry of e-mail exchanges, we came to realize that Thomson Scientific has no explicit process for deciding which articles other than original research articles it deems as citable. We conclude that science is currently rated by a process that is itself unscientific, subjective, and secretive.

During the course of our discussions with Thompson Scientific, PLoS Medicine's potential impact factor— based on the same articles published in the same year—seesawed between as much as 11 (when only research articles are entered into the denominator) to less than 3 (when almost all article types in the magazine section are included, as Thomson Scientific had initially done).

So the JIF not an objectively determined metric. Rather, it is determined by negotiation and committee. Not the kind of metric you would use to objectively assess anything.

3. The JIF can be inflated by delays between online and print publications.

Many articles have two publication dates: upon acceptance, the article gets published in the journal's online site. This is known as a "pre-publication" or "prepub". The second date can be a few weeks or months later, when the publication coincides with the journal's print date. The latter is often used as the "official" date of publication. However, the article has existed for a while before the official date, gaining readership and perhaps citations. An article by Tort and colleagues shows that this lag is a highly significant contributor to a journal's JIF.

II. The JIF does not measure what people think it should measure

4. The JIF is abused as a surrogate measure and as a predictive metric

The JIF is commonly used as a measure by hiring committees, promotion and tenure committees and some grant review panels to predict a scientist's success. The logic in doing so is as follows: if she publishes in a "good place", then her research has merit, and will have impact in the scientific community. Furthermore, if she published in good places, she will continue to do so. Those assessing institutions have good reason to need a predictive measure: universities and grant agencies invest money in their faculty, and would like to know if these scientists would be successful. But the JIF is a poor metric for assessing, let alone predicting, an individual scientist's impact. In fact, many scientific councils urge universities, institutes and grant agencies NOT to use the JIF for these purposes. Moreover, Thomson-Reuters themselves maintain:

In the case of academic evaluation for tenure it is sometimes inappropriate to use the impact of the source journal to estimate the expected frequency of a recently published article.

Again, the impact factor should be used with informed peer review. Citation frequencies for individual articles are quite varied.

That is a very qualified statement, and since Thomson-Reuters have a vested interest in the use of the JIF in the community, I cannot blame them. I contend that it is totally inappropriate and unscientific to use the JIF of publications as a predictor for an individual's impact, now or in the future. That is simply not what the JIF measures!

Furthermore, JIF is a poor predictor for individual article impact. Since the distribution of citations-per-paper is so skewed, it follows that one cannot predict how well a paper will be cited based on the impact factor of the journal it was accepted in. Yet people continue to assess article importance based on "where it got into". This type of assessment as a surrogate measure is poor practice, and it biases reader's impression as to the merit of the article.

III. The "Impact Factor culture" is bad for science and science publication

5. JIF bias is exacerbated by poor editorial policies

Some journals limit the number of citations that can be used in an article. This encourages a more subtle form of bad science writing practice. The author, faced with a limit on the number of possible cited items, would tend to cite review articles rather than the original research articles. The reason being that one review article may cover material in several research articles. Consequently, this inflates the number of citations review articles get, and "steals citations" from research articles. As a result, journals which publish review articles (either fully or partially) get a higher impact factor. As it is, review articles are more highly cited than research articles, but limitations on number of citations exacerbate this situation. The dynamic I described renders the JIF even more unreliable, favoring review articles over research articles.

6. Papers get retracted more from high JIF ranking journals

Again, referring to Brembs's paper: JIF is a statistical predictor for the retraction rate in a journal. In other words, the higher a journal's JIF, the higher the frequency of papers that are retracted from that journal. Brembs does not report on why that is. There are probably several contributing factors: a high-profile publication gets to be read ans scrutinized more; competitors in the field may be "out to get you". But also, there may be pressure to publish high and therefore "cut corners" when performing research. We don't know whether retractions from high ranking journals are due to a higher fraction of poor papers than in low ranking journals, or whether that is because the papers in high ranking journals tend to be scrutinized more carefully be more people. Be that as it may, the correlation of JIF and retraction rate is disturbing.

7. I suspect that the extraordinary importance placed on the JIF delays scientific progress

Because of the strong incentives to publish in high-impact journals, researchers can sequester their findings for years, delaying actual communication of their research. It can take an extraordinarily long amount of time to publish findings in a high-impact journal. These journals have very low acceptance rates, so a paper can get delayed for years while the research and papers get revised and resubmitted over and over. Even worse, research would get tailored towards what is perceived as impactful and "sexy". Another problem, raised recently by the editor in chief of Science, Bruce Alberts, is the overcrowding of fashionable fields:

But perhaps the most destructive result of any automated scoring of a researcher's quality is the "me-too science" that it encourages. Any evaluation system in which the mere number of a researcher's publications increases his or her score creates a strong disincentive to pursue risky and potentially groundbreaking work, because it takes years to create a new approach in a new experimental context, during which no publications should be expected. Such metrics further block innovation because they encourage scientists to work in areas of science that are already highly populated, as it is only in these fields that large numbers of scientists can be expected to reference one's work, no matter how outstanding.

Caveat: I could not find research supporting these points, and I really only wrote it based on personal experience and conversations with colleagues. If anyone out there knows of supporting research (how would you even begin is an interesting question), let me know.

Concluding with an excerpt from an editorial by Kai Simmons, published in 2008, unfortunately still true today:

There are no numerical shortcuts for evaluating research quality. What counts is the quality of a scientist's work wherever it is published. That quality is ultimately judged by scientists, raising the issue of the process by which scientists review each others' research. However, unless publishers, scientists, and institutions make serious efforts to change how the impact of each individual scientist's work is determined, the scientific community will be doomed to live by the numerically driven motto, "survival by your impact factors."

38.

Deep impact: unintended consequences of journal rank

Björn Brembs, Katherine Button and Marcus Munafò
Front. Hum. Neurosci., 24 June 2013

– ◊ –

Most researchers acknowledge an intrinsic hierarchy in the scholarly journals ("journal rank") that they submit their work to, and adjust not only their submission but also their reading strategies accordingly. On the other hand, much has been written about the negative effects of institutionalizing journal rank as an impact measure. So far, contributions to the debate concerning the limitations of journal rank as a scientific impact assessment tool have either lacked data, or relied on only a few studies. In this review, we present the most recent and pertinent data on the consequences of our current scholarly communication system with respect to various measures of scientific quality (such as utility/citations, methodological soundness, expert ratings or retractions). These data corroborate previous hypotheses: using journal rank as an assessment tool is bad scientific practice. Moreover, the data lead us to argue that any journal rank (not only the currently-favored Impact Factor) would have this negative impact. Therefore, we suggest that abandoning journals altogether, in favor of a library-based scholarly communication system, will ultimately be necessary. This new system will use modern information technology to vastly improve the filter, sort and discovery functions of the current journal system.

Introduction

Science is the bedrock of modern society, improving our lives through advances in medicine, communication, transportation, forensics, entertainment and countless other areas. Moreover, today's global problems cannot be solved without scientific input and understanding. The more our society relies on science, and the more our population becomes scientifically literate, the more important the reliability [i.e., veracity and integrity, or, “credibility” (Ioannidis, 2012)] of scientific research becomes. Scientific research is largely a public endeavor, requiring public trust. Therefore, it is critical that public trust in science remains high. In other words, the reliability of science is not only a societal imperative, it is also vital to the scientific community itself. However, every scientific publication may in principle report results which prove to be unreliable, either unintentionally, in the case of honest error or statistical variability, or intentionally in the case of misconduct or fraud. Even under ideal circumstances, science can never provide us with absolute truth. In Karl Popper's words: “Science is not a system of certain, or established, statements” (Popper, 1995). Peer-review is one of the mechanisms which have evolved to increase the reliability of the scientific literature.

At the same time, the current publication system is being used to structure the careers of the members of the scientific community by evaluating their success in obtaining publications in high-ranking journals. The hierarchical publication system (“journal rank”) used to communicate scientific results is thus central, not only to the composition of the scientific community at large (by selecting its members), but also to science's position in society. In recent years, the scientific study of the effectiveness of such measures of quality control has grown.

Retractions and the Decline Effect

A disturbing trend has recently gained wide public attention: The retraction rate of articles published in scientific journals, which had remained stable since the 1970's, began to increase rapidly in the early 2000's from 0.001% of the total to about 0.02% (Figure 1A). In 2010 we have seen the creation and popularization of a website dedicated to monitoring retractions (http://retractionwatch.com), while 2011 has been described as the "the year of the retraction" (Hamilton, 2011). The reasons suggested for retractions vary widely, with the recent sharp rise potentially facilitated by an increased willingness of journals to issue retractions, or increased scrutiny and error-detection from online media. Although cases of clear scientific misconduct initially constituted a minority of cases (Nath et al., 2006; Cokol et al., 2007; Fanelli, 2009; Steen, 2011a; Van Noorden, 2011; Wager and Williams, 2011), the fraction of retractions due to misconduct has risen sharper than the overall retraction rate and now the majority of all retractions is due to misconduct (Steen, 2011b; Fang et al., 2012).

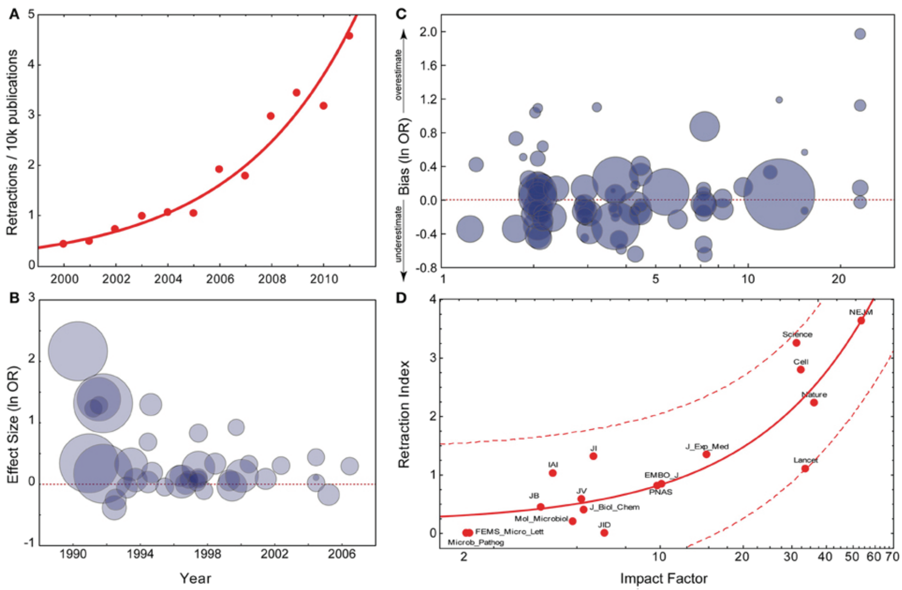


Figure 1. Current trends in the reliability of science. (A) Exponential fit for PubMed retraction notices (data from pmretract.heroku.com). (B) Relationship between year of publication and individual study effect size. Data are taken from Munafò et al. (2007), and represent candidate gene studies of the association between DRD2 genotype and alcoholism. The effect size (y-axis) represents the individual study effect size (odds ratio; OR), on a log-scale. This is plotted

against the year of publication of the study (x-axis). The size of the circle is proportional to the IF of the journal the individual study was published in. Effect size is significantly negatively correlated with year of publication. (C) Relationship between IF and extent to which an individual study overestimates the likely true effect. Data are taken from Munafò et al. (2009), and represent candidate gene studies of a number of gene-phenotype associations of psychiatric phenotypes. The bias score (y-axis) represents the effect size of the individual study divided by the pooled effect size estimated indicated by meta-analysis, on a log-scale. Therefore, a value greater than zero indicates that the study provided an over-estimate of the likely true effect size. This is plotted against the IF of the journal the study was published in (x-axis), on a log-scale. The size of the circle is proportional to the sample size of the individual study. Bias score is significantly positively correlated with IF, sample size significantly negatively. (D) Linear regression with confidence intervals between IF and Fang and Casadevall's Retraction Index (data provided by Fang and Casadevall, 2011).

Retraction notices, a metric which is relatively easy to collect, only constitute the extreme end of a spectrum of unreliability that is inherent to the scientific method: we can hardly ever be entirely certain of our results (Popper, 1995). Much of the training scientists receive aims to reduce this uncertainty long before the work is submitted for publication. However, a less readily quantified but more frequent phenomenon (compared to rare retractions) has recently garnered attention, which calls into question the effectiveness of this training. The "decline-effect," which is now well-described, relates to the observation that the strength of evidence for a particular finding often declines over time (Simmons et al., 1999; Palmer, 2000; Møller and Jennions, 2001; Ioannidis, 2005b; Møller et al., 2005; Fanelli, 2010; Lehrer, 2010; Schooler, 2011; Simmons et al., 2011; Van Dongen, 2011; Bertamini and Munafo, 2012; Gonon et al., 2012). This effect provides wider scope for assessing the unreliability of scientific research than retractions alone, and allows for more general conclusions to be drawn.

Researchers make choices about data collection and analysis which increase the chance of false-positives (i.e., researcher bias) (Simmons et al., 1999; Simmons et al., 2011), and surprising and novel effects are more likely to be published than studies showing no effect. This is the well-known phenomenon of publication bias (Song et al., 1999; Møller and Jennions, 2001; Callaham, 2002; Møller et al., 2005; Munafò et al., 2007; Dwan et al., 2008; Young et al., 2008; Schooler, 2011; Van Dongen, 2011). In other words, the probability of getting a paper published might be biased toward larger initial effect sizes, which are revealed by later studies to be not so large (or even absent entirely), leading to the decline effect. While sound methodology can help reduce researcher bias (Simmons et al., 1999), publication bias is more difficult to address. Some journals are devoted to publishing null results, or have sections devoted to these, but coverage is uneven across disciplines and often these are not particularly high-ranking or well-read (Schooler, 2011; Nosek et al., 2012). Publication therein is typically not a cause for excitement (Giner-Sorolla, 2012; Nosek et al., 2012), leading to an overall low frequency of replication studies in many fields (Kelly, 2006; Carpenter, 2012; Hartshorne and Schachner, 2012; Makel et al., 2012; Yong, 2012). Publication bias is also exacerbated by a tendency for journals to be less likely to publish replication studies (or, worse still, failures to replicate) (Curry, 2009; Goldacre, 2011; Sutton, 2011; Editorial, 2012; Hartshorne and Schachner, 2012; Yong, 2012). Here we argue that the counter-measures proposed to improve the reliability and veracity of science such as peer-review in a hierarchy of journals or methodological training of scientists may not be sufficient.

While there is growing concern regarding the increasing rate of retractions in particular, and the unreliability of scientific findings in general, little consideration has been given to the infrastructure by which scientists not only communicate their findings but also evaluate each other as a potential contributing factor. That is, to what extent does the environment in which science takes place contribute to the problems described above? By far the most common metric by which publications are evaluated, at least initially, is the perceived prestige or rank of the journal in which they appear. Does the pressure to publish in prestigious, high-ranking journals contribute to the unreliability of science?

The Decline Effect and Journal Rank

The common pattern seen where the decline effect has been documented is one of an initial publication in a high-ranking journal, followed by attempts at replication in lower-ranked journals which either failed to replicate the original findings, or suggested a much weaker effect (Lehrer, 2010). Journal rank is most commonly assessed using Thomson Reuters' Impact Factor (IF), which has been shown to correspond well with subjective ratings of journal quality and rank (Gordon, 1982; Saha et al., 2003; Yue et al., 2007; Sønderstrup-Andersen and Sønderstrup-Andersen, 2008). One particular case (Munafò et al., 2007) illustrates the decline effect (Figure 1B), and shows that early publications both report a larger effect than subsequent studies, and are also published in journals with a higher IF. These observations raise the more general question of whether research published in high-ranking journals is inherently less reliable than research in lower-ranking journals.

As journal rank is also predictive of the incidence of fraud and misconduct in retracted publications, as opposed to other reasons for retraction (Steen, 2011a), it is not surprising that higher ranking journals are also more likely to publish fraudulent work than lower ranking journals (Fang et al., 2012). These data, however, cover only the small fraction of publications that have been retracted. More important is the large body of the literature that is not retracted and thus actively being used by the scientific community. There is evidence that unreliability is higher in high-ranking journals as well, also for non-retracted publications: A meta-analysis of genetic association studies provides evidence that the extent to which a study over-estimates the likely true effect size is positively correlated with the IF of the journal in which it is published (Figure 1C) (Munafò et al., 2009). Similar effects have been reported in the context of other research fields (Ioannidis, 2005a; Ioannidis and Panagiotou, 2011; Siontis et al., 2011).

There are additional measures of scientific quality and in none does journal rank fare much better. A study in crystallography reports that the quality of the protein structures described is significantly lower in publications in high-ranking journals (Brown and Ramaswamy, 2007). Adherence to basic principles of sound scientific (e.g., the CONSORT guidelines: http://www.consort-statement.org), or statistical methodology have also been tested. Four different studies on levels of evidence in medical and/or psychological research have found varying results. While two studies on surgery journals found a correlation between IF and the levels of evidence defined in the respective studies (Obremskey et al., 2005; Lau and Samman, 2007), a study of anesthesia journals failed to find any statistically significant correlation between journal rank and evidence-based medicine principles (Bain and Myles, 2005) and a study of seven medical/psychological journals found highly varying adherence to statistical guidelines, irrespective of journal rank (Tressoldi et al., 2013). The two surgery studies covered an IF range between 0.5 and 2.0, and 0.7 and 1.2, respectively, while the anesthesia study covered the range 0.8–3.5. It is possible that any correlation at the lower end of the scale is abolished when higher rank journals are included. The study by Tressoldi and colleagues, which included very high ranking journals, supports this interpretation. Importantly, if publications in higher ranking journals were methodologically sounder, then one would expect the opposite result: inclusion of high-ranking journals should result in a stronger, not a weaker correlation. Further supporting the notion that journal rank is a poor predictor of statistical soundness is our own analysis of data on statistical power in neuroscience studies (Button et al., 2013). There was no significant correlation between statistical power and journal rank (N = 650; rs = −0.01; t = 0.8; Figure 2). Thus, the currently available data seem to indicate that journal rank is a poor indicator of methodological soundness.

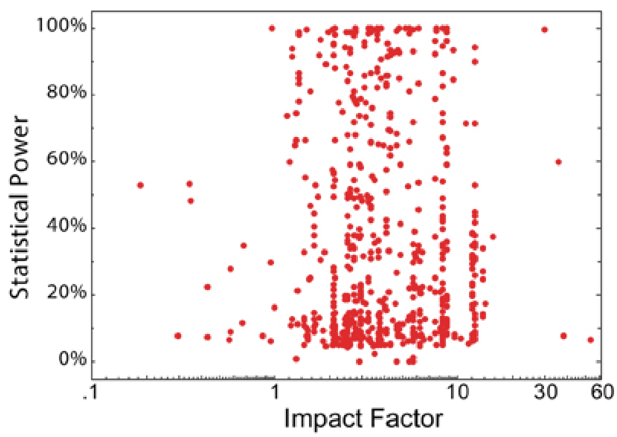


Figure 2. No association between statistical power and journal IF. The statistical power of 650 neuroscience studies (data from Button et al., 2013; 19 missing ref; 3 unclear reporting; 57 published in journal without 2011 IF; 1 book) plotted as a function of the 2011 IF of the publishing journal. The studies were selected from the 730 contributing to the meta-analyses included in Button et al. (2013), Table 1, and included where journal title and IF (2011 © Thomson Reuters Journal Citation Reports) were available.

Beyond explicit quality metrics and sound methodology, reproducibility is at the core of the scientific method and thus a hallmark of scientific quality. Three recent studies reported attempts to replicate published findings in preclinical medicine (Scott et al., 2008; Prinz et al., 2011; Begley and Ellis, 2012). All three found a very low frequency of replication, suggesting that maybe only one out of five preclinical findings is reproducible. In fact, the level of reproducibility was so low that no relationship between journal rank and reproducibility could be detected. Hence, these data support the necessity of recent efforts such as the "Reproducibility Initiative" (Baker, 2012) or the "Reproducibility Project" (Collaboration, 2012). In fact, the data also indicate that these projects may consider starting with replicating findings published in high-ranking journals.

Given all of the above evidence, it is therefore not surprising that journal rank is also a strong predictor of the rate of retractions (Figure 1D) (Liu, 2006; Cokol et al., 2007; Fang and Casadevall, 2011).

Social Pressure and Journal Rank

There are thus several converging lines of evidence which indicate that publications in high ranking journals are not only more likely to be fraudulent than articles in lower ranking journals, but also more likely to present discoveries which are less reliable (i.e., are inflated, or cannot subsequently be replicated). Some of the sociological mechanisms behind these correlations have been documented, such as pressure to publish (preferably positive results in high-ranking journals), leading to the potential for decreased ethical standards (Anderson et al., 2007) and increased publication bias in highly competitive fields (Fanelli, 2010). The general increase in competitiveness, and the precariousness of scientific careers (Shapin, 2008), may also lead to an increased publication bias across the sciences (Fanelli, 2011). This evidence supports earlier propositions about social pressure being a major factor driving misconduct and publication bias (Giles, 2007), eventually culminating in retractions in the most extreme cases.

That being said, it is clear that the correlation between journal rank and retraction rate is likely too strong (coefficient of determination of 0.77; data from Fang and Casadevall, 2011) to be explained exclusively by the decreased reliability of the research published in high ranking journals. Probably, additional factors contribute to this effect. For instance, one such factor may be the greater visibility of publications in these journals, which is both one of the incentives driving publication bias, and a likely underlying cause for the detection of error or misconduct with the eventual retraction of the publications as a result (Cokol et al., 2007). Conversely, the scientific community may also be less concerned about incorrect findings published in more obscure journals. With respect to the latter, the finding that the large majority of retractions come from the numerous lower-ranking journals (Fang et al., 2012) reveals that publications in lower ranking journals are scrutinized and, if warranted, retracted. Thus, differences in scrutiny are likely to be only a contributing factor and not an exclusive explanation, either. With respect to the former, visibility effects in general can be quantified by measuring citation rates between journals, testing the assumption that if higher visibility were a contributing factor to retractions, it must also contribute to citations.

Journal Rank and Study Impact

Thus far we have presented evidence that research published in high-ranking journals may be less reliable compared with publications in lower-ranking journals. Nevertheless, there is a strong common perception that high-ranking journals publish "better" or "more important" science, and that the IF captures this well (Gordon, 1982; Saha et al., 2003). The assumption is that high-ranking journals are able to be highly selective and publish only the most important, novel and best-supported scientific discoveries, which will then, as a consequence of their quality, go on to be highly cited (Young et al., 2008). One way to reconcile this common perception with the data would be that, while journal rank may be indicative of a minority of unreliable publications, it may also (or more strongly) be indicative of the importance of the majority of remaining, reliable publications. Indeed, a recent study on clinical trial meta-analyses found that a measure for the novelty of a clinical trial's main outcome did correlate significantly with journal rank (Evangelou et al., 2012). Compared to this relatively weak correlation (with all coefficients of determination lower than 0.1), a stronger correlation was reported for journal rank and expert ratings of importance (Allen et al., 2009). In this study, the journal in which the study had appeared was not masked, thus not excluding the strong correlation between subjective journal rank and journal quality as a confounding factor. Nevertheless, there is converging evidence from two studies that journal rank is indeed indicative of a publication's perceived importance.

Beyond the importance or novelty of the research, there are three additional reasons why publications in high-ranking journals might receive a high number of citations. First, publications in high-ranking journals achieve greater exposure by virtue not only of the larger circulation of the journal in which they appear, but also of the more prominent media attention (Gonon et al., 2012). Second, citing high-ranking publications in one's own publication may increase its perceived value. Third, the novel, surprising, counter-intuitive or controversial findings often published in high-ranking journals, draw citations not only from follow-up studies but also from news-type articles in scholarly journals reporting and discussing the discovery. Despite these four factors, which would suggest considerable effects of journal rank on future citations, it has been established for some time that the actual effect of journal rank is measurable, but nowhere near as substantial as indicated (Seglen, 1994, 1997; Callaham, 2002; Chow et al., 2007; Kravitz and Baker, 2011; Hegarty and Walton, 2012; Finardi, 2013) and as one would expect if visibility were the exclusive factor driving retractions. In fact, the average effect sizes roughly approach those for journal rank and unreliability, cited above.

The data presented in a recent analysis of the development of these correlations between IF-based journal rank and future citations over the period from 1902–2009 (with IFs before the 1960's computed retroactively) reveal two very informative trends (Figure 3, data from (Lozano et al., 2012). First, while the predictive power of journal rank remained very low for the entire first two thirds of the twentieth century, it started to slowly increase shortly after the publication of the first IF data in the 1960's. This correlation kept increasing until the second interesting trend emerged with the advent of the internet and keyword-search engines in the 1990's, from which time on it fell back to pre-1960's levels until the end of the study period in 2009. Overall, consistent with the citation data already available, the coefficient of determination between journal rank and citations was always in the range of ~0.1 to 0.3 (i.e., quite low). It thus appears that indeed a small but significant correlation between journal rank and future citations can be observed. Moreover, the data suggest that most of this small effect stems from visibility effects due to the influence of the IF on reading habits (Lozano et al., 2012), rather than from factors intrinsic to the published articles (see data cited above). However, the correlation is so weak that it cannot alone account for the strong correlation between retractions and journal rank, but instead requires additional factors, such as the increased unreliability of publications in high ranking journals cited above. Supporting these weak correlations between journal rank and future citations are data reporting classification errors (i.e., whether a publication received too many or too few citations with regard to the rank of the journal it was published in) at or exceeding 30% (Starbuck, 2005; Chow et al., 2007; Singh et al., 2007; Kravitz and Baker, 2011). In fact, these classification errors, in conjunction with the weak citation advantage, render journal rank practically useless as an evaluation signal, even if there was no indication of less reliable science being published in high ranking journals.

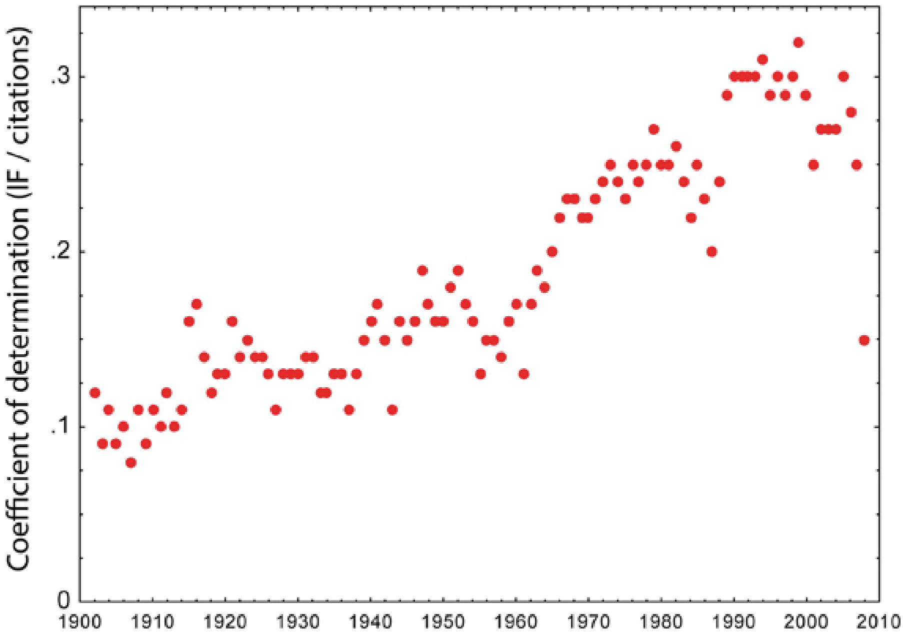


Figure 3. Trends in predicting citations from journal rank. The coefficient of determination (R2) between journal rank (as measured by IF) and the citations accruing over 2 years after publications is plotted as a function of publication year in a sample of almost 30 million publications. Lozano et al. (2012) make the case that one can explain the trends in the predictive value of journal rank by the publication of the IF in the 1960's (R2 increase is accelerating) and the widespread adoption of internet searches in the 1990's (R2 is dropping). The data support the interpretation that reading habits drive the correlation between journal rank and citations more than any inherent quality of the articles. IFs before the invention of the IF have been retroactively computed for the years before the 1960's.

The only measure of citation count that does correlate strongly with journal rank (negatively) is the number of articles without any citations at all (Weale et al., 2004), supporting the argument that fewer articles in high-ranking journals go unread. Thus, there is quite extensive evidence arguing for the strong correlation between journal rank and retraction rate to be mainly due to two factors: there is direct evidence that the social pressures to publish in high ranking journals increases the unreliability, intentional or not, of the research published there. There is more indirect evidence, derived mainly from citation data, indicating that increased visibility of publications in high ranking journals may potentially contribute to increased error-detection in these journals. With several independent measures failing to provide compelling evidence that journal rank is a reliable predictor of scientific impact or quality, and other measures indicating that journal rank is at least equally if not more predictive of low reliability, the central role of journal rank in modern science deserves close scrutiny.

Practical Consequences of Journal Rank

Even if a particular study has been performed to the highest standards, the quest for publication in high-ranking journals slows down the dissemination of science and increases the burden on reviewers, by iterations of submissions and rejections cascading down the hierarchy of journal rank (Statzner and Resh, 2010; Kravitz and Baker, 2011; Nosek and Bar-Anan, 2012). A recent study seems to suggest that such rejections eventually improve manuscripts enough to yield measurable citation benefits (Calcagno et al., 2012). However, the effect size of such resubmissions appears to be of the order of 0.1 citations per article, a statistically significant but, in practical terms, negligible effect. This conclusion is corroborated by an earlier study which failed to find any such effect (Nosek and Bar-Anan, 2012). Moreover, with peer-review costs estimated in excess of 2.2 billion € (US$ ~2.8b) annually (Research Information Network, 2008), the resubmission cascade contributes to the already rising costs of journal rank: the focus on journal rank has allowed corporate publishers to keep their most prestigious journals closed-access and to increase subscription prices (Kyrillidou et al., 2012), creating additional barriers to the dissemination of science. The argument from highly selective journals is that their per-article cost would be too high for author processing fees, which may be up to 37,000€ (US$ 48,000) for the journal Nature (House of Commons, 2004). There is also evidence from one study in economics suggesting that journal rank can contribute to suppression of interdisciplinary research (Rafols et al., 2012), keeping disciplines separate and isolated.

Finally, the attention given to publication in high-ranking journals may distort the communication of scientific progress, both inside and outside of the scientific community. For instance, the recent discovery of a "Default-Mode Network" in rodent brains was, presumably, made independently by two different sets of neuroscientists and published only a few months apart (Upadhyay et al., 2011; Lu et al., 2012). The later, but not the earlier, publication (Lu et al., 2012) was cited in a subsequent high-ranking publication (Welberg, 2012). Despite both studies largely reporting identical findings (albeit, perhaps, with different quality), the later report has garnered 19 citations, while the earlier one only 5, at the time of this writing. We do not know of any empirical studies quantitatively addressing this particular effect of journal rank. However, a similar distortion due to selective attention to publications in high-ranking journals has been reported in a study on medical research. This study found media reporting to be distorted, such that once initial findings in higher-ranking journals have been refuted by publications in lower ranking journals (a case of decline effect), they do not receive adequate media coverage (Gonon et al., 2012).

Impact Factor—Negotiated, Irreproducible, and Unsound

The IF is a metric for the number of citations to articles in a journal (the numerator), normalized by the number of articles in that journal (the denominator). However, there is evidence that IF is, at least in some cases, not calculated but negotiated, that it is not reproducible, and that, even if it were reproducibly computed, the way it is derived is not mathematically sound. The fact that publishers have the option to negotiate how their IF is calculated is well-established—in the case of PLoS Medicine, the negotiation range was between 2 and about 11 (Editorial, 2006). What is negotiated is the denominator in the IF equation (i.e., which published articles which are counted), given that all citations count toward the numerator whether they result from publications included in the denominator or not. It has thus been public knowledge for quite some time now that removing editorials and News-and-Views articles from the denominator (so called "front-matter") can dramatically alter the resulting IF (Moed and Van Leeuwen, 1995, 1996; Baylis et al., 1999; Garfield, 1999; Adam, 2002; Editorial, 2005; Hernán, 2009). While these IF negotiations are rarely made public, the number of citations (numerator) and published articles (denominator) used to calculate IF are accessible via Journal Citation Reports. This database can be searched for evidence that the IF has been negotiated. For instance, the numerator and denominator values for Current Biology in 2002 and 2003 indicate that while the number of citations remained relatively constant, the number of published articles dropped. This decrease occurred after the journal was purchased by Cell Press (an imprint of Elsevier), despite there being no change in the layout of the journal. Critically, the arrival of a new publisher corresponded with a retrospective change in the denominator used to calculate IF (Table 1). Similar procedures raised the IF of FASEB Journal from 0.24 in 1988 to 18.3 in 1989, when conference abstracts ceased to count toward the denominator (Baylis et al., 1999).

Journal: current biology	Published items 2000	Published items 2001	Published items 2002	Sum published items	Citations in preceding 2 years	IF
JCR science edition 2002	504	528	n.c.	1032	7231	7.007
JCR science edition 2003	n.c.	300	334	634	7551	11.910

Most of the rise in IF is due to the reduction in published items.
Note the discrepancy between the number of items published in 2001 between the two consecutive JCR Science Editions.
*n.c.: year not covered by this edition. Raw data see **Figure A1**.*

Table 1. Thomson Reuters' IF calculations for the journal "Current Biology" in the years 2002/2003.

In an attempt to test the accuracy of the ranking of some of their journals by IF, Rockefeller University Press purchased access to the citation data of their journals and some competitors. They found numerous discrepancies between the data they received and the published rankings, sometimes leading to differences of up to 19% (Rossner et al., 2007). When asked to explain this discrepancy, Thomson Reuters replied that they routinely use several different databases and had accidentally sent Rockefeller University Press the wrong one. Despite this, a second database sent also did not match the published records. This is only one of a number reported errors and inconsistencies (Reedijk, 1998; Moed et al., 1996).

It is well-known that citation data are strongly left-skewed, meaning that a small number of publications receive a large number of citations, while most publications receive very few (Seglen, 1992, 1997; Weale et al., 2004; Editorial, 2005; Chow et al., 2007; Rossner et al., 2007; Taylor et al., 2008; Kravitz and Baker, 2011). The use of an arithmetic mean as a measure of central tendency on such data (rather than, say, the median) is clearly inappropriate, but this is exactly what is used in the IF calculation. The International Mathematical Union reached the same conclusion in an analysis of the IF (Adler et al., 2008). A recent study correlated the median citation frequency in a sample of 100 journals with their 2-year IF and found a very strong correlation, which is expected due to the similarly left-skewed distributions in most journals (Editorial, 2013). However, at the time of this writing, it is not known if using the median (instead of the mean) improves any of the predominantly weak predictive properties of journal rank. Complementing the specific flaws just mentioned, a recent, comprehensive review of the bibliometric literature lists various additional shortcomings of the IF more generally (Vanclay, 2011).

Conclusions

While at this point it seems impossible to quantify the relative contributions of the different factors influencing the reliability of scientific publications, the current empirical literature on the effects of journal rank provides evidence supporting the following four conclusions: (1) journal rank is a weak to moderate predictor of utility and perceived importance; (2) journal rank is a moderate to strong predictor of both intentional and unintentional scientific unreliability; (3) journal rank is expensive, delays science and frustrates researchers; and, (4) journal rank as established by IF violates even the most basic scientific standards, but predicts subjective judgments of journal quality.

Caveats

While our latter two conclusions appear uncontroversial, the former two are counter-intuitive and require explanation. Weak correlations between future citations and journal rank based on IF may be caused by the poor statistical properties of the IF. This explanation could (and should) be tested by using any of the existing alternative ranking tools available (such as Thomson Reuters' Eigenfactor, Scopus' SCImagoJournalRank, or Google's Scholar Metrics etc.) and computing correlations with the metrics discussed above. However, a recent analysis shows a high correlation between these ranks, so no large differences would be expected (Lopez-Cozar and Cabezas-Clavijo, 2013). Alternatively, one can choose other important metrics and compute which journals score particularly high on these. Either way, since the IF reflects the common perception of journal hierarchies rather well (Gordon, 1982; Saha et al., 2003; Yue et al., 2007; Sønderstrup-Andersen and Sønderstrup-Andersen, 2008), any alternative hierarchy that would better reflect article citation frequencies might violate this intuitive sense of journal rank, as different ways to compute journal rank lead to different hierarchies (Wagner, 2011). Both alternatives thus challenge our subjective journal ranking. To put it more bluntly, if perceived importance and utility were to be discounted as indirect proxies of quality, while retraction rate, replicability, effect size overestimation, correct sample sizes, crystallographic quality, sound methodology and so on counted as more direct measures of quality, then inversing the current IF-based journal hierarchy would improve the alignment of journal rank for most and have no effect on the rest of these more direct measures of quality.

The subjective journal hierarchy also leads to a circularity that confounds many empirical studies. That is, authors use journal rank, in part, to make decisions of where to submit their manuscripts, such that well-performed studies yielding groundbreaking discoveries with general implications are preferentially submitted to high-ranking journals. Readers, in turn, expect only to read about such articles in high-ranking journals, leading to the exposure and visibility confounds discussed above and at length in the cited literature. Moreover, citation practices and methodological standards vary in different scientific fields, potentially distorting both the citation and reliability data. Given these confounds one might expect highly varying and often inconclusive results. Despite this, the literature contains evidence for associations between journal rank and measures of scientific impact (e.g., citations, importance, and unread articles), but also contains at least equally strong, consistent effects of journal rank predicting scientific unreliability (e.g., retractions, effect size, sample size, replicability, fraud/misconduct, and methodology). Neither group of studies can thus be easily dismissed, suggesting that the incentives journal rank creates for the scientific community (to submit either their best or their most unreliable work to the most high-ranking journals) at best cancel each other out. Such unintended consequences are well-known from other fields where metrics are applied (Hauser and Katz, 1998).

Therefore, while there are concerns not only about the validity of the IF as the metric of choice for establishing journal rank but also about confounding factors complicating the interpretation of some of the data, we find, in the absence of additional data, that these concerns do not suffice to substantially question our conclusions, but do emphasize the need for future research.

Potential Long-Term Consequences of Journal Rank

Taken together, the reviewed literature suggests that using journal rank is unhelpful at best and unscientific at worst. In our view, IF generates an illusion of exclusivity and prestige based on an assumption that it predicts scientific quality, which is not supported by empirical data. As the IF aligns well with intuitive notions of journal hierarchies (Gordon, 1982; Saha et al., 2003; Yue et al., 2007), it receives insufficient scrutiny (Frank, 2003) (perhaps a case of confirmation bias). The one field in which journal rank is scrutinized is bibliometrics. We have reviewed the pertinent empirical literature to supplement the largely argumentative discussion on the opinion pages of many learned journals (Moed and Van Leeuwen, 1996; Lawrence, 2002, 2007, 2008; Bauer, 2004; Editorial, 2005; Giles, 2007; Taylor et al., 2008; Todd and Ladle, 2008; Tsikliras, 2008; Adler and Harzing, 2009; Garwood, 2011; Schooler, 2011; Brumback, 2012; Sarewitz, 2012) with empirical data. Much like dowsing, homeopathy or astrology, journal rank seems to appeal to subjective impressions of certain effects, but these effects disappear as soon as they are subjected to scientific scrutiny.

In our understanding of the data, the social and psychological influences described above are, at least to some extent, generated by journal rank itself, which in turn may contribute to the observed decline effect and rise in retraction rate. That is, systemic pressures on the author, rather than increased scrutiny on the part of the reader, inflate the unreliability of much scientific research. Without reform of our publication system, the incentives associated with increased pressure to publish in high-ranking journals will continue to encourage scientists to be less cautious in their conclusions (or worse), in an attempt to market their research to the top journals (Anderson et al., 2007; Giles, 2007; Shapin, 2008; Munafò et al., 2009; Fanelli, 2010). This is reflected in the decline in null results reported across disciplines and countries (Fanelli, 2011), and corroborated by the findings that much of the increase in retractions may be due to misconduct (Steen, 2011b; Fang et al., 2012), and that much of this misconduct occurs in studies published high-ranking journals (Steen, 2011a; Fang et al., 2012). Inasmuch as journal rank guides the appointment and promotion policies of research institutions, the increasing rate of misconduct that has recently been observed may prove to be but the beginning of a pandemic: It is conceivable that, for the last few decades, research institutions world-wide may have been hiring and promoting scientists who excel at marketing their work to top journals, but who are not necessarily equally good at conducting their research. Conversely, these institutions may have purged excellent scientists from their ranks, whose marketing skills did not meet institutional requirements. If this interpretation of the data is correct, a generation of excellent marketers (possibly, but not necessarily, also excellent scientists) now serve as the leading figures and role models of the scientific enterprise, constituting another potentially major contributing factor to the rise in retractions.

The implications of the data presented here go beyond the reliability of scientific publications—public trust in science and scientists has been in decline for some time in many countries (Nowotny, 2005; European Commission, 2010; Gauchat, 2010), dramatically so in some sections of society (Gauchat, 2012), culminating in the sentiment that scientists are nothing more than yet another special interest group (Miller, 2012; Sarewitz, 2013). In the words of Daniel Sarewitz: “Nothing will corrode public trust more than a creeping awareness that scientists are unable to live up to the standards that they have set for themselves” (Sarewitz, 2012). The data presented here prompt the suspicion that the corrosion has already begun and that journal rank may have played a part in this decline as well.

Alternatives

Alternatives to journal rank exist—we now have technology at our disposal which allows us to perform all of the functions journal rank is currently supposed to perform in an unbiased, dynamic way on a per-article basis, allowing the research community greater control over selection, filtering, and ranking of scientific information (Hönekopp and Khan, 2011; Kravitz and Baker, 2011; Lin, 2012; Priem et al., 2012; Roemer and Borchardt, 2012; Priem, 2013). Since there is no technological reason to continue using journal rank, one implication of the data reviewed here is that we can instead use current technology and remove the need for a journal hierarchy completely. As we have argued, it is not only technically obsolete, but also counter-productive and a potential threat to the scientific endeavor. We therefore would favor bringing scholarly communication back to the research institutions in an archival publication system in which both software, raw data and their text descriptions are archived and made accessible, after peer-review and with scientifically-tested metrics accruing reputation in a constantly improving reputation system (Eve, 2012). This reputation system would be subjected to the same standards of scientific scrutiny as are commonly applied to all scientific matters and evolve to minimize gaming and maximize the alignment of researchers' interests with those of science [which are currently misaligned (Nosek et al., 2012)]. Only an elaborate ecosystem of a multitude of metrics can provide the flexibility to capitalize on the small fraction of the multi-faceted scientific output that is actually quantifiable. Such an ecosystem would evolve such that the only evolutionary stable strategy is to try and do the best science one can.

The currently balkanized literature, with a lack of interoperability and standards as one of its many detrimental, unintended consequences, prevents the kind of innovation that gave rise to the discover functions of Amazon or eBay, the social networking functions of Facebook or Reddit and course the sort and search functions of Google—all technologies virtually every scientist uses regularly for all activities but science. Thus, fragmentation and the resulting lack of access and interoperability are among the main underlying reasons why journal rank has not yet been replaced by more scientific evaluation options, despite widespread access to article-level metrics today. With an openly accessible scholarly literature standardized for interoperability, it would of course still be possible to pay professional editors to select publications, as is the case now, but after publication. These editors would then actually compete with each other for paying customers, accumulating track records for selecting (or missing) the most important discoveries. Likewise, virtually any functionality the current system offers would easily be replicable in the system we envisage. However, above and beyond replicating current functionality, an open, standardized scholarly literature would place any and all thinkable scientific metrics only a few lines of code away, offering the possibility of a truly open evaluation system where any hypothesis can be tested. Metrics, social networks and intelligent software then can provide each individual user with regular, customized updates on the most relevant research. These updates respond to the behavior of the user and learn from and evolve with their preferences. With openly accessible, interoperable literature, data and software, agents can be developed that independently search for hypotheses in the vast knowledge accumulating there. But perhaps most importantly, with an openly accessible database of science, innovation can thrive, bringing us features and ideas nobody can think of today and nobody will ever be capable of imagining, if we do not bring the products of our labor back under our own control. It was the hypertext transfer protocol (http) standard that spurred innovation and made the internet what it is today. What is required is the equivalent of http for scholarly literature, data and software.

Funds currently spent on journal subscriptions could easily suffice to finance the initial conversion of scholarly communication, even if only as long-term savings. One avenue to move in this direction may be the recently announced Episcience Project (Van Noorden, 2013). Other solutions certainly exist (Bachmann, 2011; Birukou et al., 2011; Kravitz and Baker, 2011; Kreiman and Maunsell, 2011; Zimmermann et al., 2011; Beverungen et al., 2012; Florian, 2012; Ghosh et al., 2012; Hartshorne and Schachner, 2012; Hunter, 2012; Ietto-Gillies, 2012; Kriegeskorte, 2012; Kriegeskorte et al., 2012; Lee, 2012; Nosek and Bar-Anan, 2012; Pöschl, 2012; Priem and Hemminger, 2012; Sandewall, 2012; Walther and Van den Bosch, 2012; Wicherts et al., 2012; Yarkoni, 2012), but the need for an alternative system is clearly pressing (Casadevall and Fang, 2012). Given the data we surveyed above, almost anything appears superior to the status quo.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments

Neil Saunders was of tremendous value in helping us obtain and understand the PubMed retraction data for Figure 1A. Ferric Fang and Arturo Casadeval were so kind as to let us use their retraction data to re-plot their figure on a logarithmic scale (Figure 1D). We are grateful to George A. Lozano, Vincent Larivière, and Yves Gingras for sharing their citation data with us (Figure 3). We are indebted to John Ioannidis, Daniele Fanelli, Christopher Baker, Dwight Kravitz, Tom Hartley, Jason Priem, Stephen Curry and three anonymous reviewers for their comments on an earlier version of this manuscript, as well as to Nikolaus Kriegeskorte and Erin Warnsley for their helpful comments and suggestions during the review process here. Marcus R. Munafo is a member of the UK Centre for Tobacco Control Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged. Björn Brembs was a Heisenberg-Fellow of the German Research Council during the time most of this manuscript was written and their support is gratefully acknowledged as well.

References

Adam, D. (2002). The counting house. Nature 415, 726–729. doi: 10.1038/415726a

Adler, N. J., and Harzing, A.-W. (2009). When knowledge wins: transcending the sense and nonsense of academic rankings. Acad. Manag. Learn. Edu. 8, 72–95. doi: 10.5465/AMLE.2009.37012181

Adler, R., Ewing, J., and Taylor, P. (2008). Joint Committee on Quantitative Assessment of Research: Citation Statistics (A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathemat. Available online at: http://www.mathunion.org/fileadmin/IMU/Report/CitationStatistics.pdf

Allen, L., Jones, C., Dolby, K., Lynn, D., and Walport, M. (2009). Looking for landmarks: the role of expert review and bibliometric analysis in evaluating scientific publication outputs. PLoS ONE 4:8. doi: 10.1371/journal.pone.0005910

Anderson, M. S., Martinson, B. C., and De Vries, R. (2007). Normative dissonance in science: results from a national survey of u.s. Scientists. JERHRE 2, 3–14. doi: 10.1525/jer.2007.2.4.3

Bachmann, T. (2011). Fair and open evaluation may call for temporarily hidden authorship, caution when counting the votes, and transparency of the full prepublication procedure. Front. Comput. Neurosci. 5:61. doi: 10.3389/fncom.2011.00061

Bain, C. R., and Myles, P. S. (2005). Relationship between journal impact factor and levels of evidence in anaesthesia. Anaesth. Intensive Care 33, 567–570.

Baker, M. (2012). Independent Labs to Verify High-Profile Papers. Nature News. Available online at: http://www.nature.com/doifinder/10.1038/nature.2012. doi: 10.1038/nature.2012.11176

Bauer, H. H. (2004). Science in the 21st century?: knowledge monopolies and research cartels. J. Scient. Explor. 18, 643–660.

Baylis, M., Gravenor, M., and Kao, R. (1999). Sprucing up one's impact factor. Nature 401, 322.

Begley, C. G., and Ellis, L. M. (2012). Drug development: raise standards for preclinical cancer research. Nature 483, 531–533. doi: 10.1038/483531a

Bertamini, M., and Munafo, M. R. (2012). Bite-size science and its undesired side effects. Perspect. Psychol. Sci. 7, 67–71. doi: 10.1177/1745691611429353

Beverungen, A., Bohm, S., and Land, C. (2012). The poverty of journal publishing. Organization 19, 929–938. doi: 10.1177/1350508412448858

Birukou, A., Wakeling, J. R., Bartolini, C., Casati, F., Marchese, M., Mirylenka, K., et al. (2011). Alternatives to peer review: novel approaches for research evaluation. Front. Comput. Neurosci. 5:56. doi: 10.3389/fncom.2011.00056

Brown, E. N., and Ramaswamy, S. (2007). Quality of protein crystal structures. Acta Crystallogr. D Biol. Crystallogr. 63, 941–950. doi: 10.1107/S0907444907033847

Brumback, R. A. (2012). "3.. 2.. 1.. Impact [factor]: target [academic career] destroyed!": just another statistical casualty. J. Child Neurol. 27, 1565–1576. doi: 10.1177/0883073812465014

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376. doi: 10.1038/nrn3475

Calcagno, V., Demoinet, E., Gollner, K., Guidi, L., Ruths, D., and De Mazancourt, C. (2012). Flows of research manuscripts among scientific journals reveal hidden submission patterns. Science (N.Y.) 338, 1065–1069. doi: 10.1126/science.1227833

Callaham, M. (2002). Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA 287, 2847–2850. doi: 10.1001/jama.287.21.2847

Carpenter, S. (2012). Psychology's bold initiative. Science 335, 1558–1561. doi: 10.1126/science.335.6076.1558

Casadevall, A., and Fang, F. C. (2012). Reforming science: methodological and cultural reforms. Inf. Immun. 80, 891–896. doi: 10.1128/IAI.06183-11

Chow, C. W., Haddad, K., Singh, G., and Wu, A. (2007). On using journal rank to proxy for an article's contribution or value. Issues Account. Edu. 22, 411–427. doi: 10.2308/iace.2007.22.3.411

Cokol, M., Iossifov, I., Rodriguez-Esteban, R., and Rzhetsky, A. (2007). How many scientific papers should be retracted? EMBO Reports 8, 422–423. doi: 10.1038/sj.embor.7400970

Collaboration, O. S. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. Perspect. Psychol. Sci. 7, 657–660. doi: 10.1177/1745691612462588

Curry, S. (2009). Eye-opening Access. Occam's Typwriter: Reciprocal Space. Available online at: http://occamstypewriter.org/scurry/2009/03/27/eye_opening_access/

Dwan, K., Altman, D. G., Arnaiz, J. A., Bloom, J., Chan, A.-W., Cronin, E., et al. (2008). Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PLoS ONE 3:e3081. doi: 10.1371/journal.pone.0003081

Editorial. (2005). Not-so-deep impact. Nature 435, 1003–1004. doi: 10.1038/4351003b

Editorial. (2006). The impact factor game. It is time to find a better way to assess the scientific literature. PLoS Med. 3:e291. doi: 10.1371/journal.pmed.0030291

Editorial. (2012). The well-behaved scientist. Science 335, 285–285.

Editorial. (2013). Beware the impact factor. Nat. Materials 12:89. doi: 10.1038/nmat3566

European Commission. (2010). Science and Technology Report. 340 / Wave 73.1 – TNS Opinion and Social.

Evangelou, E., Siontis, K. C., Pfeiffer, T., and Ioannidis, J. P. A. (2012). Perceived information gain from randomized trials correlates with publication in high-impact factor journals. J. Clin. Epidemiol. 65, 1274–1281. doi: 10.1016/j.jclinepi.2012.06.009

Eve, M. P. (2012). Tear it down, build it up: the research output team, or the library-as-publisher. Insights UKSG J. 25, 158–162. doi: 10.1629/2048-7754.25.2.158

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS ONE 4:e5738. doi: 10.1371/journal.pone.0005738

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. PLoS ONE 5:e10271. doi: 10.1371/journal.pone.0010271

Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. Scientometrics 90, 891–904. doi: 10.1007/s11192-011-0494-7

Fang, F. C., and Casadevall, A. (2011). Retracted science and the retraction index. Inf. Immun. 79, 3855–3859. doi: 10.1128/IAI.05661-11

Fang, F. C., Steen, R. G., and Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. Proc. Natl. Acad. Sci. U.S.A. 109, 17028–17033. doi: 10.1073/pnas.1212247109

Finardi, U. (2013). Correlation between journal impact factor and citation performance: an experimental study. J. Inf. 7, 357–370. doi: 10.1016/j.joi.2012.12.004

Florian, R. V. (2012). Aggregating post-publication peer reviews and ratings. Front. Comput. Neurosci. 6:31. doi: 10.3389/fncom.2012.00031

Frank, M. (2003). Impact factors: arbiter of excellence? JMLA 91, 4–6.

Garfield, E. (1999). Journal impact factor: a brief review. CMAJ 161, 979–980.

Garwood, J. (2011). A conversation with Peter Lawrence, Cambridge. “The Heart of Research is Sick”. LabTimes 2, 24–31.

Gauchat, G. (2010). The cultural authority of science: public trust and acceptance of organized science. Pub. Understand. Sci. 20, 751–770. doi: 10.1177/0963662510365246

Gauchat, G. (2012). Politicization of Science in the public sphere: a study of public trust in the United States, 1974 to 2010. Am. Sociol. Rev. 77, 167–187. doi: 10.1177/0003122412438225

Ghosh, S. S., Klein, A., Avants, B., and Millman, K. J. (2012). Learning from open source software projects to improve scientific review. Front. Comput. Neurosci. 6:18. doi: 10.3389/fncom.2012.00018

Giles, J. (2007). Breeding cheats. Nature 445, 242–243. doi: 10.1038/445242a

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards Grease the way through the publication bottleneck but undermine science. Perspect. Psychol. Sci. 7, 562–571. doi: 10.1177/1745691612457576

Goldacre, B. (2011). I Foresee that Nobody Will Do Anything About this Problem. Bad Science. Available online at: http://www.badscience.net/2011/04/i-foresee-that-nobody-will-do-anything-about-this-problem

Gonon, F., Konsman, J. -P., Cohen, D., and Boraud, T. (2012). Why most biomedical findings Echoed by newspapers turn out to be false: the case of attention deficit hyperactivity disorder. PLoS ONE 7:e44275. doi: 10.1371/journal.pone.0044275

Gordon, M. D. (1982). Citation ranking versus subjective evaluation in the determination of journal hierachies in the social sciences. J. Am. Soc. Inf. Sci. 33, 55–57. doi: 10.1002/asi.4630330109

Hamilton, J. (2011). Debunked Science: Studies Take Heat In 2011. NPR. Available online at: http://www.npr.org/2011/12/29/144431640/debunked-science-studies-take-heat-in-2011

Hartshorne, J. K., and Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. Front. Comput. Neurosci. 6:8. doi: 10.3389/fncom.2012.00008

Hauser, J. R., and Katz, G. M. (1998). Metrics: you are what you measure!. Eur. Manag. J. 16, 517–528. doi: 10.1016/S0263-2373(98)00029-2

Hegarty, P., and Walton, Z. (2012). The consequences of predicting scientific impact in psychology using journal impact factors. Perspect. Psychol. Sci. 7, 72–78. doi: 10.1177/1745691611429356

Hernán, M. A. (2009). Impact factor: a call to reason. Epidemiology 20, 317–318; discussion 319–20. doi: 10.1097/EDE.0b013e31819ed4a6

Hönekopp, J., and Khan, J. (2011). Future publication success in science is better predicted by traditional measures than by the h index. Scientometrics 90, 843–853. doi: 10.1007/s11192-011-0551-2

House of Commons. (2004). Scientific publications: free for all? Tenth Report of Session 2003-2004, vol II: Written evidence, Appendix 138.

Hunter, J. (2012). Post-publication peer review: opening up scientific conversation. Front. Comput. Neurosci. 6:63. doi: 10.3389/fncom.2012.00063

Ietto-Gillies, G. (2012). The evaluation of research papers in the XXI century. The open peer discussion system of the world economics association. Front. Comput. Neurosci. 6:54. doi: 10.3389/fncom.2012.00054

Ioannidis, J. P. A. (2005a). Contradicted and initially stronger effects in highly cited clinical research. JAMA 294, 218–228. doi: 10.1001/jama.294.2.218

Ioannidis, J. P. A. (2005b). Why most published research findings are false. PLoS Med. 2:e124 doi: 10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. Perspect. Psychol. Sci. 7, 645–654. doi: 10.1177/1745691612464056

Ioannidis, J. P. A., and Panagiotou, O. A. (2011). Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA 305, 2200–2210. doi: 10.1001/jama.2011.713

Kelly, C. D. (2006). Replicating empirical research in behavioral ecology: how and why it should be done but rarely ever is. Q. Rev. Biol. 81, 221–236. doi: 10.1086/506236

Kravitz, D. J., and Baker, C. I. (2011). Toward a new model of scientific publishing: discussion and a proposal. Front. Comput. Neurosci. 5:55. doi: 10.3389/fncom.2011.00055

Kreiman, G., and Maunsell, J. H. R. (2011). Nine criteria for a measure of scientific output. Front. Comput. Neurosci. 5:48. doi: 10.3389/fncom.2011.00048

Kriegeskorte, N. (2012). Open evaluation: a vision for entirely transparent post-publication peer review and rating for science. Front. Comput. Neurosci. 6:79. doi: 10.3389/fncom.2012.00079

Kriegeskorte, N., Walther, A., and Deca, D. (2012). An emerging consensus for open evaluation: 18 visions for the future of scientific publishing. Front. Comput. Neurosci. 6:94. doi: 10.3389/fncom.2012.00094

Kyrillidou, M., Morris, S., and Roebuck, G. (2012). ARL Statistics. American Research Libraries Digital Publications. Available online at: http://publications.arl.org/ARL_Statistics

Lau, S. L., and Samman, N. (2007). Levels of evidence and journal impact factor in oral and maxillofacial surgery. Int. J. Oral Maxillofac. Surg. 36, 1–5. doi: 10.1016/j.ijom.2006.10.008

Lawrence, P. (2008). Lost in publication: how measurement harms science. Ethics Sci. Environ. Politics 8, 9–11. doi: 10.3354/esep00079

Lawrence, P. A. (2002). Rank injustice. Nature 415, 835–836. doi: 10.1038/415835a

Lawrence, P. A. (2007). The mismeasurement of science. Curr. Biol. 17, R583–R585. doi: 10.1016/j.cub.2007.06.014

Lee, C. (2012). Open peer review by a selected-papers network. Front. Comput. Neurosci. 6:1. doi: 10.3389/fncom.2012.00001

Lehrer, J. (2010). The Decline Effect and the Scientific Method. New Yorker. Available online at: http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer

Lin, J. (2012). Cracking Open the Scientific Process: "Open Science" Challenges Journal Tradition With Web Collaboration. New York Times. Available online at: http://www.nytimes.com/2012/01/17/science/open-science-challenges-journal-tradition-with-web-collaboration.html?pagewanted=all

Liu, S. (2006). Top journals' top retraction rates. Sci. Ethics 1, 92–93.

Lopez-Cozar, E. D., and Cabezas-Clavijo, A. (2013). Ranking journals: could Google scholar metrics be an alternative to journal citation reports and Scimago journal rank? ArXiv 1303. 5870, 26. doi: 10.1087/20130206

Lozano, G. A., Larivière, V., and Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. J. Am. Soc. Inf. Sci. Technol. 63, 2140–2145. doi: 10.1002/asi.22731

Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., and Yang, Y. (2012). Rat brains also have a default mode network. Proc. Natl. Acad. Sci. U.S.A. 109, 3979–3984. doi: 10.1073/pnas.1200506109

Makel, M. C., Plucker, J. A., and Hegarty, B. (2012). Replications in psychology research: how often do they really occur? Perspect. Psychol. Sci. 7, 537–542. doi: 10.1177/1745691612460688

Miller, K. R. (2012). America's Darwin Problem. Huffington Post. Available online at: http://www.huffingtonpost.com/kenneth-r-miller/darwin-day-evolution_b_1269191.html

Moed, H. F., and Van Leeuwen, T. N. (1995). Improving the accuracy of institute for scientific information's journal impact factors. J. Am. Soc. Inf. Sci. 46, 461–467. doi: 10.1002/(SICI)1097-4571(199507)46:6<461::AID-ASI5>3.0.CO;2-G

Moed, H. F., and Van Leeuwen, T. N. (1996). Impact factors can mislead. Nature 381, 186. doi: 10.1038/381186a0

Moed, H. F., Van Leeuwen, T. N., and Reedijk, J. (1996). A critical analysis of the journal impact factors of Angewandte Chemie and the journal of the American Chemical Society inaccuracies in published impact factors based on overall citations only. Scientometrics 37, 105–116. doi: 10.1007/BF02093487

Møller, A. P., and Jennions, M. D. (2001). Testing and adjusting for publication bias. Trends Ecol. Evol. 16, 580–586. doi: 10.1016/S0169-5347(01)02235-2

Møller, A. P., Thornhill, R., and Gangestad, S. W. (2005). Direct and indirect tests for publication bias: asymmetry and sexual selection. Anim. Behav. 70, 497–506. doi: 10.1016/j.anbehav.2004.11.017

Munafò, M. R., Freimer, N. B., Ng, W., Ophoff, R., Veijola, J., Miettunen, J., et al. (2009). 5-HTTLPR genotype and anxiety-related personality traits: a meta-analysis and new data. Am. J. Med. Genet. B Neuropsychiat. Genet. 150B, 271–281. doi: 10.1002/ajmg.b.30808

Munafò, M. R., Matheson, I. J., and Flint, J. (2007). Association of the DRD2 gene Taq1A polymorphism and alcoholism: a meta-analysis of case-control studies and evidence of publication bias. Mol. Psychiatry 12, 454–461. doi: 10.1038/sj.mp.4001938

Nath, S. B., Marcus, S. C., and Druss, B. G. (2006). Retractions in the research literature: misconduct or mistakes? Med. J. Aust. 185, 152–154.

Nosek, B. A., and Bar-Anan, Y. (2012). Scientific Utopia: I. opening scientific communication. Psychol. Inquiry 23, 217–243. doi: 10.1080/1047840X.2012.692215

Nosek, B. A., Spies, J. R., and Motyl, M. (2012). Scientific Utopia: II. restructuring incentives and practices to promote truth over publishability. Perspect. Psychol. Sci. 7, 615–631. doi: 10.1177/1745691612459058

Nowotny, H. (2005). Science and society. High- and low-cost realities for science and society. Science (N.Y.) 308, 1117–1118. doi: 10.1126/science.1113825

Obremskey, W. T., Pappas, N., Attallah-Wasif, E., Tornetta, P., and Bhandari, M. (2005). Level of evidence in orthopaedic journals. J. Bone Joint Surg. Am. 87, 2632–2638. doi: 10.2106/JBJS.E.00370

Palmer, A. R. (2000). Quasi-replication and the contract of error: lessons from sex ratios, heritabilities and fluctuating asymmetry. Ann. Rev. Ecol. Syst. 31, 441–480. doi: 10.1146/annurev.ecolsys.31.1.441

Popper, K. (1995). In Search of A Better World: Lectures and Essays from Thirty Years. New Edn. Routledge (December 20, 1995) (London).

Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. Front. Comput. Neurosci. 6:33. doi: 10.3389/fncom.2012.00033

Priem, J. (2013). Scholarship: beyond the paper. Nature 495, 437–440. doi: 10.1038/495437a

Priem, J., and Hemminger, B. M. (2012). Decoupling the scholarly journal. Front. Comput. Neurosci. 6:19. doi: 10.3389/fncom.2012.00019

Priem, J., Piwowar, H. A., and Hemminger, B. M. (2012). Altmetrics in the wild: using social media to explore scholarly impact. ArXiv 1203.4745

Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? Nat. Rev. Drug Disc. 10, 712. doi: 10.1038/nrd3439-c1

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., and Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: a comparison between Innovation Studies and Business and Management. Res. Policy 41, 1262–1282. doi: 10.1016/j.respol.2012.03.015

Reedijk, J. (1998). Sense and nonsense of science citation analyses: comments on the monopoly position of ISI and citation inaccuracies. Risks of possible misuse and biased citation and impact data. New J. Chem. 22, 767–770. doi: 10.1039/a802808g

Research Information Network. (2008). Activities, Costs and Funding Flows in the Scholarly Communications System Research Information Network. Report commissioned by the Research Information Network (RIN). Available online at: http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu

Roemer, R. C., and Borchardt, R. (2012). From bibliometrics to altmetrics: a changing scholarly landscape. College Res. Libr. News 73, 596–600.

Rossner, M., Van Epps, H., and Hill, E. (2007). Show me the data. J. Cell Biol. 179, 1091–1092. doi: 10.1083/jcb.200711140

Saha, S., Saint, S., and Christakis, D. A. (2003). Impact factor: a valid measure of journal quality? JMLA 91, 42–46.

Sandewall, E. (2012). Maintaining live discussion in two-stage open peer review. Front. Comput. Neurosci. 6:9. doi: 10.3389/fncom.2012.00009

Sarewitz, D. (2012). Beware the creeping cracks of bias. Nature 485, 149–149. Available online at: http://www.nature.com/doifinder/10.1038/485149a [Accessed May 10, 2012]. doi: 10.1038/485149a

Sarewitz, D. (2013). Science must be seen to bridge the political divide. Nature 493:7. doi: 10.1038/493007a

Schooler, J. (2011). Unpublished results hide the decline effect. Nature 470, 437. doi: 10.1038/470437a

Scott, S., Kranz, J. E., Cole, J., Lincecum, J. M., Thompson, K., Kelly, N., et al. (2008). Design, power, and interpretation of studies in the standard murine model of ALS. Amyotroph. Lateral Scler. 9, 4–15. doi: 10.1080/17482960701856300

Seglen, P. O. (1992). The skewness of science. J. Am. Soc. Inf. Sci. 43, 628–638. doi: 10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0

Seglen, P. O. (1994). Causal relationship between article citedness and journal impact. J. Am. Soc. Inf. Sci. 45, 1–11. doi: 10.1002/(SICI)1097-4571(199401)45:1<1::AID-ASI1>3.0.CO;2-Y

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. BMJ 314, 498–502. doi: 10.1136/bmj.314.7079.497

Shapin, S. (2008). The Scientific Life?: A Moral History of a Late Modern Vocation. Chicago: University of Chicago Press.

doi: 10.7208/chicago/9780226750170.001.0001

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). Falsepositive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychol. Sci. 22, 1359–1366. doi: 10.1177/0956797611417632

Simmons, L. W., Tomkins, J. L., Kotiaho, J. S., and Hunt, J. (1999). Fluctuating paradigm. Proc. R. Soc. B Biol. Sci. 266, 593–595. doi: 10.1098/rspb.1999.0677

Singh, G., Haddad, K. M., and Chow, C. W. (2007). Are articles in "Top" management journals necessarily of higher quality? J. Manag. Inquiry 16, 319–331. doi: 10.1177/1056492607305894

Siontis, K. C. M., Evangelou, E., and Ioannidis, J. P. A. (2011). Magnitude of effects in clinical trials published in high-impact general medical journals. Int. J. Epidemiol. 40, 1280–1291. doi: 10.1093/ije/dyr095

Sønderstrup-Andersen, E. M., and Sønderstrup-Andersen, H. H. K. (2008). An investigation into diabetes researcher's perceptions of the Journal Impact Factor—reconsidering evaluating research. Scientometrics 76, 391–406. doi: 10.1007/s11192-007-1924-4

Song, F., Eastwood, A., Gilbody, S., and Duley, L. (1999). The role of electronic journals in reducing publication bias. Inf. Health Soc. Care 24, 223–229. doi: 10.1080/146392399298429

Starbuck, W. H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. Org. Sci. 16, 180–200.
doi: 10.1287/orsc.1040.0107

Statzner, B., and Resh, V. H. (2010). Negative changes in the scientific publication process in ecology: potential causes and consequences. Freshw. Biol. 55, 2639–2653. doi: 10.1111/j.1365-2427.2010.02484.x

Steen, R. G. (2011a). Retractions in the scientific literature: do authors deliberately commit research fraud? J. Med. Ethics 37, 113–117. doi: 10.1136/jme.2010.038125

Steen, R. G. (2011b). Retractions in the scientific literature: is the incidence of research fraud increasing? J. Med. Ethics 37, 249–253. doi: 10.1136/jme.2010.040923

Sutton, J. (2011). psi study highlights replication problems. The Psychologist News. Available online at:
http://www.thepsychologist.org.uk/blog/blogpost.cfm?threadid=1984andcatid=48

Taylor, M., Perakakis, P., and Trachana, V. (2008). The siege of science. Ethics Sci. Environ. Polit. 8, 17–40. doi: 10.3354/esep00086

Todd, P., and Ladle, R. (2008). Hidden dangers of a 'citation culture'. Ethics Sci. Environ. Polit. 8, 13–16. doi: 10.3354/esep00091

Tressoldi, P. E., Giofré, D., Sella, F., and Cumming, G. (2013). High impact = high statistical standards? Not necessarily so. PLoS ONE 8:e56180. doi: 10.1371/journal.pone.0056180

Tsikliras, A. (2008). Chasing after the high impact. Ethics Sci. Environ. Polit. 8, 45–47. doi: 10.3354/esep00087

Upadhyay, J., Baker, S. J., Chandran, P., Miller, L., Lee, Y., Marek, G. J., et al. (2011). Default-mode-like network activation in awake rodents. PLoS ONE 6:e27839. doi: 10.1371/journal.pone.0027839

Van Dongen, S. (2011). Associations between asymmetry and human attractiveness: possible direct effects of asymmetry and signatures of publication bias. Ann. Hum. Biol. 38, 317–323. doi: 10.3109/03014460.2010.544676

Van Noorden, R. (2011). Science publishing: the trouble with retractions. Nature 478, 26–28. doi: 10.1038/478026a

Van Noorden, R. (2013). Mathematicians aim to take publishers out of publishing. Nat. News. Available online at: http://www.nature.com/news/mathematicians-aim-to-take-publishers-out-of-publishing-1.12243

Vanclay, J. K. (2011). Impact factor: outdated artefact or stepping-stone to journal certification? Scientometrics 92, 211–238. doi: 10.1007/s11192-011-0561-0

Wager, E., and Williams, P. (2011). Why and how do journals retract articles? An analysis of Medline retractions 1988–2008. J. Med. Ethics 37, 567–570. doi: 10.1136/jme.2010.040964

Wagner, P. D. (2011). What's in a number? J. Appl. Physiol. (Bethesda, MD.: 1985) 111: 951–953. doi: 10.1152/japplphysiol.00935.2011

Walther, A., and Van den Bosch, J. J. F. (2012). FOSE: a framework for open science evaluation. Front. Comput. Neurosci. 6:32. doi: 10.3389/fncom.2012.00032

Weale, A. R., Bailey, M., and Lear, P. A. (2004). The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor. BMC Med. Res. Methodol. 4:14. doi: 10.1186/1471-2288-4-14

Welberg, L. (2012). Neuroimaging: rats join the "default mode" club. Nat. Rev. Neurosci. 11:223. doi: 10.1038/nrn3224

Wicherts, J. M., Kievit, R. A., Bakker, M., and Borsboom, D. (2012). Letting the daylight in: reviewing the reviewers and other ways to maximize transparency in science. Front. Comput. Neurosci. 6:20. doi: 10.3389/fncom.2012.00020

Yarkoni, T. (2012). Designing next-generation platforms for evaluating scientific output: what scientists can learn from the social web. Front. Comput. Neurosci. 6:72. doi: 10.3389/fncom.2012.00072

Yong, E. (2012). Replication studies: bad copy. Nature 485, 298–300. doi: 10.1038/485298a

Young, N. S., Ioannidis, J. P. A., and Al-Ubaydli, O. (2008). Why current publication practices may distort science. PLoS Med. 5:e201.
doi: 10.1371/journal.pmed.0050201

Yue, W., Wilson, C. S., and Boller, F. (2007). Peer assessment of journal quality in clinical neurology. JMLA 95, 70–76.

Zimmermann, J., Roebroeck, A., Uludag, K., Sack, A., Formisano, E., Jansma, B., et al. (2011). Network-based statistics for a community driven transparent publication process. Front. Comput. Neurosci. 6:11. doi: 10.3389/fncom.2012.00011

39.

IMPACT FACTOR DISTORTIONS

BRUCE ALBERTS

SCIENCE 17 MAY 2013: VOL. 340 NO. 6134 P. 787

– ◊ –

This Editorial coincides with the release of the San Francisco declaration on research Assessment (DORA), the outcome of a gathering of concerned scientists at the December 2012 meeting of the American Society for Cell Biology.* To correct distortions in the evaluation of scientific research, DORA aims to stop the use of the "journal impact factor" in judging an individual scientist's work. The Declaration states that the impact factor must not be used as "a surrogate measure of the quality of individual research articles, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions." DORA also provides a list of specific actions, targeted at improving the way scientific publications are assessed, to be taken by funding agencies, institutions, publishers, researchers, and the organizations that supply metrics. These recommendations have thus far been endorsed by more than 150 leading scientists and 75 scientific organizations, including the American Association for the Advancement of Science (the publisher of Science).

Here are some reasons why:

The impact factor, a number calculated annually for each scientific journal based on the average number of times its articles have been referenced in other articles, was never intended to be used to evaluate individual scientists, but rather as a measure of journal quality. However, it has been increasingly misused in this way, with scientists now being ranked by weighting each of their publications according to the impact factor of the journal in which it appeared. For this reason, I have seen curricula vitae in which a scientist annotates each of his or her publications with its journal impact factor listed to three significant decimal places (for example, 11.345). And in some nations, publication in a journal with an impact factor below 5.0 is officially of zero value. As frequently pointed out by leading scientists, this impact factor mania makes no sense.†

The misuse of the journal impact factor is highly destructive, inviting a gaming of the metric that can bias journals against publishing important papers in fields (such as social sciences and ecology) that are much less cited than others (such as biomedicine). And it wastes the time of scientists by overloading highly cited journals such as Science with inappropriate submissions from researchers who are desperate to gain points from their evaluators.‡

But perhaps the most destructive result of any automated scoring of a researcher's quality is the "me-too science" that it encourages. Any evaluation system in which the mere number of a researcher's publications increases his or her score creates a strong disincentive to pursue risky and potentially groundbreaking work, because it takes years to create a new approach in a new experimental context, during which no publications should be expected. Such metrics further block innovation because they encourage scientists to work in areas of science that are already highly populated, as it is only in these fields that large numbers of scientists can be expected to reference one's work, no matter how outstanding. Thus, for example, in my own field of cell biology, new tools now allow powerful approaches to understanding how a large single-celled organism such as the cilate Stentor can precisely pattern its surface, creating organlike features that are presently associated only with multicellular organisms.§ The answers are likely to bring new insights into how all cells operate, including our own. But only the very bravest of young scientists can be expected to venture into such a poorly populated research area, unless automated numerical evaluations of individuals are eliminated.

The DORA recommendations are critical for keeping science healthy. As a bottom line, the leaders of the scientific enterprise must accept full responsibility for thoughtfully analyzing the scientific contributions of other researchers. To do so in a meaningful way requires the actual reading of a small selected set of each researcher's publications, a task that must not be passed by default to journal editors.

* www.ascb.org/SFdeclaration.html.

† K. Simons, Science 322, 165 (2008).

‡ B. Alberts, B. Hanson, K. L. Kelner, Science 321, 15 (2008).

§ For example, see M. Kirschner, J. Gerhart, T. Mitchison, Cell 100, 79 (2000).

40.

The "Impact Factor" and Selected Issues of Content and Technology in Humanities Scholarship Published Online

Steven Tötösy de Zepetnek

Journal of Scholarly Publishing 41.4 (2010)

– ◊ –

In this article I discuss issues of content and valuation in publishing humanities scholarship online. What I mean by content is not scholarship per se, but the "value" of an article published by a scholar as to its locus of publication. While in the U.S. and Canada peer review is accepted as a measure of quality for promotion and tenure (although humanities departments and faculties of arts at times keep their own lists of "important" journals ranked as to the importance of the publication of articles), in Europe and in Asia (for example in Spain, Belgium, increasingly in Germany, India, Taiwan, the People's Republic of China, etc.) there is a move in the humanities towards the principle of the "impact factor" as borrowed from the social sciences, the natural sciences, medicine, and other empirical and applied fields. I suspect the implementation of the impact factor in the humanities may play a role sooner rather than later also in the U.S., Canada, and Australia. It is of particular relevance that the said impact factor with regard to the humanities is tied to one particular indexing and abstracting service, the for–profit company Thomson Reuters Institute for Scientific Information Arts and Humanities Citation Index.

Importantly, the valuation of scholarship via the impact factor creates a challenge for the humanities in general and for online publishing in the field in particular.

Before I discuss the situation I outline above, I would like to note that there is, a priori, a difference between those (universities and scholars themselves) who prefer the traditional way of publishing scholarship in print and those who believe in adopting the mode of online publishing using the advantages of new media technology: it is curious situation, indeed, that while most online journals' publication of scholarship is no different from print publication except the mode of delivery, scholars and the academe — often and still — regard online publishing of less relevance, value, and importance than print publication. This has several implications, including that in some humanities departments the publication of articles in online journals — even if and despite of peer–review — are not accorded the same value as print publications. Situations thus arise in which scholars are in fact discouraged from publishing online. There is, then, a further gap between those who prefer online publishing based on subscription fees and those who argue for online publishing in open access. At the same time, owing to pressures in funding and the overall constricted situation of academic publishing in particular in the humanities, as well as the dramatic increase in published scholarship, most humanities scholars recognize the inevitability and necessity of digital publishing. This recognition is grudging, however, and with objections at many levels.1 One such issue is the said growing importance of the impact factor and its relevance for publishing humanities scholarship online.

The concept of the impact factor first arose in the 1960s with the science Citation Index (SCI) and, subsequently, with Journal Citation Reports (JCR). In its simplest terms, the impact factor is a measure of the frequency with which an article in a journal has been cited in a particular year or period. The impact factor is meant to indicate total citation frequencies and to eliminate some of the bias of counts that favor large journals over small ones, frequently issued journals over less–frequently issued ones, and older journals over newer ones. In the humanities, there are three indexing services considered increasingly as the primary carriers of the impact factor: 1) the Arts and Humanities Citation Index (AHCI, a subset of the Thomson Reuters Institute for Scientific Information [ISI] Web of Knowledge); 2) the European Reference Index for the Humanities (ERIH, recent and still in the process of development); and 3) Elsevier's Scopus, which does mostly science–based indexing, and is similar to the above ISI, a for–profit indexing and abstracting service.

With regard to the above three indexing and abstracting services for online humanities publications, the situation is not optional. For example, AHCI started to consider the indexing of articles published in online humanities journals only in 2005. An additional problem with the AHCI is that it neglects humanities journals in general, and humanities journals published online in particular. Christine L. Borgman, in her Scholarship in the Digital Age (2007), underlines the less–than–adequate coverage of humanities scholarship: "The depth of coverage in the ISI Web of Knowledge, which is among the most comprehensive online bibliography databases, is deepest in the sciences, shallower in the social sciences, and most shallow in the humanities … Indicators in the ISI citators are the least valid for the arts and humanities because they only include references made by journal articles."2 "Complicating matters further," Borgman continues, "the Science, Social Science, and Arts and Humanities Citations Indexes are a closed system consisting of references made by established journals on a list selected by the editors at Thomson Scientific. Not all journals are included, and books and conference proceedings rarely are indexed."3 Borgman reaches the conclusion that "The simplest approach to clarifying the legitimacy of digital documents is to rely on traditional quality indicators such as the imprimatur of well–regarded publishers. This approach, however, cedes much of the control for legitimization to publishers and discourages experimental forms of publication. For example, if the only publications valued for promotion, tenure, and institutional reviews are those that appear in journals with high–impact factor as measured by ISI citation statistics … journals indexed by the ISI have an inordinate power over the quality–control system."4

The above situation is an a priori handicap in the valuation of humanities scholarship, but even if we accept the inevitability of the widely sanctioned primacy of the ISI and its citation indices, the neglect of humanities journals in general and those published online in particular, the additional obstacles of being accepted by the indexing service is a further matter of relevance. The example I employ to illustrate selected issues is the journal I have been editing since 1999 and publish with Purdue University Press, CLCWeb: Comparative Literature and Culture (ISSN 1481–4374). The journal was submitted for indexing by ISI's AHCI in 2001 at which time AHCI's response was that indexing would not be possible because the journal's material is not cited in print journals (this was an argument similar to the one I received in 1999 when I applied for funding to the Social Sciences and Humanities Research Council of Canada: the response was that unless texts published in an online journal are cited in print journals and unless the online journal can show subscription fees, no funding is possible). By 2005 ISI changed its approach to the indexing of online journals in the humanities and thus I re–submitted the journal. While the content of the journal was approved speedily, AHCI told me that there are technical issues to be resolved and that it would take some time to arrive at the approval of the journal for indexing, this because CLCWeb is published in html. At the same time Purdue University Press embarked on a number of projects with regard to its digital repositories and online publications and by 2006 the Press decided to publish its online journals in pdf in DigitalCommons with a third–party publisher, the Berkeley Electronic Press. Thus, between January 2006 and January 2007 I converted with the assistance of three editorial assistants all material of CLCWeb 1999–current for publication in pdf and re–submitted the journal to AHCI for indexing. In March 2009 I was told again that the journal has been approved with regard to its scholarly content and with the technical evaluation remaining CLCWeb would likely be indexed starting in 2009. On 11 November 2009 I received the approval to index from ISI with the statement that CLCWeb has been approved for indexing in AHCI as of volume 10 issue 1 (2008) (on this, namely backdating the indexing, see below). Overall, my experience with AHCI has been not the best because of the time span to arrive at the approval of indexing journal stretching from 2001 until 2009: while there may have been legitimate issues because of matters technical, the long stretch to receive the approval to index is less than acceptable (after all, ISI is a for–profit company whose institutional subscription fee for university libraries is USD 11,000/year) if for no other reason than the urgency of the prominence of ISI and the development to accord AHCI indexed journals the prominence and importance I mention above.

The European Reference Index for the Humanities (ERIH) is an undertaking by the European Union's European Science Foundation and "aims initially to identify, and gain more visibility for top–quality European Humanities research published in academic journals in, potentially, all European languages. It is a fully peer–reviewed, Europe–wide process, in which 15 expert panels sift and aggregate input received from funding agencies, subject associations and specialist research centres across the continent. In addition to being a reference index of the top journals in 15 areas of the Humanities, across the continent and beyond, it is intended that ERIH will be extended to include book–form publications and non–traditional formats. It is also intended that ERIH will form the backbone of a fully–fledged research information system for the Humanities."5 At ERIH's inception in 2006, the humanities experts chosen to decide which journal would be included in their ranking system numbered all of four scholars: this has changed since and now there are dozens of scholars appointed to select and review journals in many fields in the humanities. CLCWeb: Comparative Literature and Culture was submitted for indexing by ERIH in 2006, at which time I received no acknowledgement or any communication; the journal was resubmitted in September 2007 and ERIH's "Junior Science Officer"—who stated just having been hired—apologized in a June 2008 e–mail that my previous submission was not acknowledged and wrote also that by the end of 2008 the selection of submitted journals would be completed and that such selections will be made every four years after that. The last communication I received from ERIH was an e–mail of 17 October 2008. I have not heard anything since about whether the journal was approved for indexing or whether the evaluation process would be still ongoing.

In 2008, across Europe as well as in Australia and on the listserve of the Council of Editors of Learned Journals (CELJ), a discussion about ERIH's process of selection erupted. Editors of humanities journals and others criticized the "numerics" and "metrics" approach applied to humanities scholarship and ERIH's A, B, or C system of journal rankings (see also the report on ERIH by Jennifer Howard in The Chronicle of Higher Education in 2008). While ERIH argued that their system of ranking is not based on descending valuation but on different categories of journals, it remained clear to most that in the end a journal ranked A would end up as a top journal regardless of the category designation.6 ERIH, similar to AHCI, does not appear to have an interest in online journals in the humanities and the current list of the so–called initial 2008 rankings of ERIH has few of such.

Elsevier's Scopus is an abstracting service similar to ISI's AHCI, that is, it is a for–profit company. In my perception, the humanities rate low on their list of priorities, and until 2008 they did not even consider to include the humanities in their indexing services;7 however, as of 2009 they are starting the indexing of humanities journals. In comparison with AHCI and ERIH, my experience with Scopus has been better in the sense that after submission of CLCWeb for indexing in December 2008 I received the decision to include the journal in March 2009 and since October 2009 the journal is listed and indexed among Scopus' Arts and Humanities journals.

With the three indexing services discussed above, one further matter remains curious: once an online journal is accepted for listing and indexing, why would these services begin indexing only in the year of acceptance? Clearly, with online journals, whose history is no older than the birth of the world wide web in 1994, the indexing of all material published would be easy and of little expense—technically speaking—precisely because all its material is available online. I put this question to all three indexing services. AHCI responded that they would remain as is with the start of indexing with the year of acceptance (although, see above, when I received the decision to index, they approved the journal for indexing with 2008, thus backdated…), ERIH did not reply, and Scopus responded that indeed my proposal does make sense and that they are looking into the indexing of all material published in an online journal and that this may happen in the future.

As mentioned above, the move in the humanities towards valuation of work via the impact factor is occurring mostly in Europe and Asia, but this move has implications also for journals published in the U.S., Canada, and Australia.8 In many instances a journal published in the U.S., for example, is missing out on good work because scholars are reluctant to publish in a journal that is not indexed in the AHCI, and in the case of CLCWeb>: Comparative Literature and Culture this has occurred several times over the years. It remains without saying that the overall reluctance of many U.S. humanities scholars to publish in an online journal — whether AHCI–indexed or not — remains an indicator of the slow pace of the humanities alongside new media technology, an issue that has not changed significantly since 2001.9

Next, I discuss briefly a number of items relevant for online publishing. One of the most important of these, which in my opinion is often misapplied, is the matter of a journal's Universal Resource Locator (URL). There is the argument against long URLs; however, in my opinion the URL of a journal, whenever possible, ought to be a "narrative" URL and not an acronym. I think the reason that so many journals have acronyms and not narrative URLs, which would make the journal immediately recognizable, has much to do with the disjointed relationship between the scholar and the "techie." To illustrate this, here is again the example of CLCWeb: Comparative Literature and Culture: when the university press that publishes the journal decided to publish in .pdf instead of the previous HTML format, the URL was designated without consultation with me, the journal's editor, and a press technology expert made up the journal's new URL. In other words, it is important that the technology side and the scholarship side of a learned journal communicate with each other, and it should also be a standard that the editor of an online journal be someone who knows new media technology.

Another important issue is mirrors and archival preservation. Here, the U.S. is relying on the multiplication of institutional and commercial depositories of archives instead of on a national policy. The Library of Congress has no depository of online journals. In contrast, most countries have been creating national depositories of electronic material since the late 1990s. In addition to the archiving of a journal's material in a national digital depository, the use of mirrors is advisable not only for this type of preservation but also for the facilitation of the speed of download in other parts of the world. A mirror of a journal published in the U.S., for example, functions thus in Europe or Asia not only as facilitating a higher speed of download but also as a depository.

In conclusion, I would like to suggest that institutions such as university presses and associations and organizations in the humanities would make efforts to persuade and pressure indexing services such as the AHCI, Elsevier's Scopus, and the European ERIH as to their approach to humanities scholarship published online (or in print): a perpetually and traditionally underfunded field of scholarship, the humanities must receive due recognition and prompt inclusion for indexing not because I lament the traditional non–recognition of the humanities as not being "practical" and applied enough; rather, I am suggesting this because, after all, universities and university libraries subscribe world wide to the said indexing services to the tunes of thousands of dollars per year and since such subscriptions are package deals thus including the humanities, it is unacceptable that the humanities remain disregarded and on the second burner whether print or online journal indexing.

Note: Research for the above article and participation in the 2008 convention of the Modern Language Association of America panel on learned journals was funded by Purdue University Press and National Sun Yat–sen University's Center for the Humanities and Social Sciences. I thank Nathan L. Grant (Saint Louis University) for his coments on the paper.

1 For more detail, see Steven Tötösy de Zepetnek, "The New Knowledge Management and Online Research and Publishing in the Humanities," CLCWeb: Comparative Literature and Culture 3.1 (2001): Perdue University Press. See also Tötösy de Zepetnek, "New Media, Publishing in the Humanities, and CLCWeb: Comparative Literature and Culture,"Formamente: Rivista Internazionale di Ricerca sul Futuro Digitale / International Research Journal on Digital Future 2.1–2 (2007): 255–73.

2 Christine L. Borgman, Scholarship in the Digital Age: Information, Infrastructure, and the Internet (Cambridge, MA: MIT Press, 2007), page 158–59.

3 Borgman, Scholarship in the Digital Age, page 64.

4 Borgman, Scholarship in the Digital Age, page 85.

5 ERIH: European Reference Index for the Humanities.(2009):

6 Jennifer Howard, The Chronicle of Higher Education (10 October 2008): "New Ratings of Humanities Journals Do More Than Rank — They Rankle,".

7 Borgman, Scholarship in the Digital Age, page 215.

8 See also Howard, "New Ratings of Humanities Journals Do More Than Rank — They Rankle," (2008).

9 See Tötösy de Zepetnek, "The New Knowledge Management and Online Research and Publishing in the Humanities," (2001) and Tötösy de Zepetnek, "New Media, Publishing in the Humanities, and CLCWeb: Comparative Literature and Culture" (2007).