Research evaluation by scientometrics

Tuan V. Nguyen, D.Sc, Ph.D

Principal Fellow and Lab Head, Garvan Institute of Medical Research

Professor, St Vincent's Clinical School, UNSW Medicine, UNSW Australia

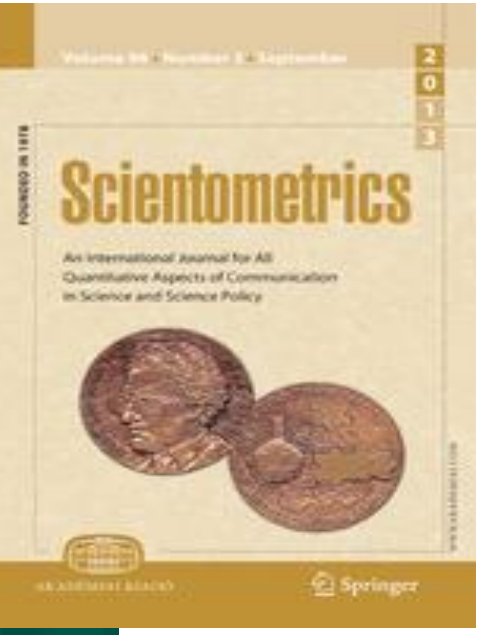
Professor of Predictive Medicine, University of Technology Sydney (UTS)

Adjunct Professor of Epidemiology and Biostatistics,

School of Medicine, University of Notre Dame Australia

Increased use of bibliometrics

- **Australia** – national research assessment (NHMRC) used bibliometric indices
- **UK** – institutional assessment is supported (not dictated) by citation analysis
- **Germany** – use of bibliometric indices is the norm
- **China** – authors are asked to publish only in ISI indexed journals
- **Norway** – counted pubications by a weighing factor
- **France**, **Canada**, **USA**, etc "crazied" about bibliometric indices



Informing Research Choices: Indicators and Judgment

The Expert Panel on Science Performance and Research Funding

Good use of bibliometrics for individual evaluation of researchers

2011/03/21 STI News

Report by the French Academy of Sciences, proposing 5 recommendations to control use of publications and limit improper use. The report was submitted to the French Ministry of Higher Education and Research on 17 January 2011.

(see pages (5-7 for recommendations).

Criteria for evaluation of an individual

- Research productivity
- Research quality
- Scientific and societal impact

Research Productivity

Research productivity

- Number of original (peer-reviewed) publications
- Reviews, book chapters
- Invited editorials, commentaries
- PhD graduates

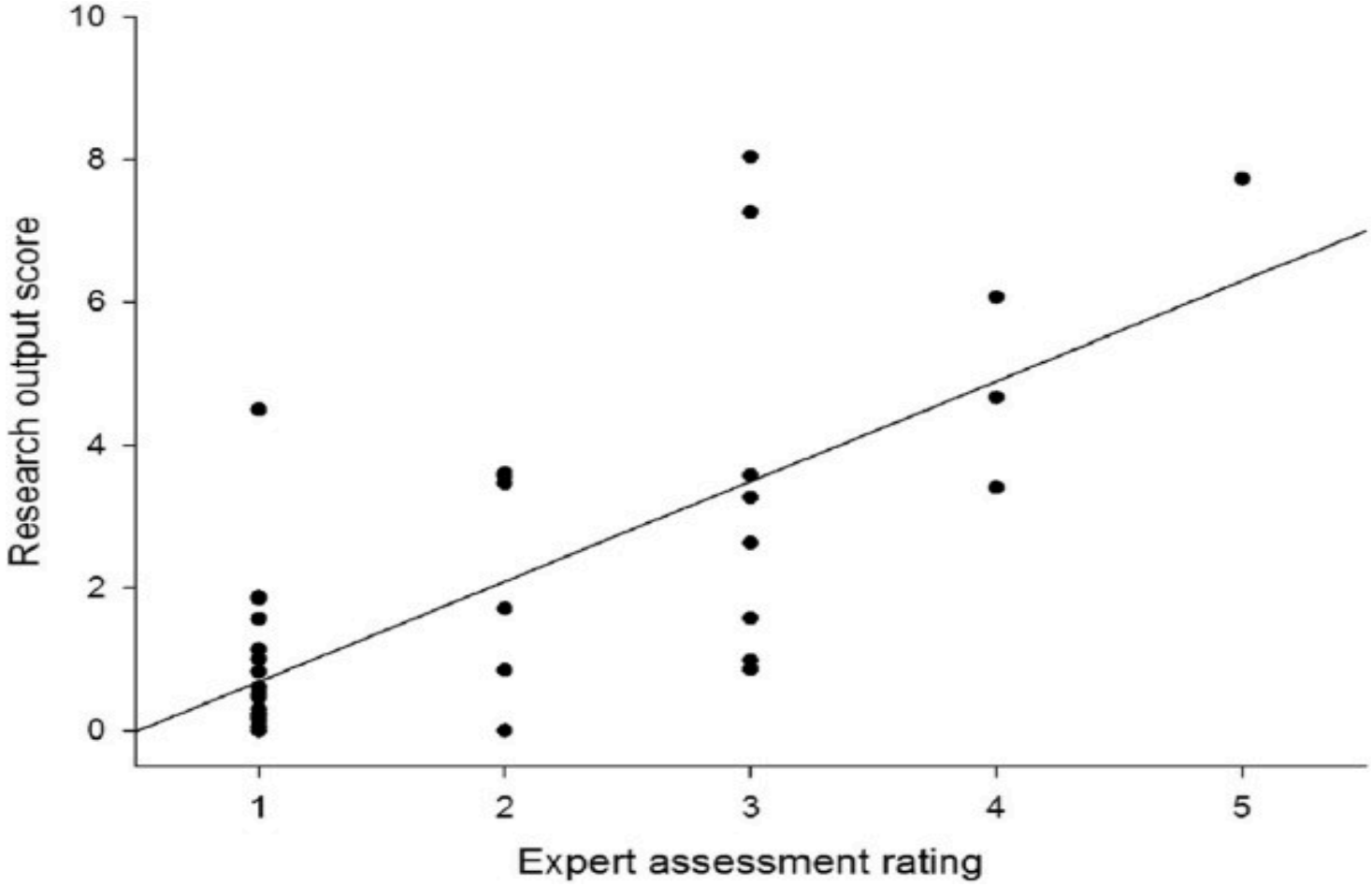
Research output score (ROS)

ROS = *p* + *s* + *g*

p: number of papers

s: number of PhD graduates

g: grant income



Wootton R. A simple generalizable method for measuring individual research productivity ... *Health Res Policy Syst 2013;11:2*

Productivity

- Output is NOT productivity
- Productivity should have a measure of *input*
- But what is input?
  - Ideas
  - Institutional support
  - Infrastructure
  - Competitive grants

Competitive grants

- Distributed through a bureaucratic process, low rates of success
- Potential bias against innovative grant (hard to find referees, regression-toward-the-mean effect)
- Grant success is not a good measure of academic productivity

A compromise measure

- **Productivity ratio** = ***Output / Input***
- D-index = ROS / Grant funding

Research Quality

“Non enim numero haec iudicantur, sed pondere” (the number does not matter, ***the quality does****)*

Marcus Tullius Cicero

But what is "quality"?

- Scientific importance of the work
- Rigor of methodology employed
- Elegance in research design and findings

Quality ≠ Impact

Measures of research quality

- Peer assessment
- Impact factor (IF)
- Number of citations
- h index and its derived measures (hc, m quotient)
- g index
- Egenfactor (EF)
- Relative citation index (normalized to field specific cites)
- Author superiority index (ASI)

Peer assessment

- A primary means has been used for many years
- **Many** serious problems
  - Subjective
  - Conflict of interest
  - Discipline (and local) favoritism
  - Insufficient competence
  - Superficial assessment

Problems with peer assessment

Peering Into Peer Review

Why don't proposals given better scores by the National Institutes of Health lead to more important research outcomes?

Peering at peer review revealed high degree of chance associated with funding of grant applications

Nancy E. Mayo[a,*], James Brophy[b], Mark S. Goldberg[a], Marina B. Klein[c], Sydney Miller[d], Robert W. Platt[e], Judith Ritchie[f]

[a]*Division of Clinical Epidemiology R4.29, McGill University Health Center, RVH Site, 687 Pine Avenue West, Montreal, H3A 1A1, Canada*
[b]*Health Technology Assessment Unit, R4.14, McGill University Health Center, RVH Site, 687 Pine Avenue West, Montreal, H3A 1A1, Canada*
[c]*Divisions of Infectious Diseases/Immunodeficiency, Royal Victoria Hospital, McGill University Health Centre, 687 Pine Avenue West, Montreal, H3A 1A1, Canada*

Peer Review Practices in Biomedical Literature: A Time for Change?

Kamal Kumar Mahawar,[1] Deepak Kejariwal,[2] Ajay Malviya,[3] Rashmi Birla[4] and Y.K.S. Viswanath,[5]
[1]Department of Surgery, Sunderland Royal Hospital, Sunderland, [2]Department of Gastroenterology, and [4]Department of Surgery, County Durham and Darlington NHS Trust, Durham, [3]Department of Orthopaedics, Wansbeck General Hospital, Ashington, and [5]Department of Surgery, James Cook University Hospital, Middlesborough, UK.

Decision of funding can be ... random

Abstract

Objective To quantify randomness and cost when choosing health and medical research projects for funding.

Design Retrospective analysis.

Setting Grant review panels of the National Health and Medical Research Council of Australia.

Participants Panel members' scores for grant proposals submitted in 2009.

Main outcome measures The proportion of grant proposals that were always, sometimes, and never funded after accounting for random variability arising from differences in panel members' scores, and the cost effectiveness of different size assessment panels.

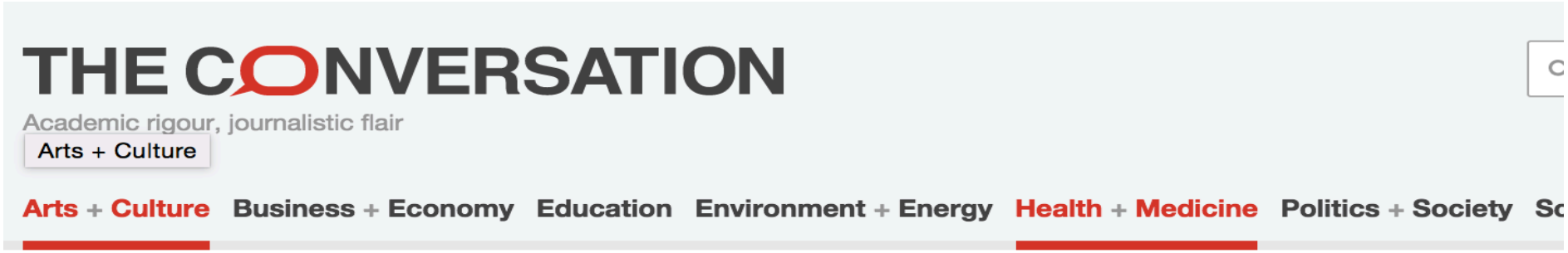
Results 59% of 620 funded grants were sometimes not funded when random variability was taken into account. Only 9% (n=255) of grant proposals were always funded, 61% (n=1662) never funded, and 29% (n=788) sometimes funded. The extra cost per grant effectively funded from the most effective system was $A18 541 (£11 848; €13 482; $19 343).

Conclusions Allocating funding for scientific research in health and medicine is costly and somewhat random. There are many useful research questions to be addressed that could improve current processes.

Graves et al. BMJ 2011;343

Fingers crossed: the role of randomness in medical research funding

September 28, 2011 6.26am AEST



A better way to award NHMRC's medical research grants

October 26, 2012 4.01pm AEDT

2. Research Quality / Peer Assessment

What about esteem indicators

Testing novel quantitative indicators of research 'quality', esteem and 'user engagement': an economics pilot study

Claire Donovan and Linda Butler

Peers decided that *novel esteem indicators* reflect individual standing, and research-oriented workload, rather than actual research quality. While these indicators correlate with research activity, and were less

One accurate measurement is worth a thousand expert opinions.

(Grace Hopper)

Impact factor

- Simple index (2-year window)
- As a **surrogate index** for research quality
- Monetary rewards
  - Garvan Institute
  - Overseas institutions

IF and monetary rewards

Table 1. Examples of monetary reward system

University	Monetary award
Guangzhou Medical University	
IF < 1	Three thousand RMB
1 ≤ IF < 2	Fifteen thousand RMB
2 ≤ IF < 3	Twenty-five thousand RMB
3 ≤ IF < 4	Thirty-five thousand RMB
4 ≤ IF < 5	Forty-five thousand RMB
5 ≤ IF < 8	Seventy thousand RMB
8 ≤ IF < 10	Ninety thousand RMB
10 ≤ IF < 15	One-hundred and thirty thousand RMB
IF ≥ 15	Three hundred thousand RMB
Zhejiang Chinese Medical University	
Nature or Science	Hundred thousand RMB
SCI papers with IF > 3	Six thousand RMB

US$7200

US$14400

US$48000

Shao JF, Shen HY. Research Evaluation (2012)

The failure of IF

1. Ecological fallacy: IF reflects the citation of a journal, not a individual paper
2. Matthew effect: Attention to high IF papers
3. Lack of transparncy
4. Irreproducible
5. Mix of publication types
6. Coercive journal self citation
7. No clear cut correlation between citation and quality
8. Can not do cross-discipline comparison
9. Long delay

Skewed distribution of citations

A typical journal citation distribution: Citations in 2011-2013 to *Nature* articles published in 2010 (made 20131103 from Scopus data)

Rule: ~60% of papers published in *any* journal are cited less than the average IF

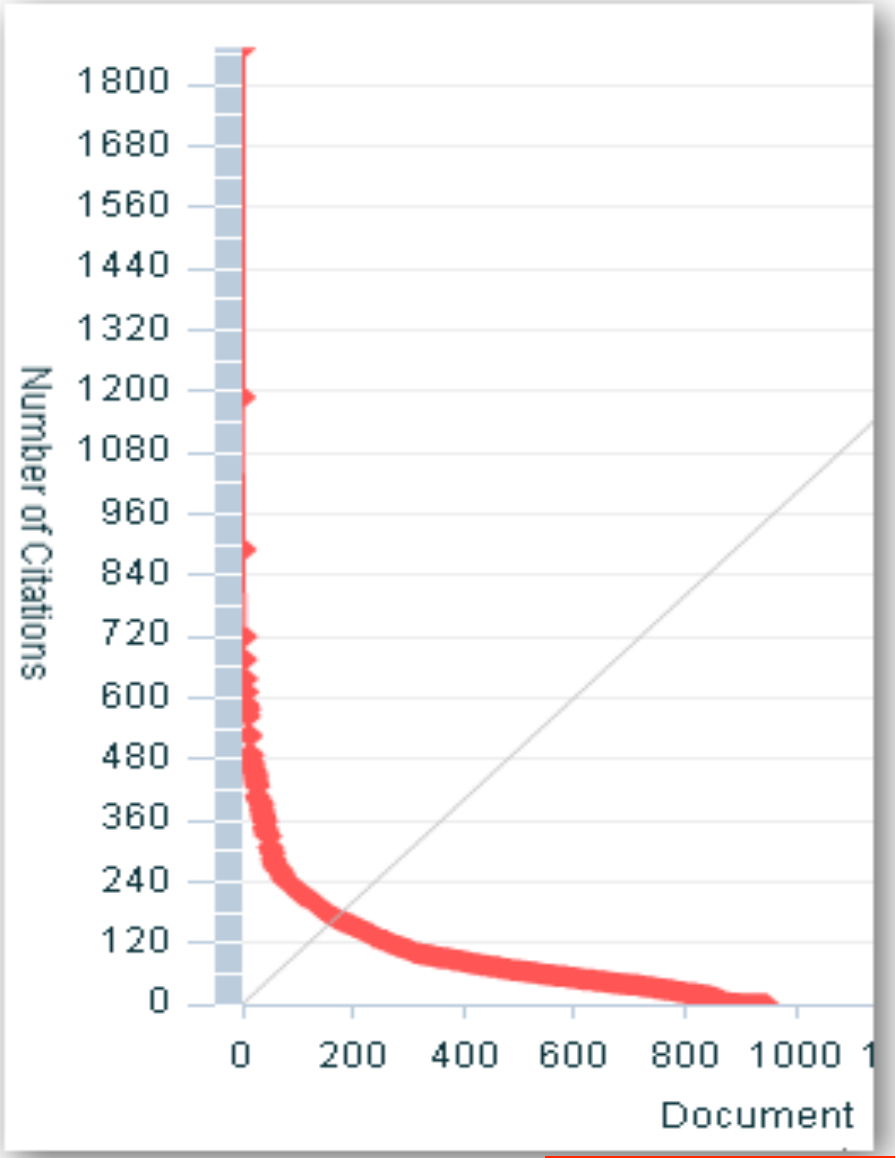
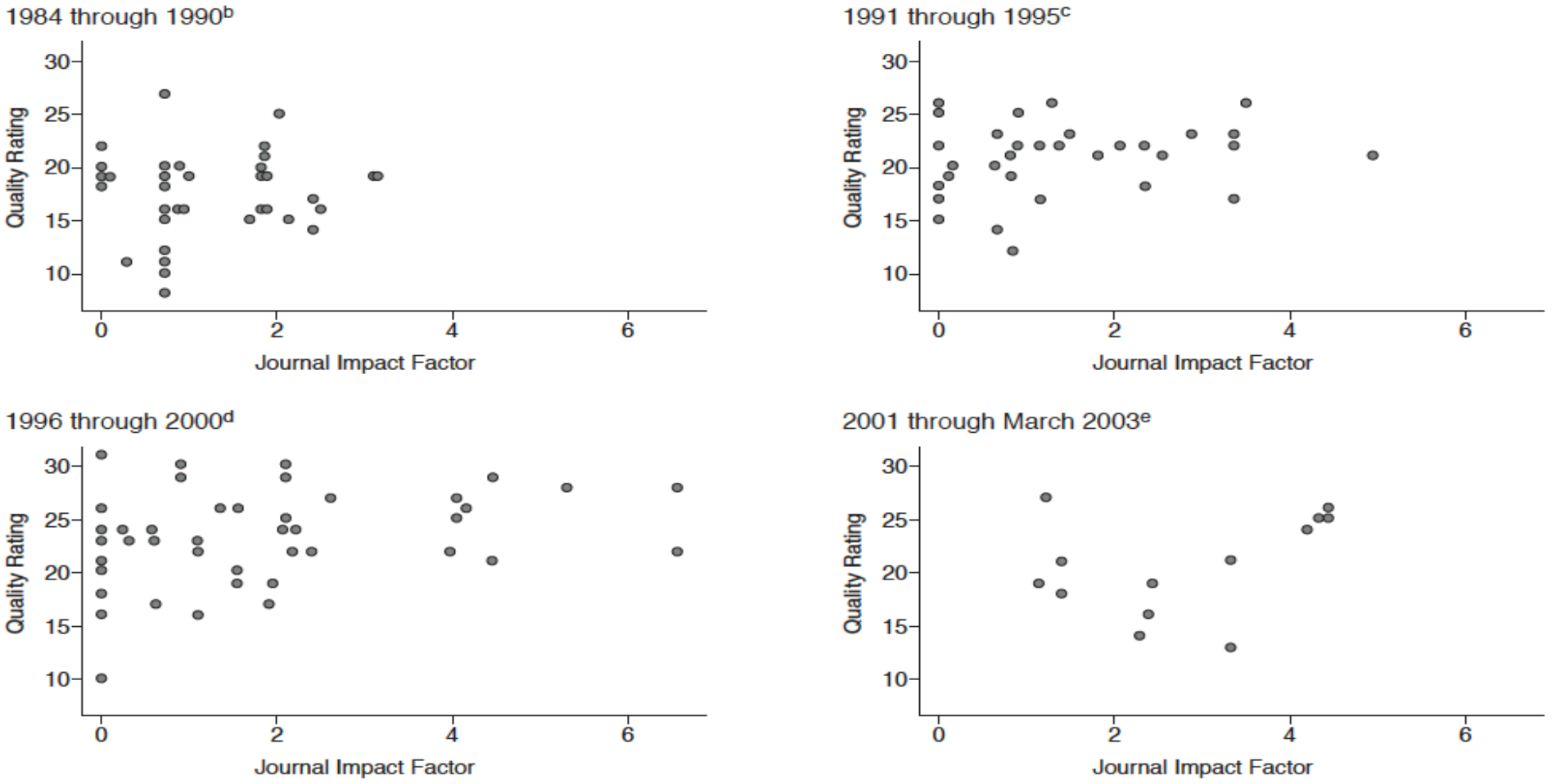


Figure 1. Correlation Between Trial Quality and Impact Factor of Journals Where These Trials Were Published, Stratified by Period of Publication[a]



"The impact factor is not a valid measure of randomized controlled trial quality" (J Clin Psychiatry 2006;67)

Nobel winner declares boycott of top science journals

Randy Schekman says his lab will no longe
Science as they distort scientific process

Do not resuscitate: the journal impact factor declared dead

May 21, 2013 11.51am AEST

It's time to let the journal impact factor die. Ben McLeod

Citation

- Total citation (very often) used as a *gold standard* measure quality
- About 60% of published papers have never been cited (P Jacso, Online Information Review 2009)
- "**Culture**" -- Citation patterns different across fields of research

Problems with citations

- Database dependency
- Does not take into account the author's position
- Citations could be unrelated to quality
- "Cultural factors" (eg US centric)
- Novel papers attract less citations than conventional papers

Time window for citation

Table 3 Spearman correlation with total citations by field (based on dataset 1)

Year	Biology	Biomedical research	Chemistry	Clinical medicine	Earth and space	Engineering and tech	Health sciences	Humanities	Mathem
1	0.174	0.295	0.229	0.258	0.284	0.203	0.244	0.199	0.171
2	0.464	0.657	0.547	0.602	0.622	0.466	0.488	0.407	0.386
3	0.656	0.812	0.739	0.767	0.777	0.636	0.647	0.541	0.571
4	0.752	0.873	0.811	0.844	0.851	0.734	0.741	0.637	0.684
5	0.810	0.906	0.852	0.886	0.888	0.792	0.813	0.711	0.750
6	0.848	0.930	0.881	0.915	0.910	0.835	0.861	0.768	0.795
7	0.874	0.943	0.899	0.930	0.925	0.861	0.887	0.804	0.826
8	0.893	0.953	0.914	0.942	0.937	0.880	0.908	0.832	0.848
9	0.907	0.960	0.926	0.950	0.945	0.895	0.923	0.852	0.868
10	0.918	0.966	0.935	0.957	0.952	0.906	0.933	0.869	0.883

J Wang. Citation time window choice for research impact evaluation. Scientometrics 2013

H index

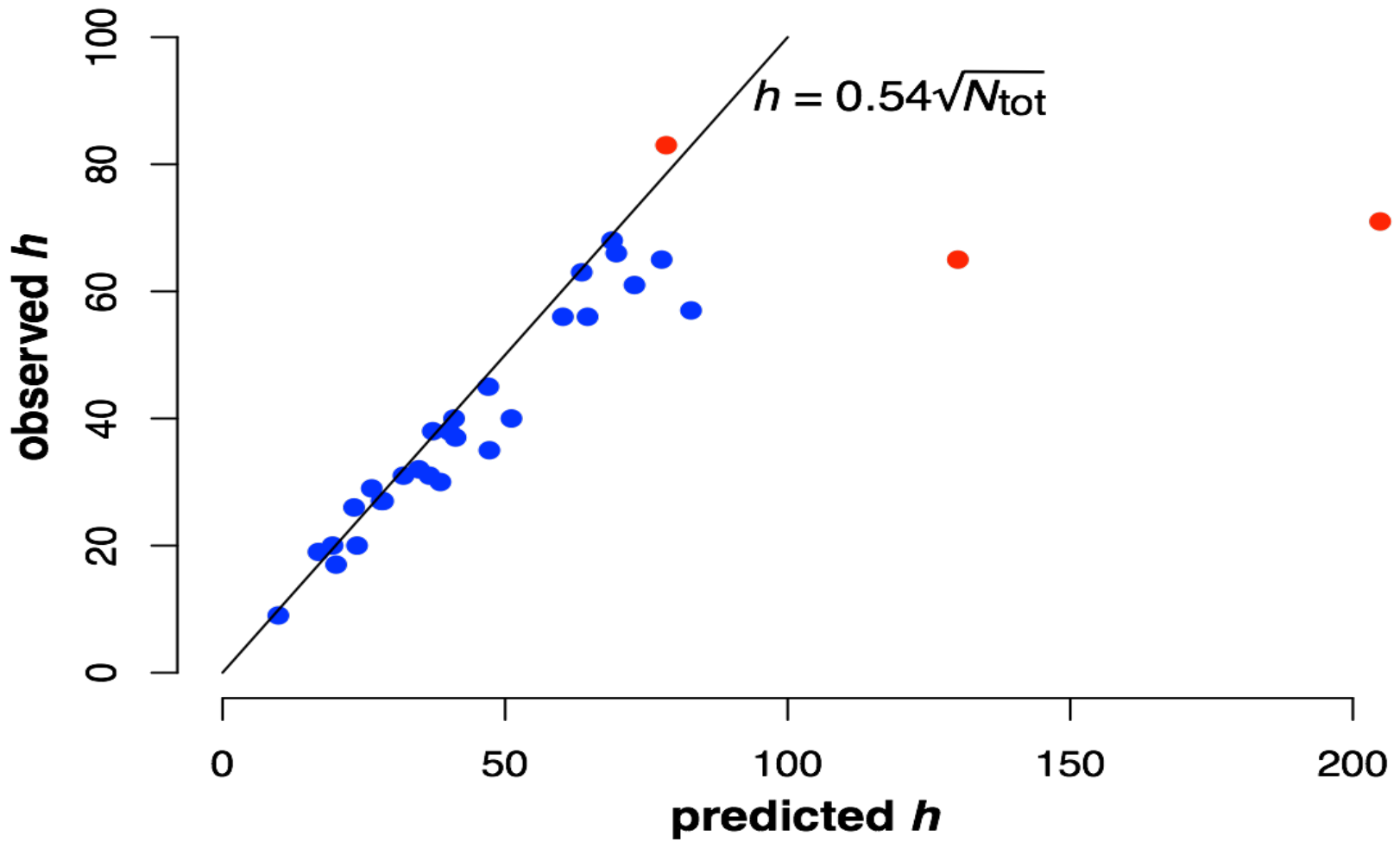
- Preferable to other indices (# papers, citations)
- However, it has deficiencies
  - **Field dependency**
  - Database dependency
  - Never decreased with advancing age → favor old people
  - Affected by the total number of papers

Normalisation of h index to Physics

Iglesias J, et al. Scaling the index for different scientific ISI fields. Scientometrics 2007

ISI Fields	*Power Law*	*Stretched Exponential* 100 papers	200 papers	500 papers	1000 papers
Agricultural Sciences	1.27	1.20	1.24	1.30	1.35
Biology & Biochemistry	0.60	0.77	0.73	0.68	0.64
Chemistry	0.92	0.95	0.94	0.93	0.92
Clinical Medicine	0.76	0.86	0.83	0.80	0.77
Computer Science	1.75	1.97	—	—	—
Economics & Business	1.32	1.23	1.28	1.36	1.42
Engineering	1.70	1.79	—	—	—
Environment/Ecology	0.88	0.93	0.92	0.90	0.88
Geosciences	0.88	0.93	0.91	0.89	0.88
Immunology	0.52	0.73	0.68	0.63	0.58
Materials Science	1.36	1.29	1.35	1.44	—
Mathematics	1.83	—	—	—	—
Microbiology	0.63	0.79	0.75	0.71	0.67
Molecular Biology&Genetics	0.44	0.68	0.64	0.57	0.53
Neuroscience&Behavior	0.56	0.75	0.71	0.66	0.62
Pharmacology&Toxicology	0.84	0.90	0.89	0.86	0.85
Physics	1.00	1.00	1.00	1.00	1.00
Plant & Animal Science	1.08	1.05	1.06	1.07	1.08
Psychiatry/Psychology	0.88	0.93	0.91	0.90	0.88
Social Sciences, general	1.60	1.58	1.72	—	—
Space Science	0.74	0.85	0.82	0.79	0.76

Relationships between h index and citations



Yong A. Critique of Hirsch's Citation Index: A Combinatorial Fermi Problem. *Notices of the AMS* 2014;61:1040-1050

Distribution of H index (biomedical science)

DOI: 10.1111/eci.12171

PERSPECTIVE

A list of highly influential biomedical researchers, 1996–2011

Kevin W. Boyack*, Richard Klavans†, Aaron A. Sorensen‡ and John P.A. Ioannidis§

*SciTech Strategies Inc., Albuquerque, NM 87122, USA, †SciTech Strategies Inc., Berwyn, PA 19312, USA, ‡Temple University School of Medicine, Philadelphia, PA 19140, USA, §Stanford University School of Medicine, Stanford, CA 94305, USA

- 15 million authors (1996-2011)
- 149655 (1%) have h index ≥ 20
- 45752 have h index ≥ 30
- 15385 have h index ≥ 40
- 5185 have h index ≥ 50
- 1773 have h index ≥ 60
- 717 have h index ≥ 70
- 281 have h index ≥ 80

Eigenfactor (EF)

"*Tell me who your friends are, I will tell you who you are*"

- EF takes into account the importance of the journals that cited the work
- AI (Article Influence) = EF / #papers

www.eigenfactor.org

g-index (1)

- Used for distinguishing quality, giving more weight to highly cited papers
- *g* = 20 means that 20 papers of an author have a total citations of at least 400

(1) *Egghe L. Theory and practice of the g-index. Scientometrics. 2006;69:131–152*

View of the Council of Canadian Academies

Indicator	Valid indicator of quality?
Weighted publication counts	Yes
Citation	Yes
External support	No
Esteem measures	No
Webometrics	No
Peer review assessment	Yes / No

Impact

Impact

- Scientific impact
- Societal impact

Societal impact of medical research

- Informing policies (citations on guidelines, govt. policy, development of medicines)
- Building capacity (training; development)
- Relationship between research and health outcomes and cost savings
- Healthier workforce

What should we do?

Assessing individual scientist: productivity

- Research productivity = weighted publication counts

  weight = (EF, citation, rank of journal in the field)

- Make **societal impact** a provision in recruitment and promotion

Assessing individual scientist: quality

- IF is definitely not a good index – don't resurrect it!
- Peer review and esteem indicators are not objective and have many problems
- Citation is more appropriate, but requires time window AND field-specific *normalisation*
- H index may be ok, but must be field-specific normalised and active duration of research

How to reward?

- Do not reward based on IF
- Reward based on
  - citation (year 3)
  - Impact

Final words

- All indices have problems, but some are better than others
- There exists NO perfect metric; we should make the best of the current indices (citation, h index, impact)

Return to essential values of science

IRER Principle

- Importance
- Rigor
- Elegance
- Reproducibility