Panel 4

How Should the Scientometrics of the Chair Candidates be Assessed?

Chair: **Gregory J. Gores, M.D.**[1]
Vice Chair: **Paul Wouters, Ph.D.**[2]
Members: **Alison Abbott, Ph.D.**[3]
Patrick Bossuyt, Ph.D.[4]
Margit Osterloh, Ph.D.[5]

[1] Dean for Research, Mayo Clinic, Rochester, MN, USA
[2] Director of CWTS, Centre for Science and Technology Studies, Leiden University, The Netherlands
[3] Correspondent, Nature, UK
[4] Clinical Epidemiology, University of Amsterdam, The Netherlands
[5] Management Science, Warwick Business School, UK

Correspondence:
Chair: Gregory J. Gores, M.D (gores.gregory@mayo.edu)
Vice Chair: Paul Wouters, Ph.D. (p.f.wouters@cwts.leidenuniv.nl)

Abstract

Although scientometrics are easy to obtain and rank candidates by, they have several deficiencies and should not be heavily weighted in the candidate selection process. Therefore scientometric indicators should be used as supplements to qualitative approaches only. At the individual level, scientometric information can best be used in an integrated form with other types of evidence, such as in individual researcher portfolios or enriched Curricula Vitae. Career development is a key responsibility of a department chair. Thus, having successful mentees reflects positively on the candidate, and therefore should be a facet of the candidates' assessment. Finally, candidates for research leadership positions must be able to delineate an approach for assuring research integrity of the department and institution, should be well educated in research methods, and must be able to set adequate incentives for his or her coworkers.

Introduction

The essential characteristic of a university as an institution is the combination of teaching, knowledge dissemination and research. Those who teach are also expected to advance our understanding of the world. The word "university" itself stems from the Latin universitas scholarium, or universitas magistrorum et scholarium: a corporation or community of masters and students. This definition has two defining elements: the gathering of masters and students, and the fact that they form a community. This community also speaks out, and with some authority, on issues of learning and research.

The leader of this community would typically also be a member of that community. Given the prominent place of science in modern universities, it comes as no surprise that leading figures in a school should be masters themselves: they should be, or should have been, researchers themselves.

So we expect heads of school to have a credible track record in scientific research. As inevitable and self-evident as this seems, putting this principle in practice is less straightforward. How then should one evaluate whether a researcher is qualified as a future head of school? What kind of evidence would demonstrate that the candidate is, or has been, a good researcher?

A causal inspection of practices in schools around the world shows a wide variety in approaches for answering this question. In a sense, these ways also reflect the different value systems and principles for appraising scientific activity altogether. One the one end we find quantitative, bibliometric approaches based on quantitative analyses of scientific output. Although initially developed to search for scientific information, these methods are also used to evaluate research institutions and more recently also individual researchers[1]. Selection committees would simply count the scientific output, expressed in numbers of articles in peer-reviewed journals. The articles could be weighted by the impact factor of these respective journals: publications in high impact journals receive more weight. Alternatively, selection committees may also look at citations: has the work been cited? A variety of citation measures have been used to assess the performance of researchers, among others: the total number of citations, the Hirsch Index, and the Impact Factor of the journals, in which the candidate has published.

On the other end of the spectrum we find selection committees who rely on a more qualitative approach. Here candidates may be invited to submit a kind of portfolio: a selection of key articles, limited in number, which should testify for their qualities as a researcher. The five or ten articles may be accompanied by a description of the history of the research in it, and a description of its reception, how it influenced other research, and improved understanding. At this end, there is no simple criterion for adequate or superior researchers; it is less clear, when a candidate is qualified as an experience, or leading researcher. In this document, we describe the relative merits and shortcomings of examining scientometrics, research integrity, and academic genealogy in selecting chairman candidates in an academic medical center.

Methods

Our team was assembled by the organizing committee for this conference, and met by teleconference to refine our questions. Information was obtained from authoritative works identified by digital searches and known in our respective disciplines. Sections of the document were written by content experts, assembled, and edited by the writing team. The final document was approved by all authors. Recommendations we dichotomized as weak or strong as suggested in the organization committee instructions.

How should the scientometrics of the chair candidate be assessed?

Heads of medical research schools are expected to be exemplary leaders in all aspects of scientific work. In addition to other qualities, they must also have a credible track record in scientific research. Therefore, information about the number of their publications, the quality and influence of those publications, the quality and influence of the journals or publishers with which they have published, the amount of externally funded research they have been able to acquire, as well as their capabilities to lead large research teams and to engage in, and inspire to, collaboration with other scientific researchers at the international level is clearly relevant in the selection process of the best candidates for the position of medical chairs/heads of medical research schools. This includes both descriptive and more advanced forms of statistical information about their track record.

Scientometric information can be used to inform human judgment about the scientific potential of a candidate, but not to replace this expert judgment [2,3]. There are a number of reasons for this. First of all, scientometric indicators always require interpretation within the framework of the specific context of the medical school involved. Three dimensions are particularly important: the mission of the school involved (both in terms of its scientific and its societal goals); the medical fields and sub-disciplines involved; and the specific leadership capacities that are most important for the further development of the school given the institutional and organizational situation. The interaction between these three dimensions will determine, which scientometric information is most important, what types of indicators are needed, and how the scores on these indicators should be interpreted. Second, the coverage of the available databases should be taken into account. The two citation indexes with international coverage are focused on the experimental and natural sciences. Their coverage of the humanities is insufficient and their coverage of the technical and social sciences varies strongly. Practice-oriented journals and books are not included in these databases (with some exceptions). This means, that not all research performed in medical research schools can be measured by the standard scientometric database and standardized analytical tools to which universities may have subscriptions. Third, the technical properties of the available scientometric indicators need to be taken into account. For example, citation indicators such as the total number of citations or the Hirsch Index should not be used to compare research across different medical sub-disciplines because the publication and epistemic cultures between these fields may differ very strongly. A researcher working in a field, where chapters in books are important will by definition have lower scores on these indicators than a researcher in a field that prioritizes publications in international scientific journals. A field that is crucial for medical practice will have a different make-up than a field that is focused on experimental work in the laboratory. The current practice of using those indicators that happen to be the most easily available (most often a version of the Hirsch Index or the Journal Impact Factor), as the sole criteria in the selection process of scientific researchers and in career assessments, must be avoided [2] and DORA (San Francisco Declaration on Research Assessment 2012 [4]. As is explained in the DORA declaration, the impact factor of a journal is not a measure of the citation impact of an individual author.

Fourth, the interpretation of the scientific track record of a candidate should take into account the specific practices regarding authorship of publications [5]. These vary highly by field. In many medical disciplines, the authors of a paper may not all have performed the research. Some forms of authorship are actually forms of acknowledgement of their support (either by delivering data or in institutional ways). It is, therefore, necessary to know the actual role of the candidate in the scientific work that has been published under his name. Often, this information is not (yet) available in the standard publication databases.

This means that selection committees should have some elementary basic knowledge of scientometric information in general, and of the scientometric characteristics of their field in particular, in order to be able to interpret the statistical information in the best way. It is particularly important that they are aware that scientometric indicators are not direct measurements of scientific or scholarly quality. At best, these indicators are proxies that measure part of what counts as quality. Important aspects of scientific or scholarly quality are not measurable. The number of citations is the building block of most scientometric indicators. Hence, it is important to realize that the number of citations is not only determined by the scientific quality of the work. Many other factors may influence the number of citations: the publication venue, the coverage of the database, the international visibility of the research group, the topicality of the research program, the availability of literature reviews in the field (which may capture many citations from original research), the social networking capabilities of the researcher, as well as forms of malpractice such as citation cartels and outright fraudulent publications practices. Some of these factors may be relevant for the selection process, but some of them may not. Therefore, selection committees should organize expert support (e.g, by bibliometric experts or specialists at the university library), if they wish to use scientometric statistical evidence or indicators. The number of citations in research publications is a good measure of the impact the work has had on the scientific community that works in the same field. It may also show interdisciplinary outreach. It is not a direct measure of the contribution to the body of knowledge, which will often take much more time to be settled. Neither is it a direct or indirect measure of all dimensions of the scientific or scholarly quality. The kind of contribution that the publications of a candidate have made to the field can best be determined by experts in the field, who are also the best placed to interpret the publication record of those publications.

A specific aspect of the design of indicators is the level of aggregation at which they are reliable [6]. Most scientometric indicators require substantial amounts of publications and the numbers of publications and citations will often simply be too low at the level of the individual researcher, even if they are - in the context of their field - very productive. Therefore, the most appropriate form in which scientometric information can be used in the selection of researchers for a particular post is the individual scientific portfolio. Basically, this is an enriched form of the Curriculum Vitae. The European research project ACUMEN has developed a possible design for such a portfolio [7]. Scientometric centers have also started to deliver individual researcher portfolios on demand. Such a portfolio gives information about: the expertise of the researcher; the teaching and publication activity; the impact or influence of this work (both scientifically and societally); and last but not least the collaboration network in which the candidate engages (e.g. the network of co-authors; co-PIs on grants; the geographic or disciplinary spread of these authors, etc.).

In these portfolios, scientometric indicators are used as positional information rather than as accountability criteria. In this way, scientometric information on the one hand will enrich the available information the selection committee has at its disposal and may lead to better decisions. It will help reduce bias in selection decisions (which is the main risk, if the selection is only based on intuitive judgment of peers). It may alert selection committees to unexpected aspects of the candidate's performance that they would otherwise miss. And it will help to make the selection process more systematic and thereby increase its quality. On the other hand, scientometric indicators may change the behavior of researchers. They may lead to goal displacement (the scores become the goal rather than the means) or event to attempts to "game the system." In the long term this may undermine the originality and quality of their research [8, 9].

Is the academic genealogy of the chair candidate important?

Disciplines are propagated through knowledge transfer activities. These knowledge transfer activities can be traced using academic genealogy approaches, the organization of a family tree of scholars according to supervisory relationships. Five classifications of academic genealogy have been identified including honorific, egotistical, historical, paradigmatic, and analytic. ***Honorific*** genealogy is usually performed to honor a senior individual such as a "festschrift". ***Egotistical*** genealogy usually follows a single individual or a small set of scholars and traces their antecedents. ***Historical*** genealogy uses academic genealogy as a platform for historical analysis of events and knowledge. ***Paradigmatic*** genealogy is used to define how new ideas are transmitted and involves tracing how ideas move through mentors, collaborators, and colleagues overtime. Finally, ***analytic*** genealogy uses analysis of databases and rigorous statistical analysis. In selecting department chairs it is probably optimal to use both an analytic and an egotistical approach. These combine approaches are not mutually exclusive.

One of the major job descriptions of a department chair is to promote career development of junior faculty. Although academic curriculum vitae and bibliographies usual trace the genealogic roots of the candidate, they seldom are complete, rigorous, or detailed in their approach to assessing the mentees, who have trained with the potential candidate. What we know from analytic genealogic studies is that this mentorship fecundity is most prominent during their first two-thirds of an individual's career and drops off during the later third [10]. Thus, accessing mentees fecundity should not disadvantage the usual chairperson candidate, which often has half to two-thirds of his/her career completed. Interestingly, it is also noted that mentorship fecundity does not correlate well with individual citations in the literature [11]. Thus, measuring mentorship fecundity adds a dimension to assessing the candidates' academic contributions, which may be missed by simply focusing on manuscripts and citations. This approach to assessing candidates for academic medicine has neither been standardized nor subjected to an evaluation. Nonetheless, we propose that candidates do list the people, who have trained with them and their current status. The success of the mentees would reflect positively on the mentorship abilities of the candidate, and conversely, a large number of unsuccessful trainees may speak volumes about the ability of the candidate to evaluate talent and/or develop careers.

We note that a large number of databases are being developed to access academic genealogy such as ProQuest, Dissertations and Thesis database, the Academic Family Tree, Neurotree, etc. Visualizing the results of such information is complex and mimics the structure of genetic family trees. Software available to accomplish this task includes Microsoft Visio, and other programs associated with the databases described above.

How is research integrity of the chair candidate examined?

In the 1990s some spectacular cases of scientific fraud, particularly in clinical and bio medical sciences, prompted funding agencies around the world to formulate guidelines for research integrity (see: www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/ [12]. The various guidelines speak to the necessity of applying rigorous scientific methodology, and creating open and honest working atmospheres, in which young scientists are appropriately trained [13].

Problems, however, continue to emerge, bringing clinical and biomedical research into disrepute. Research agencies are acutely aware of the increasingly public concern that so much of the literature is tainted with results that turn out to be irreproducible, and data sets that turn out to have been manipulated by authors.

In 2005 epidemiologists John Ioannidis [14] provocatively claimed that 'most published research findings are false,' because studies are statistically underpowered. His paper, which concludes that studies atbest capture prevailing bias, became highly influential. Indeed, in 2011, scientists from the pharmaceutical company Bayer reported they could reproduce barely a quarter of some 67 preclinical studies conducted in-house in other pharmaceutical companies [15]. Others, at Amgen, reported that they could repeat only six of 53 basic research studies in cancer [16]. A recent study by EMBO Journal carried out image analysis on submitted papers. It concluded that 20% of the papers had problems such as manipulated gels! Editor Bernd Pulverer wrote that this was either intended to beautify – already poor scientific practice – or, worse, to consciously or unconsciously fit data better to the hypothesis under investigation [17].

Others have shown that basic design of clinical studies can also go wrong, similarly supporting prevailing bias. In a BMJ article earlier this year, cardiologist Darrel Francis et al. analyzed for discrepancies 133 papers reporting on early-phase clinical trials using autologous stem cells to treat heart attack and heart failure. They found that only trials containing design, reporting or statistical errors, or other types of discrepancies, showed positive outcomes. Error-free trials showed no benefit at all [18].

Dismal outcomes are not always the outcome of deliberate scientific fraud. A second elicitor of malpractice is the effect of wrong incentives. If career development and income is dependent on scientometric indicators, like the number of citations or articles published in high-impact-journals, consciously or unconsciously goal displacement sets in. To maximize these numbers, avoidable waste might be produced [15], or sloppy research or even HARKing (Hypothesizing After Results are Known) might be triggered. Moreover, intrinsically motivated "taste for science" is crowded out. Such behavior can be avoided by qualitative evaluations executed by experts in the field. Evaluations by experts provide helpful feedbacks that foster competence and strengthen intrinsic instead of extrinsic incentives. Leaders, therefore, should communicate that they never use scientometric numbers as performance indicators but as supplements to qualitative peer judgments only.

A third elicitor of malpractice is the lack of knowledge on statistical methodologies and ethics. Alarm about the poor reproducibility of biomedical research results led to a hearing by the US Congress in March 2013. In his testimony, former AAAS (American Association for the Advancement of Science) president Bruce Alberts, said that 'research institutes and universities should place more emphasis on short courses, that teach research methodology, ethics, and important technical skills such as how to avoid statistical errors to all their research trainees.'
http://biochemistry.ucsf.edu/labs/alberts/Editorials/Testimony.pdf.

Leadership of medical faculties therefore needs to be committed to research integrity, should be able to set adequate incentives for his or her coworkers, and should be dedicated to a good education in research methods. When examining an individual's candidacy for a chairperson position; their approach to ensuring these characteristics should be explored. Candidates with a well-developed plan or who delineate such a plan will help in protecting the institutions research integrity – a clear leadership objective.

References

1. Moed HF. Citation Analysis in Research Evaluation: Springer, 2005: 348.
2. Adler R, Ewing J, Taylor P. Citation Statistics. In: International Mathematical Union (IMU). International Council of Industrial and Applied Mathematics (ICIAM). Institute of Mathematical Statistics (IMS); 2008. p. 26.
3. Moed HF. The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. Science and Public Policy 2007;34.
4. Plume A. San Francisco Declaration on Research Assessment (DORA) - Elsevier's view. Elsevier Connect 2013.
5. Marusic A, Bosnjak L, Jeroncic A. A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. PloS one 2011;6:e23477.
6. Glanzel W, Wouters P. The dos and don'ts in individual level bibliometrics. In. Vienna; 2013.
7. Wouters P, et.al. Guidelines for Good Evaluation Practives with the ACUMEN Portfolio. http://research-acumen.eu/wp-content/uploads/D6.14-Good-Evaluation-Practices.pdf 2014.
8. Cronin B, Sugimoto CR, editors. Beyond Bibliometrics: MIT press; 2014.
9. Enders J. The Changing Governance of the Sciences. The Advent of Research Evaluation Systems. Sociology of the Sciences Yearbook 2009;47:3.
10. Malmgren RD, Ottino JM, Nunes Amaral LA. The role of mentorship in protege performance. Nature 2010;465:622-626.
11. Sugimoto CR, Russell TG, Meho LI, Marchionini G. MPACT and citation impact: Two sides of the same scholarly coin? Library & Information Science Research 2008;30:273-281.
12. Nielsen OH. [New guidelines on good scientific practice]. Ugeskrift for laeger 2009;171:1847-1849.
13. Abbott A. Science comes to terms with the lessons of fraud. Nature 1999;398:13-17.
14. Ioannidis JP. Why most published research findings are false. PLoS medicine 2005;2:e124.
15. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? Nature reviews. Drug discovery 2011;10:712.
16. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature 2012;483:531-533.
17. Pulverer B. STAP dance. The EMBO journal 2014;33:1285-1286.
18. Nowbar AN, Mielewczik M, Karavassilis M, Dehbi HM, Shun-Shin MJ, Jones S, Howard JP, et al. Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis. BMJ 2014;348:g2688.