Arrogance and ignorance in assessing scientific quality in a multidisciplinary academic medical centre

J.A.E. Spaan

Several academic institutions in the Netherlands and elsewhere develop indices to rank their scientists which will impact evaluation and steering of research. An important part of these indices is based on bibliometric indices. The development of such ranking indices is often seen as the prerogative of management and is kept out of the process where scientific instruments should be presented and evaluated: peer-reviewed journals. In this case the index of the author's institution is criticised both for the evasion of discussion as for the lack of compensation for bias related to discipline, gender and personal history. Furthermore, it is argued that the ranking based on 'numbers' rather than scientific contributions is detrimental to the motivation of the staff suffering under the several modi of bias, is counterproductive for interdisciplinary achievements and discourages young researchers in less scoring disciplines to find their way in the medical academic arena. (Neth Heart J 2010;18:319-22).

Keywords: Academic Medical Centers; Bibliometrics; Biomedical Research; Netherlands Research Personnel

Apparently our managers and policy makers in the academic community feel the need for some objective measure of quality of research, obviously to be used as decision instrument at some point in time. Such a measure will therefore result in some feedback on the choices researchers will

make to optimise their performance in terms of such a measure. It is therefore rather important to discuss publicly the quality of these measures and perform some sensitivity analysis with respect to the parameters that they contain.

The direct reason for this paper is the discussion started by a paper by Opthof and Wilde¹ demonstrating the heterogeneity of the Hirsch index (*h*-index)² in a rather homogeneous discipline of cardiology and in doing so only marginally criticised the quality index developed at the Academic Medical Center (AMC) in Amsterdam. This induced a response from the AMC³ explaining its rationale behind the AMC index. In a response Opthof and Wilde⁴ underlined that the AMC index lacked validation as index for measuring quality of scientists.

It is difficult to avoid the impression of bringing a domestic quarrel into the public arena but that certainly is not the intention of the author. In contrast, I have great solidarity with the institute I have worked in with great satisfaction for so many years. However, the public debate on the definition of the AMC index and the process by which it was introduced in the AMC research community can be useful as a test case for the development of fair judgment of quality of scientists.

The purpose of this paper is to discuss some standards that should be fulfilled in the development of an objective measure of scientific quality and some obvious factors that bias quality indices for which proper compensation is needed. Moreover, the impact that rating of scientists has within an interdisciplinary institution will be discussed.

The need for public debate

I could not agree more with Opthof and Wilde that the discussion about assessment of scientific quality should be public and responsive following standards similar if not identical as holding for scientific publications. This is the first serious criticism to the internally developed 'AMC index' for evaluation of the research performance of PIs (principal investigators). It was developed in a small inner circle without relevant feedback from

J.A.E. Spaan

Department of Biomedical Engineering and Physics,
Academic Medical Center, University of Amsterdam,
Amsterdam, the Netherlands

Correspondence to: J.A.E. Spaan
Department of Biomedical Engineering and Physics, Academic
Medical Center, University of Amsterdam, PO Box 22660,
1100 DD Amsterdam, the Netherlands
E-mail: j.a.spaan@amc.uva.nl

the diverse scientific community within the AMC.⁵ Written criticisms on the 'AMC index' were not replied to in writing by the executive board of the AMC and its research council. There was an oral presentation of this evaluation report within the AMC to an extensive audience. A lot of criticism was ventilated on that occasion, but again without any response in writing and without any effect. Development of such an institutional 'Quality Index' is apparently seen as the prerogative of the executive board and not in need of a serious discussion with its own scientists or the scientific community in general.

As always when scientists are evaluated and ranked, those with a top score keep silent since such a ranking provides them with an 'in home' recognition and, more importantly, with some safety when re-organisation has to be implemented. It is to the 'below average ranked' in such a list to respond. However, the response of those in need of defence is not taken very seriously per definition. Only the high ranked are taken seriously and seem to be recognised as true scientists. It truly is the arrogance of numbers.

Bias in bibliometric indices

Without detailed knowledge of how the AMC index is compounded one may appreciate a few factors that in general require correction in order to avoid bias. In the first place there is not a proper correction by discipline in the index. The *highest* ranked biomedical engineer/physicist within the AMC scores *less than average* in the evaluation. It is difficult to imagine that the creativity and success in this discipline's research deserves that qualification. There is some correction for discipline in the index based on an analysis by the CWTS, Center for Science and Technology Studies, Leiden the Netherlands. However, the field correction of the CWTS does not 'look' below the journal level. This is not justified for such an interdisciplinary institute as the AMC. The citation frequency of different fields covered by one and the same journal is significant. To mention one, the *h*-index cannot only be applied to (groups of) scientists, but also to journals or to topics ('fields', albeit in another way than defined by the CWTS). Thus, in Circulation over the years 2000-2009 the *h*-index of 'atherosclerosis' was 152, whereas that of 'congenital heart disease' was 42. The CWTS correction, by ignoring this huge difference, will obviously be detrimental to those active in the field of congenital heart disease and publishing in Circulation. In other words, each field requires its own reference value to obtain a fairer system. The basic sciences have lower reference values than the clinical sciences and are therefore at a strong disadvantage when the same reference value for the two groups

is used. Hence, the arrogance of numbers results in the arrogance of disciplines and this is detrimental to interdisciplinary research.

A second bias is gender. It is well known that the Netherlands scores very low compared with other European countries when it comes to the percentage of women in high academic ranks.⁶ The AMC index does not correct for gender issues although the AMC does not form an exception on the European evaluation. The internal report only compares the scores of males and females, but this is not the same as correcting for confounding factors. Women have many more obstacles to overcome to be recognised in the world community of science.⁷

The third bias is personal history of scientists. I have nothing against a correction for 'scientific age' in itself. On the contrary, younger scientists should not be hindered by the fact that they have had less time to publish than more senior colleagues. However, 'scientific age' does not take into account events that relate to one or more major switches in research topics or time spent on teaching, clinical work and research within a scientific career. Also, some scientists may have the luck of starting in a successful group with momentum for production of publications with a successful mentor, others have to build their career from scratch, beginning with almost nothing. Moreover, in the past decades journals have worked hard to improve their impact factors and therefore citations for the same type of papers of a few decades ago will contribute less to the present success than the more recent ones. The solution for this need not to be so complicated when restriction is made to a more recent period as suggested by Opthof and Wilde.

In conclusion, one scores best in the ranking as a male, medical doctor or molecular biologist, starting his career in a successful group and without any disruption of that career. A major question is whether these are the factors that underlie creativity and the development of independent scientists?

AMC denial of bias in generated ranking

The internal AMC report⁵ took issue towards the bias-related criticism by claiming that no statistically significant bias towards scientific age, gender and discipline could be found in the AMC ranking dataset. However, this conclusion is too sad for words. All tested groups contained such an enormous internal variability that it was *a priori* excluded to find statistical 'in between' differences. The mere fact that there were no statistical differences between the basic scientists, clinical scientists and epidemiological scientists as whole groups, does not imply that the variability *within* those groups can be taken as a parameter of scientific quality. The applied method is a clear-cut example how

to arrive methodologically correctly at the wrong conclusion. The fact that proven sources of bias in development and evaluation of scientific fields cannot be detected in a dataset does not imply that bias is not playing a role. The method chosen for analysis is to blame, not the absence of statistical significance in testing the wrong model defined by the choice of groups.

Discussion and opinion

A major shortcoming of the AMC index is that it consists of three factors that are supposed to measure different characteristics: Production, Relevance and Viability. However, these three factors are strongly correlated and not independent apart from criticism on the way these characteristics have been measured. Moreover, it is not even attempted to score creativity and the risk the scientist takes in choosing his/her actual research topic. Similarly, the development of a new technique sometimes takes years before becoming productive, which should be taken into account. Ignorance is therefore an important characteristic of such an evaluation procedure.

It makes little sense to state over and over again that scientists working in certain fields score 'better' than others within the same university medical centre. Biomedical engineers within the AMC should be compared with biomedical engineers outside the AMC and oncologists within the AMC with those outside the AMC. It is as simple as that. A ranking system that does not take away the 'scientific field bias' de-motivates younger scientists in the 'under privileged disciplines' to pursue a career within the medical sciences.⁸

Basically, the author has no problems with the position of the high scorers on the list. However, should we blow a whistle when, incidentally, information is obtained of a 'topscorer' whose publication list results in a *h*-index correlating for 84% with the list of someone else? Similarly, what should one do with the incidentally obtained knowledge of a frequently cited paper published more than once? Why should I care, in any case, about other scientists steering the indices which are thought to manifest the success of their careers? Why can a researcher not be provided with an atmosphere to do what he/she is supposed to do: being creative and innovative in the field he/she was asked to explore in the first place?

More importantly, how good are the toppers at the AMC anyway? We have no Nobel Prize laureate that I know of. In the list of *ISI highly cited.com* we have only one representative at the AMC and that is Guido Tytgat, a retired professor. Will we create better performance on these external indices by putting 'all our eggs in one basket'? In my negative moments I think that these rankings were invented so that our managers have something to

talk about when having a drink together.

The worst consequence of the ranking is its counter-productivity in having different disciplines working together. Why would a highly ranking scientist take the effort to understand what a lower ranked 'creature' on that list is doing? Liaisons between groups have been made on opportunistic rather than solid scientific strategies. However, an institute as the AMC needs interaction between disciplines because of the opportunities it creates. We had these opportunities in our different research institutes that were abandoned by the present executive board. I miss the Cardiovascular Research Institute, CRIA. The AMC believes that it is better off by betting on the toppers according to its own index. Our research council primarily has 'topscorers' under its members but with little diversity in disciplines. Undoubtedly, the external jury that will be composed to judge the science done at the AMC will be limited to the coterie of these toppers and basic science disciplines will be judged as irrelevant. The real danger is that a research policy based predominantly on a ranking system as proposed by the AMC will become a self-fulfilling prophecy. The AMC will end up with – at first glance – higher scoring scientists, but will miss research opportunities that should be stimulated so our institute can really distinguish itself in the world of science rather than becoming victim of the 'arrogance of numbers'. The cause of the economic crisis may be a warning for what may become of science when arrogance of numbers becomes the leading factor in science policy.

Let me end by stating clearly: I am not against publishing and even using numbers as *h*-index, number of papers and citations or the impact factors. However, we should not forget that bibliometric indices are used and pushed for commercial purposes by database companies, publishers and institutes offering evaluation services to universities. Therefore, these indices should not be used as a shortcut to evaluate scientific success.⁹ Application of such indices to rank scientists induces perverse feedback mechanisms that negatively affect the general functioning of research.¹⁰ The AMC and the other scientific institutes should abandon the practice of ranking based on a numerical scale. I understand and accept that some evaluation of staff is needed. That indeed should incorporate some sort of citation analysis but this should never become predominant over common sense of leaders that understand their responsibility to science. As things now stand, there is great danger that arrogance and ignorance are going to steer science in the near future. Let's hope for the best.

Conflict of interest

Jos A.E. Spaan is Professor of Medical Physics at the AMC, the University of Amsterdam, Editor

in Chief of Medical & Biological Engineering & Computing, Fellow of IAMBE, AHA and AIMBE. The opinion provided in this paper is clearly influenced by all these relationships. ■

References

- 1 Opthof T, Wilde AA. The Hirsch-index: a simple, new tool for the assessment of scientific output of individual scientists: The case of Dutch professors in clinical cardiology. *Neth Heart J*. 2009;17:145-54.
- 2 Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*. 2005;102:16569-72.
- 3 Van Kammen J, van Lier RAW, Gunning-Schepers LJ. Assessing scientific quality in a multidisciplinary academic medical centre. *Neth Heart J*. 2009;17:500.
- 4 Opthof T, Wilde AA. Assessment of scientific quality is complicated. *Neth Heart J*. 2009;17:501-2.
- 5 Anonymous. Advice of the Research Council on the evaluation of the AMC research 2008. Internal report.
- 6 European Commission ISBN 92-79-01566-4. She figures: Women and Science, statistics and indicators, 2009.
- 7 Symonds MR, Gemmell NJ, Braisher TL, Gorringer KL, Elgar MA. Gender differences in publication output: towards an unbiased metric of research performance. *PLoS One*. 2006;1:e127.
- 8 Spaan JA. Biomedical Engineering and bibliometric indices for scientific quality. *Med Biol Eng Comput*. 2009;47:1219-20.
- 9 Simons K. The misused impact factor. *Science*. 2008;322:165.
- 10 Laloë F, Mosseri R. Bibliometric evaluation of individual researchers: Not even right ... not even wrong. *Europhysics News*. 2009;40:27-9.