Comparison of scientists of the Brazilian Academy of Sciences and of the National Academy of Sciences of the USA on the basis of the *h*-index

R. Mugnaini[1,3], A.L. Packer[1,2] and R. Meneghini[1,2]

[1]BIREME, OPAS-OMS, Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, São Paulo, SP, Brasil
[2]Departamento de Informática em Saúde, Universidade Federal de São Paulo, São Paulo, SP, Brasil
[3]Núcleo de Pesquisa de Produção Científica, Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, SP, Brasil

Correspondence to: R. Meneghini, DIS, UNIFESP, Rua Botucatu, 862, 04023-901 São Paulo, SP, Brasil
Fax: +55-11-5575-8868. E-mail: rogerio.meneghini@bireme.org

A new scientometric indicator, the *h*-index, has been recently proposed (Hirsch JE. *Proc Natl Acad Sci* 2005; 102: 16569-16572). The index avoids some shortcomings of the calculation of the total number of citations as a parameter to evaluate scientific performance. Although it has become known only recently, it has had widespread acceptance. A comparison of the average *h*-index of members of the Brazilian Academy of Sciences (BAS) and of the National Academy of Sciences of the USA (NAS-USA) was carried out for 10 different areas of science. Although, as expected, the comparison was unfavorable to the members of the BAS, the imbalance was distinct in different areas. Since these two academies represent, to a significant extent, the science of top quality produced in each country, the comparison allows the identification of the areas in Brazil that are closer to the international stakeholders of scientific excellence. The areas of Physics and Mathematics stand out in this context. The heterogeneity of the *h*-index in the different areas, estimated by the median dispersion of the index, is significantly higher in the BAS than in the NAS-USA. No elements have been collected in the present study to provide an explanation for this fact.

Key words: Scientometrics; Bibliometrics; Index *h*; Science evaluation; Scientific academy

Introduction

A new scientometric indicator has been recently proposed by Hirsch to evaluate the performance of individual scientists. It has been named *h*-index and is defined as follows: a scientist has index *h* if *h* of his or her *P* papers has at least *h* citations each and the other (*P* - *h*) papers have ≤*h* citations each (1). For example, if 20 papers of an author are ranked by the number of citations, and if the 10 most cited papers have at least 10 citations each and the remaining 10 papers have less than 10 citations, the *h*-index of this author is 10.

Hirsch points to several advantages of the *h*-index versus other indicators. An important one is that the *h*-index is not influenced as much by a small number of "big hits" as it is by the total number of citations, which may not be representative of the individual if, for instance, he or she is a co-author with many others in these papers. This is of great relevance in areas in which large networks of collaborations are required, like genomics, multicentric studies of medicines and particle physics.

From the definition of *h* it can be drawn that $N = ah^2$, *N* being the total number of citations, and *a* a proportionality constant which empirically ranges from 3 to 5 (1). There-

fore, *h* varies with the square root of the number of citations. This is an interesting property of *h* for the purpose of comparing scientists. Although this is not recommended, the performance of individual scientists is frequently weighted solely by scientometric indicators. If a scientist has four times more citations than another, one would be wrongly inclined to assume that the results of the former are four times more important/relevant than the results of the latter. If one uses *h* instead, the difference drops to two times. Although not conceived for this purpose, *h* tends to reduce the differences between scientometric indicator numbers of two scientists to more realistic levels. For example, it is a frequent opinion that Americans tend to cite other Americans (2), an understandable psychosocial behavior. The use of the *h*-index certainly lessens the effects of this trend, a positive factor for comparison purposes.

In this article, we have compared the *h*-indexes of all members of the Brazilian Academy of Sciences (BAS) in its 10 categories with some members of the National Academy of Sciences of the USA (NAS-USA). The BAS is supposedly a good representative of the most prestigious Brazilian scientists and interesting data can be inferred from these comparisons in terms of the relative performance of the 10 categories in the national and international context.

Methodology

Citations to members of the BAS and of the NAS-Institute for Scientific Information (ISI) data were obtained from the Web of Science data base (Thomson-ISI). All data were collected in August 2006. All 389 full members of the BAS at this time were considered in this survey, whereas associate and foreign members were not included. BAS scientists were classified into 10 science categories: agriculture, biology, biomedicine, chemistry, earth, engineering, health, humanities, mathematics, and physics. The classification of the NAS-USA scientists is not the same and therefore groups of American scientists were sampled at random to match the categories of the BAS. The total number of members of the NAS-USA is nearly 2000. The Creative Research Systems (http://www.surveysystem. com/sscalc.htm) sample size calculator was used for this purpose. The confidence level was set at 95% and the confidence interval at 13.2%. To give an idea of the size of the samples of the 10 NAS-USA categories, their average number of members was 41.

The *h*-index for each author was calculated from the citations of all publications listed in the ISI-Web of Science. A preliminary analysis showed that the normal probability distribution could not be used to fit all the data due to the

asymmetry associated with some of them and to the presence of outliers, i.e., *h*-index values that stand out prominently. The logarithm of the *h*-indexes made it possible to produce a smoothing of the distributions which then better fitted two log-based probability distributions: the Lognormal and the Weibull (as determined by the Anderson-Darling test, performed using the Minitab 15.1.1.0 software). The exponential distribution is a special case of the Weibull distribution. In parallel, it was warranted that samples of the same category (BAS and NAS-USA) were fitted along with the same distribution type, either the Lognormal or the Weibull distribution. Since in most cases the distribution was asymmetric and different pairs (BAS and NAS-USA) fitted better one of the distributions mentioned above, the median was taken as the central trend and the median absolute deviation as a measure of dispersion. An interesting property of the Lognormal is that, the closer the fitting, the closer the geometric mean is to the median (3).

Results

If one plots the *h*-index in decreasing order for each author of each category (data not shown), one sees that the decreasing order of the *h*-index is not necessarily followed by the order of citations since the proportionality constant *a* ($a = \sum cit / h^2$) varies in each instance.

To understand the meaning of *a*, Figure 1 illustrates the performances of two BAS scientists in the physics category. At first sight it looks as if the white triangle symbol curve represents the scientist with a higher *h*-index. However, the black triangle symbol curve points to an *h*-index of 22 whereas for the white triangle symbol curve the *h*-index is 19. In fact, the white triangle symbol curve begins with high hits of citations and drops quickly, skewing to values near zero and keeping this trend for more than 100 articles (Figure 1 shows only up to 60 articles). The black triangle symbol curve indicates lower citation values but runs smoothly to values in the range of 10-20 citations for more than 100 articles. Therefore, even with a lower number of total citations (given by the areas under the curves) the black triangle curve scientist has a higher *h*-index than the white triangle curve scientist. The difference between them is reflected by the values of the proportionality constant *a* which is much higher for the latter. High *a* values indicate big hits of citation rates for a few papers followed by a large number of low citation rates. This pattern of the curves is not uncommon and may represent a scientist that obtained these big hits with a fortunate collaboration, a short-term tenure with a strong group or participated in multi-author networks in distinct areas.

Quite clearly, the distinct categories have distinct average h-indexes. This is expected since the h-index reflects the pattern of total number of citations, which varies significantly among different areas (4). Table 1 compiles the results of the 10 categories. It can be seen that the index h medians in the different categories follow approximately a similar trend in both academies: high values for biomedical, health and chemical sciences, intermediate values for physics, biological sciences, and agriculture, and low values for mathematics and human sciences.

Low h-indexes in the humanities are common to the BAS and the NAS-USA. This seems to be due to the traditional mode of communication in this area of knowledge, which makes more use of books and proceedings of meetings when compared to natural sciences. These publications are not covered by ISI. However, a look at Google Scholar shows how citations flow intensely through this path of communication in human sciences. A recent study has shown that Brazilian researchers in the area of human sciences extensively follow this pattern of communication (5), giving little attention to publishing in journals.

In the case of mathematics the situation is different. A preference for journals is clear but the rate of publication is much slower than in most of the other areas. Most of the papers in mathematics are on demonstration of theorems, and this may take years to be accomplished.

Table 2 shows how the median h-index values for members of the BAS compared to those of the NAS-USA in the different categories. Physics and mathematics display the better scores for the BAS, attaining 43% of the median h-indexes of the NAS-USA in both cases. The lowest relative performances are for engineering and human sciences: in the former, the average h-index of BAS members

is only 20% of that of the NAS-USA. The explanation lies in the fact that the engineers of NAS-USA are much more devoted to basic scientific topics in physics and chemistry while the engineers of the BAS are more professionally oriented in their research lines. In the case of human sciences the median for BAS scientists is 19% of that of the NAS-USA scientists. This low score seems to rest on an even more persistent trend among Brazilian human science researchers not to publish their mostly discursive works in journals, while the tendency in the USA is a growing emphasis on quantitative work and on publishing in the more than 2000 authoritative journals currently available.
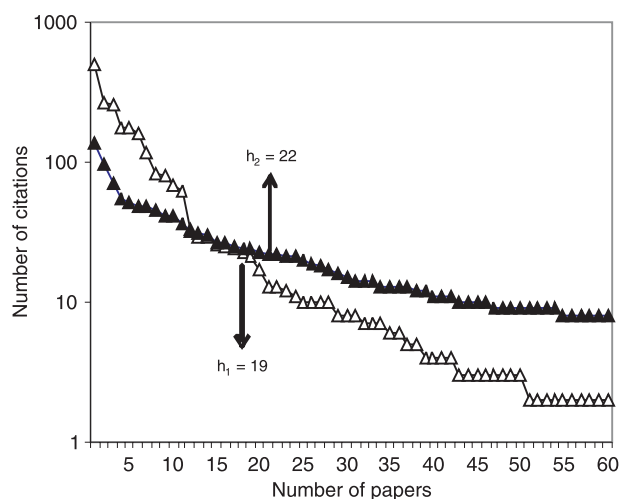

Figure 1. Number of citations as a function of number of papers for two members of the BAS in the area of physics. The purpose is to illustrate how high citations of few papers might not be reflected into higher h-indexes.

Table 1. Median and median absolute deviation of the h-indexes for various categories of members of the Brazilian Academy of Sciences and of the National Academy of Sciences of the USA.

Categories	Brazilian Academy of Sciences			National Academy of Sciences, USA		
	Median	Median absolute deviation	Deviation/ Median	Median	Median absolute deviation	Deviation/ Median
Biomedical sciences	22	7	0.32	66	15	0.23
Health sciences	20	4	0.20	82.5	24	0.29
Chemistry	18	5.5	0.31	56	9	0.16
Physics	16	3.5	0.22	37	9	0.24
Biological sciences	12	3	0.25	44	8	0.18
Agriculture	10	3	0.30	36	9	0.25
Earth sciences	9	4	0.44	37	9	0.24
Engineering	8	3.5	0.44	39.5	18	0.46
Mathematics	8	3	0.38	18.5	5.5	0.30
Human sciences	3	1	0.33	16	4	0.25

One point to be emphasized is the widespread distribution of *h*-indexes among BAS members in each area, using the relative deviation of the median as a dispersion parameter. When comparing the dispersions of the *h*-index between members of the two academies (Table 2, column 3), we notice that they vary among categories, but are on average 30% higher for the members of the BAS, being particularly high for chemistry and earth sciences.

Discussion

The *h*-index has attracted much attention ever since it was proposed in 2005. It has been conceived to be used for individual evaluation purposes, and typical *h*-index values have been suggested for advancement to tenure and full professorship in American Universities and to membership in the NAS-USA. In the latter case $h \approx 45$ has been proposed (1). Apart from resistance by the scientific community to use indexes to assess individual performance in science, the *h*-index has elicited positive reactions (6-9). Furthermore, the possibility of using the *h*-index to evaluate journals, with some evident advantages in relation to the impact factor, has been proposed (10). Proposals to overcome some shortcomings have also been published (7,11).

The good correlation detected by comparing *h*-indexes and peer judgment (9) encourages scientometric analysis based on individual performances measured by the *h*-index. Keeping in mind that a single number cannot measure all nuances of the achievements of a given scientist, it is important to point out that the present article was based on the median *h*-index for collectivities, in relation to which individual *h*-index values can be obtained. The collectivities are all the 389 members of the BAS and samples of members of the NAS-USA (which has more than 2000

Table 2. Comparison and dispersion of the *h*-indexes for the Brazilian Academy of Sciences (BAS) and the National Academy of Sciences of the USA (NAS-USA).

Categories	Median BAS/Median NAS-USA	Deviation/Median BAS/NAS-USA
Biomedical sciences	0.33	1.39
Health sciences	0.24	0.69
Chemistry	0.32	1.94
Physics	0.43	0.92
Biological sciences	0.27	1.39
Agriculture	0.28	1.20
Earth sciences	0.24	1.83
Engineering	0.20	0.96
Mathematics	0.43	1.27
Human sciences	0.19	1.32

members).

The literature is abundant in scientometric studies comparing the accomplishments of nations in science and technology (4). The present article focuses on the performance of the BAS in different scientific categories. This Academy represents the most prestigious (but certainly does not contain all the best) members of the scientific community in Brazil. Comparison between achievements of members of this Academy with that of an internationally recognized scientific academy like the NAS-USA is appropriate for two reasons: first, it allows to set each of the categories contemplated by the BAS in an international context and to ponder on the reasons for imbalances in some areas; second, the *h*-index allows a more realistic comparison of performance, as noted above. The simple fact that it overcomes some shortcomings of total citations, an indicator usually employed for the purpose of such assessments, is one motive for this type of study based on individual performance. Actually, the proportionality of the *h*-index with the square root of the total citations is a corollary of the way the *h*-index is defined. Most of those dealing with scientometric indicators have frequently come across data that let them consider, for instance, whether scientist A, who has N total citations, is indeed twice more proficient than scientist B, who has N/2 total citations. It is very likely that this will not be the conclusion of their peers. Then, scientist A having an *h*-index only 41% (1.41, square root of 2) higher than scientist B is probably a more reliable and acceptable figure.

The widespread distribution of the *h*-index among academics of the BAS - deviation/median 30% higher than for the academics of the NAS-USA - is an indication of a wide heterogeneity in each category. The reasons for this are not in the scope of the present study but certainly deserve attention and further studies. Two points that may deserve attention in future analyses of this considerable dispersion are i) the existence of sub-areas inside the same category with distinct trends of citations, and ii) criteria of member selection of the BAS that do not necessarily correspond to the merit of the performance of the investigator in particular categories.

The comparison of categories between BAS and NAS-USA favors two areas in Brazil, namely physics and mathematics. Physics is a strong scientific area in Brazil, as shown in other studies (12-14). Biomedicine is another area considered to be strong in terms of total Brazilian publications in the mainstream literature (15). However, the citations per paper are much lower than those for the USA. In fact, the prevalence of the USA over all other countries in biomedical research is very well known (4), a fact that should be considered in the comparison between

the members of the two academies in this area.

In some areas, members of the BAS reach *h*-indexes that match those of their colleagues of the NAS-USA. In fact, a few members of the BAS are also members of the NAS-USA. However, these are not necessarily those with high *h*-indexes in their respective categories, indicating that the NAS-USA uses other indicators to select foreign members.

Acknowledgments

References

1. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A* 2005; 102: 16569-16572.
2. Altbach P. The dilemmas of ranking. *Intern Higher Educ* 2006; 42.
3. Blackwood LG. The lognormal distribution, environmental data, and radiological monitoring. *Environ Monit Assess* 1992; 21: 193-210.
4. Moed HF. *Citation analysis in research evaluation*. Netherlands: Springer; 2005.
5. Mugnaini R, Meneghini R, Packer AL. Citation profiles in Brazilian journals of the SciELO database in different scientific areas. *Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics*. Madrid: CINDOC-CSIC; 2007. p 904-905.
6. Popov SB. A parameter to quantify dynamics of a researcher's scientific activity. http://arxiv.org/abs/physics/0508113.
7. Batista PD, Campiteli MG, Kinouchi O, Martinez AS. A complementary index to quantify an individual's scientific research output. *Scientometrics* 2006; 68: 179-189.
8. Bornmann L, Daniel HD. Does the *h*-index for ranking of scientists really work? *Scientometrics* 2005; 65: 391-392.
9. van Raan AFJ. Comparison of the Hirsch-index with stand-

ard bibliometric indicators and with peer judgment. *Scientometrics* 2006; 67: 491-502.
10. Braun T, Glanzel W, Schubert A. A Hirsch-type index for journals. *The Scientist* 2005; 19: 8.
11. Jin BH, Liang LM, Rousseau R, Egghe L. The *R*- and *AR*-indices: Complementing the *h*-index. *Chin Sci Bull* 2007; 52: 855-863.
12. Meneghini R, Packer AL. Articles with authors affiliated to Brazilian institutions published from 1994 to 2003 with 100 or more citations: II - identification of thematic nuclei of excellence in Brazilian science. *An Acad Bras Cienc* 2006; 78: 855-883.
13. Meneghini R. Performance of Brazilian scientists and the pattern of scientific training. A comparison between physicists and chemists. *Cienc Cult* 1995; 47: 45-49.
14. Packer AL, Meneghini R. Articles with authors affiliated to Brazilian institutions published from 1994 to 2003 with 100 or more citations: I - the weight of international collaboration and the role of the networks. *An Acad Bras Cienc* 2006; 78: 841-853.
15. Leta J, Glänzel W, Thijs B. Science in Brazil. Part 2: Sectoral and institutional research profiles. *Scientometrics* 2006; 67: 87-105.