Rev Saúde Pública 2011;45(3)

Julio Cesar Rodrigues Pereira[I]

Bruna Bronhara[II]

H-index of Collective Health professors in Brazil

ABSTRACT

OBJECTIVE: To estimate reference values and the hierarchy function of professors engaged in Collective Health in Brazil by analyzing the distribution of the h-index.

METHODS: From the Portal of Coordination for the Improvement of Higher Education Personnel (*Portal da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*), 934 authors were identified in 2008, of whom 819 were analyzed. The h-index of each professor was obtained through the Web of Science (WoS) using search algorithms controlling for namesakes and alternative spellings of their names. For each Brazilian region and for the country as a whole, we adjusted an exponential probability density function to provide the population parameters and rate of decline by region. Ranking measures were identified using the complement of the cumulative probability function and the hierarchy function among authors according to the h-index by region.

RESULTS: Among the professors analyzed, 29.8% had no citation record in WoS (h=0). The mean h for the country was 3.1, and the region with greatest mean was the southern region (h=4.7). The median h for the country was 3.1, and the greatest median was for the southern region (3.2). Standardizing populations to one hundred, the first rank in the country was h=16, but stratification by region shows that, within the northeastern, southeastern and southern regions, a greater value is necessary for achieving the first rank. In the southern region, the index needed to achieve the first rank was h=24.

CONCLUSIONS: Most of the Brazilian Collective Health authors, if assessed on the basis of the WoS h-index, did not exceed h=5. Regional differences exist, with the southeastern and northeastern regions being similar and the southern region being outstanding.

DESCRIPTORS: Authorship and Co-Authorship in Scientific Publications. Credit system and Researcher Evaluation. Scientific Production. Bibliometric indicators. Public Health. Brazil.

INTRODUCTION

The h-index has attracted wide interest in the academic community since its introduction by Hirsch in 2005.[6] Its attractiveness arises from the possibility to sort scientists on the basis of a single number. This yields an advantage over other indexes that are based on citations, such as those based on the total number of publications, total number of citations or the number of citations per publication.[2] Bibliographic databases such as the Web of Science (Thomson Reuters) and Scopus (Elsevier B.V.) have incorporated this calculation for use in evaluating an author's scientific production. The h-index has become an item on the curriculum vitae (CV) of researchers, as is shown by its adoption by the Lattes Platform of the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (National Council for Scientific and Technological Development).

[I] Departamento de Epidemiologia. Faculdade de Saúde Pública (FSP). Universidade de São Paulo (USP). São Paulo, SP, Brasil

[II] Programa de Pós-Graduação em Saúde Pública. FSP- USP. São Paulo, SP, Brasil

Correspondence:
Julio Cesar Rodrigues Pereira
Departamento de Epidemiologia
Faculdade de Saúde Pública da USP
01246-904 São Paulo, SP, Brasil
E-mail: julio@lee.dante.br

Received: 12/6/2009
Approved: 12/16/2010

Article available from: www.scielo.br/rsp

The h-index quantifies the cumulative production of an author[6], incorporating information about his/her publication record and evaluation by the corresponding scientific community (the impact of citations).[5,12] According to Hirsch's definition[6], "A scientist has index h if h of his or her N*p* papers have at least h citations each and the other (N*p* – h) papers have h ≤ citations each." Therefore, the index measures the number of articles of an author having at least as many citations as the cardinality of the set of articles, e.g., an author who has ten articles published, of which five have at least five citations, has an h-index of 5.

As a bibliometric indicator, the h-index has attracted the attention of Scientometric academics, who have analyzed the advantages and disadvantages of the index and study new opportunities for scientific production modeling. Since 1995, articles analyzing and modeling the index have accumulated in specialized journals: Scientometrics logs 55 of these articles, 23 of which were published in 2009 [search algorithm on WoS: Publication Name=(scientometrics) AND Topic=(H index)]. Journals from various fields of knowledge have devoted editorials to the h-index, and the first editorial was encountered in 2005; 26 editorials in 22 journals were found in WoS [Topic=(H index) AND Year Published=(2008) Refined by: Document Type=(editorial material)].

Despite this interest, the h-value of a given author lacks meaning and does not help in the judgment of merit; this can only be done by comparison with reference values in each field of knowledge. In fact, to contribute semantic content to values of h, Hirsch's original article describes the h-index of notable authors in his field, which is Physics. In Brazil, at least three initiatives for the identification of h-reference values exist.[1,8,10]

In 2006, Batista et al[1] studied Brazilian scientific publications registered by the WoS from 1970 to 2004 for Physics, Chemistry, Mathematics and Biomedical and Life Sciences and determined the highest values of h found in each area. Batista et al.[1] proposed a new indicator, in which the h-index is weighted by the number of co-authors, which attracted wide attention from Scientometric researchers.

Mugnaini et al[10] provided reference values to judge the magnitude of a given h-index when comparing authors of Academies of Sciences of the United States and Brazil in the following fields: Biomedical Sciences, Health Sciences, Chemistry, Physics, Biology, Agriculture, Earth Sciences Engineering, Mathematics and Human Science.

Luz et al[8] found a high correlation between h and other bibliometric indicators in the graduate programs of five institutions of higher education based on the institutional h-index, irrespective of the field of knowledge. In fact, Van Raan[12] found an association not only between different numerical indicators but also with the judgments of peers in research groups in Chemistry.

This study aims to estimate the reference values and hierarchy function of graduate researchers in Collective Health based on an analysis of the distribution parameters of the h-index.

METHODS

The sample size of the scientific production in Collective Health is inaccurate, and it is not identifiable either by institutional affiliation or by publishing vehicle. We examined the set of all graduate researchers in Collective Health in the country to obtain a sample of authors. The names and affiliations of the graduate programs were accessed through the records of the Coordination for the Improvement of Higher Level Personnel (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*) in the public domain on the internet.[a] The following options were selected: 1) Registration of students; 2) Book of indicators and 3) Collective Health for the year 2008, resulting in the sampling of all Higher Education Institutions (HEIs) and their programs in Collective Health in Brazil. For each HEI, we selected the Faculty option, resulting in the assembly of a list of all professors in Collective Health with information regarding their institutional affiliation, field and academic title. These data formed the database on faculty in Collective Health in Brazil.

Publications of professors were sorted based on the number of "times cited" obtained from the WoS database. The h-index obtained on the "citation report" page was recorded. For each name, we considered different versions of name spelling identified in the citations of CV Lattes and in the "author index" of WoS. The main difficulties of this phase were the presence of homonyms and different name formats used in bibliographic citations. Homonym cases were solved by considering institutional affiliation, recognizing the group by co-authors, consistency of the investigation field and comparison with the Lattes database. For the different bibliographic citation formats, we included the possible names by using an asterisk at the end of capital letters, aiming for a more sensitive search. For example, if the fictitious name "João Adalberto Gonçalvez Silva" were registered as Silva J, Silva JA or Silva JAG on CV Lattes, the name would be queried in WoS as Silva J*, and the information used for solving homonyms would be included in the filter page of WoS for searching the author's h-index. In the case of different authors

[a] Ministério da Educação. CAPES – Caderno de Avaliação. Brasília; 2007 [cited 2008 Mar]. Available from: http://conteudoweb.capes.gov.br/conteudoweb/CadernoAvaliacaoServlet?acao=filtraArquivo&ano=2008&codigo_ies=&area=22

having the same name in citations, such publications were excluded, and the h-index was automatically recalculated. Publications were compared with those identified in CV Lattes to ensure the validity of the information obtained.

Search algorithm and validation strategies were tested for each professor from March to November 2008. After query standardization on WoS, we proceeded with the collection of updated data in November 2009.

Figure 1 shows the frequency distribution of h-values based on region and suggests a methodological strategy for analysis. The dotted line describes an exponential decay curve, a Lotka characteristic[4] (Lotka's Law[7]) of the h distribution. The theoretical exponential probability distribution and the Pareto are both able to generalize this type of frequency distribution; we chose the first distribution for the adjustment of events from h=0. The exponential probability density function and cumulative distribution function are described as follows:

$f(x) = \lambda e^{-\lambda x}$ e $F(x) = 1 - e^{-\lambda x}$.

With the assistance of the SPSS statistical package, we fitted the density functions to the frequency data of each region of Brazil. The quality of fit of each function was described by the complement of the residual variance divided by the total variance (R^2 adjusted), and estimates of the decline rate (λ) were assessed based on the 95% confidence interval (95% CI) and the descriptive level obtained using Anova.

To define a hierarchy function of h according to the event, we resorted to the complement of the cumulative distribution percentiles:

$\text{rank}_{1}^{100} = \textit{rounding to}(100e^{-\lambda h})$

Null h values (zero percentiles) corresponded to the last position of a supposed set of discrete and ordered values of 100 h. Values between the 98.5 and 99.49 percentiles indicated first place (both extreme values were rounded to 99 and 100-99=1), and percentiles beyond 99.49 were rounded to zero and considered *hors concours* – very rare occurrences of 0.5% or less. This statistically suggests a strange element in the set, albeit in the sense of positively highlighting the high performance. The second place corresponded to the percentile values between 97.5 and 98.49 (rounded to 98) and so on. We obtained different order positions among authors in a given set that would reduce the total number of authors to 100. This strategy seeks to balance the hierarchy of exceptional authors and authors with h = 1, providing a distance between authors and last place, as such a position should be reserved for those who do not have any cited articles.

RESULTS

The h-index of 934 authors dispersed over the region, HEI and program are described in Table 1.

Figures 2 and 3 show that the southeastern (SE) and northeastern (NE) regions have more programs and professors: we found an average of 35 professors per program in the SE and 22 in the NE regions. The southern (S) region, although having a smaller number of programs and professors, showed an average of professors per program (15/program) that was more similar to the NW than to the SE. There is only one program with 21 authors in the central-western (CW) region. In the northern (N) region, there is a master's degree program in Collective Health at the Federal University of Acre (approved by the National Board of Education [CNE], Ministry of Education and Culture [MEC] ordinance 458, DOU 04-11-2008 – Endorsed CES/CNE 28/2008, 04-10-2008), but there is no "book of indicators" that allows the identification of authors.

In Table 2, we recorded the results of the analysis of the h-index distribution by region and for the country as a whole. For all regions, we reached a satisfactory adjustment to the exponential probability density function with parameter λ and with statistical significance. For the function adjustment to the data of each region, repeated records of authors from more than one program were ignored. The first line of Table 2 reports the number of authors' records contributed by each region.

DISCUSSION

The S and SE regions have the lowest proportion of h-indices equaling zero. However, the SE region has a definite shortcoming, having the greatest rate of decline (28% on average for every unit increase in the value of h). A greater rate of decline indicates a larger drop of probability density from h=0 and, consequently, a reduction of the probability of occurrence of higher values of h. Thus, if h=19 places the author at rank 1 in the SE region, this position would require h=14 in the S region.

After adjusting the exponential probability density function, the regions of greatest similarity are the SE and NE regions: their λ parameters of the density function are similar, with a large overlapping of confidence intervals. As a corollary, their means and medians are similar, as are the hierarchy positions for a given h for these two regions.

The hierarchy function in each region (Table 2) aids in the assessment of the position in a given region and for a particular value of h. For example, for h-index = 10 for a hypothetical author in the SE region, we have the following calculation:

Table 1. Graduate school professors in Collective Health according to region, higher education institution and program. Brazil, 2008.

Region/ Higher Education Institution	Program	nº of Professors	% Region	% Brazil
Central-West				
Universidade Federal de Mato Grosso	Collective Health	21	100	2.2
Total		21	100	2.2
Northeast				
Center of Studies in Collective Health/ Aggeu Research Center	Public Health[a]	22	11	2.4
Magalhães/Fiocruz - Nesc/CPqAM	Public Health	34	17	3.6
Universidade Estadual do Ceará	Public Health	16	8	1.7
Universidade Estadual de Feira de Santana	Collective Health	20	10	2.1
Universidade Federal da Bahia	Collective Health[a]	28	14	3
	Collective Health	37	18.5	4
Universidade Federal do Ceará	Public Health	13	6.5	1.4
Universidade Federal de Pernambuco	Collective Health	15	7.5	1.6
Universidade de Fortaleza	Collective Health	15	7.5	1.6
Total		200	100.0	21.4
Southeast				
Faculdade de Ciências Médicas da Santa Casa de São Paulo	Collective Health	15	2.4	1.6
Oswaldo Cruz Foundation	Maternal and Child Health	18	2.9	1.9
	Women and Children Health	30	4.8	3.2
	Public Health[a]	50	8.0	5.4
	Public Health	141	22.7	15.1
	Public Health and Environment	22	3.5	2.4
Universidade Estadual do Rio de Janeiro	Collective Health	47	7.6	5.0
Universidade Federal do Espírito Santo	Collective Health Attention	15	2.4	1.6
Universidade Federal de Minas Gerais	Public Health	17	2.7	1.8
Universidade Federal do Rio de Janeiro	Collective Health	19	3.1	2.0
Universidade Estácio de Sá	Family Health	17	2.7	1.8
Universidade Estadual Paulista Júlio de Mesquita Fo./Botucatu	Collective Health	23	3.7	2.5
Universidade Estadual de Campinas	Collective Health	28	4.5	3.0
Universidade Federal de São Paulo	Health Sciences	16	2.6	1.7
Universidade Católica de Santos	Collective Health	13	2.1	1.4
Universidade de São Paulo	Preventive Medicine	18	2.9	1.9
	Public Health	117	18.8	12.5
	Community Health	16	2.6	1.7
Total		622	100.0	66.5
South				
Universidade Estadual de Londrina	Collective Health	12	13.2	1.3
Universidade Federal de Pelotas	Epidemiology	12	13.2	1.3
Universidade Federal do Rio Grande do Sul	Epidemiology	23	25.3	2.5
Universidade Federal de Santa Catarina	Public Health	17	18.7	1.8
Universidade Luterana do Brasil	Collective Health	17	18.7	1.8
Universidade do Vale do Rio Sinos	Collective Health	10	11.0	1.1
Total		91	100.0	9.7
Brazil	34 programs	934		100.0

[a] Professional Master's degree

Table 2. Characteristics of the h-index distribution of graduate degree professors in Collective Health. Brazil, 2008.

Geographic Area	Central-West	Northeast	Southeast	South	Brazil
nº of authors	21	171	542	90	819
Proportion of authors with h=0 (%)	47.6	34.9	28.0	25.6	29.8
h: minimum	0	0	0	0	0
h: maximum	2	17	26	51	51
λ(rate of decline)	0.50	0.31	0.28	0.22	0.33
Lower bound (IC 95%)	0.24	0.25	0.25	0.18	0.29
Upper bound (IC 95%)	0.77	0.36	0.30	0.25	0.36
Descriptive level of λp(Λ > λ)	1.1E-02	2.0E-09	6.2E-16	3.3E-12	3.9E-29
Adjusted density function	$0.5e^{-0.5h}$	$0.18e^{-0.31h}$	$0.28e^{-0.28h}$	$0.17e^{-0.17h}$	$0.26e^{-0.26h}$
R^2	0.86	0.89	0.96	0.89	0.84
h: average (1/λ)	2	3.3	3.6	4.7	3.1
h: percentile 25 ln(0.75)/- λ	0.6	0.9	1	1.3	0.9
h: median (ln(0.5)/- λ	1.4	2.3	2.5	3.2	2.1
h: percentile 75 ln(0.25)/- λ	2.8	4.5	5	6.4	4.3
h: percentile 95 ln(0.05)/- λ	6	9.8	10.8	13.9	9.2
Hierarchy function: complement of integral of $f(h)$[a]	$rank_1^{100} \approx (100e^{-0.5h})$	$rank_1^{100} \approx (100e^{-0.31h})$	$rank_1^{100} \approx (100e^{-0.28h})$	$rank_1^{100} \approx (100e^{-0.17h})$	$rank_1^{100} \approx (100e^{-0.26h})$
h of rank #1	9	17	19	24	16
Exceptional h-index (above of percentile 99.5)	11	18	20	25	17

[a]≈ refers to rounding to integers

$rank_1^{100} \approx (100e^{-0,28.10}) = 6$.

This means that if there were 100 authors in Collective Health in the SE region, this specific author would be ranked sixth. In this region, h=10 corresponds to the 93.92 percentile, whose complement 6.08 yields 6 when rounded. In the CW region, whose average h-index is 2, h=10 corresponds to the first place tied with the authors with h=9 (in both cases, the rank function yields 1 as the result). In the NE region, this author would be in fifth place, and in the S region, this author would be in eleventh place. Again, there are similarities between the SE and NE regions.

In previous studies,[11] more similarities between the NE and S regions were found. These regions registered the highest annual growth rates of publications and citations, less dispersion of research interests (i.e.,

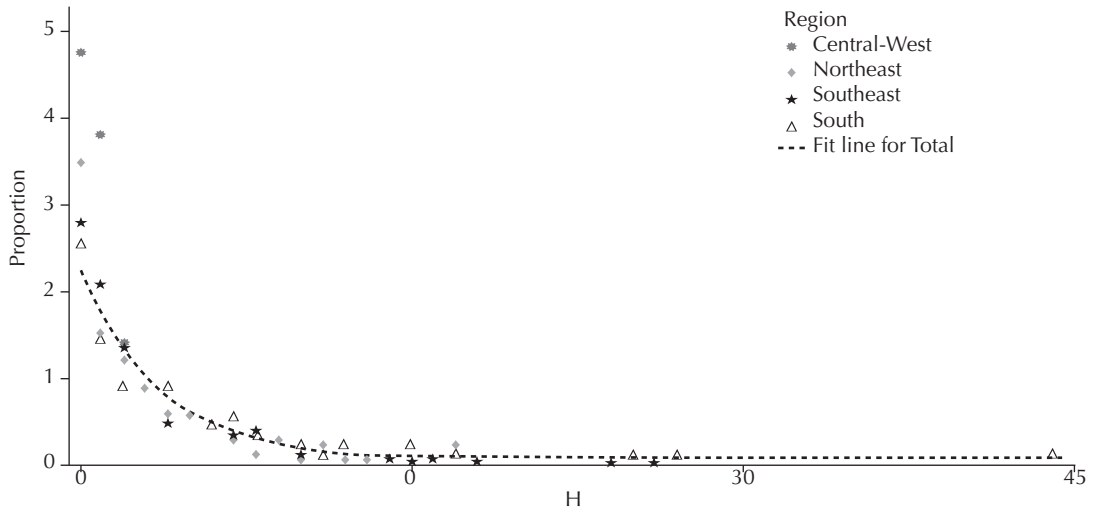

Figure 1. Frequency of the h-index of graduate school professors in Collective Health by region. Brazil, November 2009.

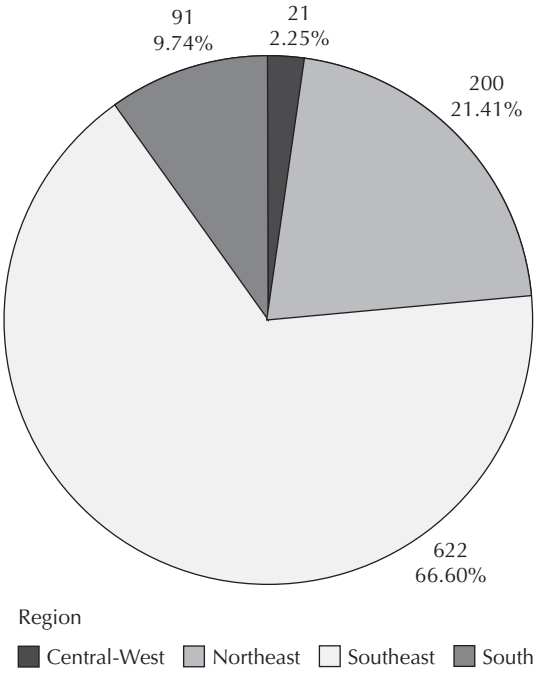

Figure 2. Graduate school professors in Collective Health by region. Brazil, 2008.

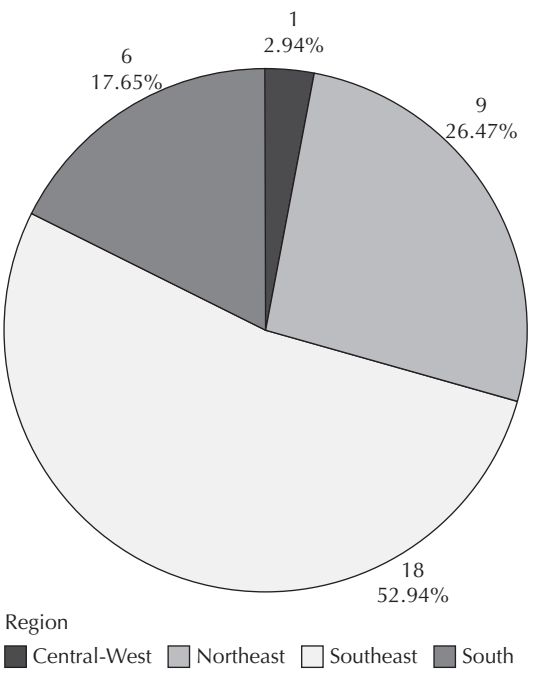

Figure 3. Distribution of graduate programs in Collective Health by region. Brazil, November 2008.

the highest values of the Shannon E-index), a higher proportion of authors cited and a greater engagement in the fields of Experimental and Clinical Medicine. These apparently paradoxical results may be explained based on changes in WoS in 2007 and 2008 (the years separating both studies) in which, seeking to respond to the competition established by Scopus, WoS more than doubled the number of Brazilian journals indexed, with a consequent sudden increase in production records and citations.[b] The earlier study covered WoS records until December 2005, at which time it indexed 26 Brazilian journals. In 2007, that number rose to 63, and in 2008, it reached its current value of 103.[9] WoS also started to record conference proceedings, which should have also extended the recognition of Brazilian scientific production.

However, the dominance of the NE region among the other regions of Brazil is remarkable. The Ministry of Science and Technology has been developing partnerships with research foundations to promote the decentralization of national scientific production, with increased investments in scholarships for states in the N, NE and CW regions. Since it was created in 2003, the Regional Scientific Development Grant (*Bolsa de Desenvolvimento Científico Regional*) has aimed at attracting and retaining doctors from deprived areas of the country. In 2007, the Brazilian government implemented the law 11.540/2007, which regulates the National Fund for Scientific and Technological Development. According to this law, at least 40% the total funds allocated to the Ministry of Science and Technology will be applied to programs promoting the qualification and the scientific and technological development of the N and NE regions, including their areas of regional development agency coverage. Initiatives such as this can explain the scientific distinction in Collective Health reached by the NE region.

As a limitation of this study, authors in Collective Health in Brazil may not be perfectly represented in the population studies, because these studies were restricted to graduate programs. Brazilian scientific production has had a significant contribution from Public Health professionals who, being exclusive to the management of the Unified Health System (*Sistema Único de Saúde*), preserve the interests of this research. Examples of such production include various publications, such as the journals administered by the Ministry of Health, e.g., the Epidemiological Bulletin, Mental Health Bulletin, HR Health Book and others.[c] However, analysis of the h-index behavior of graduate researchers may have provided reference values for evaluation or comparison purposes of the cumulative scientific production of each region, and this can be used as a reference for judging output.

[b] Meneghini R. Inusitado aumento da produção científica. In: Tendências e Debates. *Folha Sao Paulo*. 12 de maio de 2009, p.3.
[c] Ministério da Saúde. Periódicos Institucionais. Brasília;[s.d.][citado 2011 mar 21]. Disponível em: http://bvsms.saude.gov.br/php/level.php?lang=pt&component=44&item=79

The value of the h-index from the 'citation report' page underestimates the real value of h of authors whose works are not part of the publication records of WoS. The estimate of h can be refined via a 'cited references search', which will also be limited to citations of published articles that are registered in WoS. Any inaccuracy of this metric does not compromise comparisons of measurements taken under the same assumption. The h-index can also be obtained on BV Scopus and Google Scholar, resulting in different values. It is thus inappropriate to compare values of h from different sources.

The h-index has limitations that are the basis for a critical interpretation of the scientific production of an author. Examples are its dependence on the number of years of scientific activity,[6] which hinders comparisons of the h-index of young researchers with that of older researchers, an excessive use of self-citation (which can inflate the value of the h-index)[13] and the possibility of underestimating the production of "selective authors", i.e., authors who publish fewer papers but ones that have remarkable international impact and receive many citations.[3] Moreover, evaluation of the productivity of scientific researchers cannot be restricted to the use of a single indicator. A single number cannot provide more than a rough approximation of an individual's multifaceted profile, and many other factors should be considered in combination when evaluating a researcher.[6] The h-index is a tool to evaluate scientific researchers.

The previous[11] and present studies agree in concluding that the NE region has equaled the "Sul maravilha" ("southern wonder"), a phrase coined by Henfil (Henrique de Souza Filho, 1944 – 1988). If he were still alive, maybe his character Grauna would acknowledge a "Nordeste maravilha" ("northeastern wonder"), at least in Collective Health.

REFERENCES

1. Batista PD, Campiteli MG, Kinouchi O, Martinez AS. Is it possible to compare researchers with different scientific interests? *Scientometrics*. 2006;68(1):179-89. DOI:10.1007/s11192-006-0090-4
2. Bornmann L, Hans-Dieter D. What do we know about the h index? *J Am Soc Inf Sci Technol*. 2007;58(9):1381-5. DOI:10.1002/asi.20609
3. Costas R, Bordons M. The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level. *J Informetrics*. 2007;1(3):193-203. DOI:10.1016/j.joi.2007.02.001
4. Egghe L. Modelling successive h-indices. *Scientometrics*. 2008;77(3):377-387. DOI:10.1007/s11192-007-1968-5
5. Glanzel W. On the h-index – a mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*. 2006;67(2):315-21. DOI:10.1007/s11192-006-0102-4
6. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*. 2005;102(46):16569-72. DOI:10.1073/pnas.0507655102
7. Lotka AJ. The frequency distribution of scientific productivity. *J Wash Acad Sci*. 1926;16(12):317-24.
8. Luz MP, Marques-Portella C, Mendlowicz M, Gleiser S, Coutinho ESF, Figueira I. Institutional h-index: the performance of a new metric in the evaluation of Brazilian Psychiatric Pos-graduation Programs. *Scientometrics*. 2008;77(2):361-8. DOI:10.1007/s11192-007-1964-9
9. Marques F. Muito calor, pouca luz. *Pesqui FAPESP*. 2009;160:28-30.
10. Mugnaini R, Packer AL, Meneghini R. 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. *Braz J Med Biol Res*. 2008;41(4):258-62. DOI:10.1590/S0100-879X2008000400001
11. Pereira JCR, Vasconcellos JP, Furusawa L, Barbati AM. Who's Who and what's what in Brazilian Public Health Sciences. *Scientometrics* 2007; 73(1): 37-52. DOI: 10.1007/s11192-007-1787-8
12. Van Raan AFJ. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*. 2006;67(3):491-502. DOI:10.1007/s11192-006-0066-4
13. Zhivotovsky LA, Krutovsky KV. Self-citation can inflate h-index. *Scientometrics*. 2008;77(2):373-5. DOI:10.1007/s11192-006-1716-2

Research funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (Process nº 308502/2006-0).