H-Index of the Citing Articles: a Contribution to the Evaluation of Scientific Production of Experienced Researchers

Claudio Gil Soares de Araújo1,2
Aline Sardinha1,4

1. Graduate Program in Physical Education, Gama Filho University. Rio de Janeiro, RJ - Brazil
2. Exercise Medicine Clinic, CLINIMEX. Rio de Janeiro, RJ - Brazil
3. Panic and Breathing Laboratory, Psychiatry Institute, Federal University of Rio de Janeiro. Rio de Janeiro, RJ - Brazil
4. National Institute of Science and Technology – Translational Medicine (CNPq), RJ - Brazil

Mailing address:
Rua Siqueira Campos, 93/101 – 22031-070 – Rio de Janeiro, RJ, Brazil
E-mail: cgaraajo@iis.com.br

ABSTRACT

Background: there has been an increasing interest and need in evaluating the quality of scientific papers written by researchers. Therefore, the most important measurements used are the number of citations, the average number of citations per article published in indexed journals and the H-index. Those bibliometric indicators present, however, limitations when assessing the potential impact of the publications of a given researcher, particularly among the most productive ones. Objective: To propose and to demonstrate the use of a new scienciometric strategy – the H-index of the citing articles -, that allows better discrimination of the impact of a given researcher to the body of knowledge in the respective research area. Methods: Research data – Web of Science - from 13 of the most productive Brazilian PhD researchers in the field of exercise and sports sciences were analyzed, including: number of publications, number of citations, average number of citations per article, H-index and the proposed H-index of citing articles (the H-index extracted from the ranking of articles that cited the papers published by a given researcher). Data from four other researchers – Brazilian and foreign – were used as reference for comparisons. Results: The researcher’s H-index and the H-index of the citing articles were associated (r=0.92; p<0.01). However, when the six most productive authors with H-index above 7 were analyzed, the association between the two indexes disappeared (r=0.35; p=0.49). Conclusions: The H-index of the citing articles can be useful to discriminate the scientific production of authors with a high number of published papers. It is suggested that this algorithm should be adopted by Brazilian and foreign financing and scientific production evaluation agencies.

Keywords: scientometrics, citations, research, evaluation, impact factor.

INTRODUCTION

The production of scientific knowledge is considered fruitful when the results and innovations brought by the research are able to reach the scientific community and general public as well. The habitual way of making the research results and new proposals and scientific analyses available is through the publishing of articles in specialized vehicles, commonly termed scientific journals. Generally, such vehicles are characterized by rigid editorial policy, supported by academically skilled and impartial editors and revisors, with the aim to select and allow only the publishing of methodologically valid and scientifically relevant articles. From this point, a process in which the researchers try to have their contributions in excellence and with high credit journals published is established. However, what does define the researcher or a scientific journal quality? In order to try to answer this question, a need to generate objective assessment criteria emerges.

Science metrics – or scientometrics – is a field in the information sciences which has become increasingly popular among the scientific providing agencies, the higher education institutions and the researchers of many different knowledge areas1-3. In fact, as reported by Bergstrom4, the analysis of bibliography citations has already been reported in 1927, but it was only through a proposal made by the researcher Eugene Garfield and published in the Science journal in 19555, that this bibliography metrics has become the basic reference6. Thus, the inclusion of information or data of a given article published in the text of a new article, and consequently its citation on the bibliography reference list of this new article is currently the main element present in the most used algorithms of scientometrics6,7.

From the increasing information systematization and more recently the use of powerful mathematical tools, it has become objectively possible to find, quantify and value the number of times a given scientific article is subsequently cited. As a natural consequence of this metrics, the impact factor of the journals and their multiple variations, representing in a simplistic analysis, the mean of the citations of the published articles by a given journal in the two years following its publishing, has appeared8. The discussion on the positive side and the flaws or limitations of the impact factor dates from long ago9-14, but still remains on the spotlight, as proved by many relevant and fairly recent articles15-22. In this context, it is relevant to highlight that the citations of the articles published by Brazilian authors in national journals have already been analyzed for many years based on the data available by the ISI system (Institute of Scientific Information)15,23-28.

Thus, it is interesting to observe that many health-related Brazilian journals, specially the most qualified ones, sometime on the last years, have published some articles on this issue, either to present a point of view22,29-36 or to report the first factor impact.
calculated by the Web of Science\(^{(37-39)}\). The issue of the impact factor has taken an even greater meaning in Brazil, since it has been used and extremely valued by the Brazilian governmental providing agencies and especially, in the evaluation of scientific production of the professors linked to the post-graduation programs by CAPES (Coordination for the Advancement of Higher Education Teaching)\(^{(40)}\). However, this practice has been suitably questioned by the main editors of Brazilian scientific journals, notably those which act in the health field\(^{(21,33,41,42)}\) and even by the lay press\(^{(43)}\).

Although some of the main indicators of scientific production have shown to be sufficiently consistent and valid, able not only to resist to massive criticism for over 50 years, but also become progressively notorious and increasingly used, it is natural that other indicators of scientific production have come out in the last years. Among the new indicators, the\(H\)-index proposed by the physicist Hirsch in 2005\(^{(46)}\) and its subsequent variations\(^{(45-47)}\), is perhaps the one which has attracted the most attention in the academic field. The \(H\)-index, of less intuitive comprehension, also makes use of citations obtained by the published articles without establishing a temporal limit though, being hence the citations obtained from its publishing or availability on the site of the journal (publish ahead) counted. The \(H\)-index consists of the number of published articles which received citations with a number higher or equal to \(i\)\(^{(25,48,49)}\) and can be applied both for an individual and group of researchers\(^{(25,50,51)}\), as well as journals\(^{(52)}\). For example, a researcher who has an \(H\)-index 10, implies that he/she has published at least 10 articles which received 10 or more citations from its publishing or availability (publish ahead). With conceptualization very close to the central tendency measurement known as median, the \(H\)-index is not influenced by extremes, as occurs for instance with the mean of citations per published article. Since it occurs with its basic measure or fundamental unit – the citation –, the \(H\)-index should be contextualized by area or subarea of knowledge for results comparison or interpretation\(^{(20,47)}\).

Although there are positive points in the metrics described above, the analysis remains being primarily quantitative, without a more qualitative component which allows valuing the academic merit or highlight. In fact, it is relevant not only that a published article becomes cited, but also especially who is citing it and in which journal the article is being cited. One of the most recent alternatives for this issue is the proposition and analysis of the Eigenfactor\(^{(45)}\) (www.eigenfactor.org, for further information), which incorporates a quality pondering based on one algorithm; however, its high mathematical complexity makes it extremely difficult to simply read data and minimize its wider application.

In this same logic of metric analysis, it can hence be verified the potential multiplying and disseminating factor of a given scientific production by the number of citations the articles cited are obtaining. This metrics can be defined as the \(H\)-index of the citing articles and can be obtained in a relatively simple and fast manner through the analysis and decreasing order of data made available on the Web of Science – Thomson-Reuters.

In practice, we can see that some researchers accumulate an important number of citations. However, there are those whose articles are frequently cited by exponent pairs and in high impact journals, and others whose the respective articles are cited in its great majority in journals of lower impact or less rigid editorial policy. Thus, it may be appropriate to identify the scientific metrics which allows evaluating not only the quantitative aspect of the citations but also which aggregates a valuing qualitative component.

The aim of the present article is to propose and demonstrate the discrimination potential of a new bibliography metrics – the \(H\)-index of the citing articles – through the results of data analysis of a selected sample of Brazilian researchers in the physical exercise and sports field.

**METHODS**

Seventeen authors were selected for analysis of the citations and data of scientific production. Out of the 15 Brazilian authors, 13 primarily research on the biological field of the physical exercise, with Physiology and/or Exercise and Sports Medicine themes. These 13 PhD researchers (10 of which are beholders of CNPq productivity in research scholarships – levels 1A and 2), range in age and career time, from senior advisors, and are probably some of the most productive in the country in the field, to young novice Doctors. The data of four other authors of the health field – two foreigners and two Brazilians – were obtained to complement the analyses and allow the establishment of some reference points for comparisons.

On December, 2010 several metrics relevant to the study were researched on the data base Web of Science (Thomson-Reuters), accessed through the CAPES journal link, including: number of listed articles, number of citations, number of citations of the most cited article, mean of citations per listed article, \(H\)-index of the researcher, number of articles published in the 2006-2010 quinquennium, number of articles published in 2010 and number of citations in 2010. Using tools from the system itself, a list of articles which cited the researcher’s articles was carefully identified and compiled and subsequently ordered from the highest to the lowest number of citations, allowing hence visual \(H\)-index of the citing articles; that is, the \(H\)-number of articles which received at least \(H\)-citations. Special care should be taken to avoid that occasional synonyms would artificially inflate the data. Some complementary data on the undergraduate field, higher education institution of current relationship and status as CNPq scholarship beholder were obtained from the analysis of the resumes available on the Lattes Platform.

The statistical analyses were limited to conventional descriptive procedures and the determination of Pearson correlation coefficients. Whenever appropriate, the significance criterion of 5% of probability was applied. The Prism 5.04 software (Graph-Pad, USA) was used for figures calculations and designing.

**RESULTS**

Table 1 illustrates the results of the study for the 13 Brazilian researchers in the physical exercise and sports field. The profile and volume of scientific production, a result of conscious choice of the researchers, substantially ranged from seven to 199 articles published and listed on the Web of Science, including from two to 1,106 citations in the entire professional career. The \(H\)-index of the researchers also ranged in 20 times, going from one to 20.

For 12 out of the 13 researchers, the \(H\)-index of the citing articles is higher than the \(H\)-index of the researcher him/herself,
and can be up to three to four times higher. Figure 1 presents that there is a tendency of behavior similarity in the data of the H-indices of the researcher and the citing articles, reflected by a correlation coefficient of 0.92 (p < 0.001); however, when an analysis is performed using only the data of the six researchers with the highest scientific production (Figure 2) and, therefore, with higher H-indices, it is observed that the association between the two metrics of the H-index disappears (r = 0.35; p = 0.49), indicating that for the researchers with H-index higher than 7, the H-index of the citing articles may be extremely variable in practice, ranging between 12 and 32. Table 2 presents the associations observed between the several scientific metrics and the H-index of the citing articles for the 13 Brazilian PhD researchers of the physical exercise and sports field who were analyzed in the present study.

Looking back at the data in table 1 it is possible to observe that a specific foreign researcher very active in the Sports Cardiology field – certainly one of the most worldwide productive in terms of scientific publishing – has 830 indexed articles on the Web of Science, 20% of which in the last five years, totaling 33,523 citations. The most cited article by this author has 622 citations and his articles have mean citations of a little more than 40 times and his H-index is 97. For this researcher, due to limitations intrinsic to the database, it is not possible, through the access available by the CAPES journals link, to safely calculate his H-index of the citing articles. On the other hand, for a Canadian researcher publishing on Neurology themes and with article publishing concentration in the last five years, we identified 243 articles on the Web of Science and 2,528 citations and found H-index of 30 and H-index of citing articles of 56. Analyzing the obtained data of two Brazilian researchers from another health field, a CNPq productivity in research scholarship beholder and a PhD student – the production of both is almost all concentrated in the last five years –, it is observed that the scientific productivity metrics seems to be distinct from those observed in the exercise and sports field, with H-indices fairly high, both of the researcher and the post-graduation student – 12 and 2, respectively, as well as of the citing articles – 47 and 2, respectively.

Table 1. Bibliography indicators of 13 researchers with a PhD degree in the physical exercise and sports field.

<table>
<thead>
<tr>
<th>Researched subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNPq scholarship holder</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Indexed articles in the ISI</td>
<td>199</td>
<td>82</td>
<td>64</td>
<td>36</td>
<td>45</td>
<td>17</td>
<td>31</td>
<td>22</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Citations in the ISI</td>
<td>1106</td>
<td>850</td>
<td>435</td>
<td>381</td>
<td>161</td>
<td>194</td>
<td>64</td>
<td>59</td>
<td>74</td>
<td>27</td>
<td>21</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Most cited article</td>
<td>75</td>
<td>137</td>
<td>62</td>
<td>116</td>
<td>32</td>
<td>53</td>
<td>15</td>
<td>21</td>
<td>21</td>
<td>14</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Mean of citations</td>
<td>5.6</td>
<td>10.4</td>
<td>6.8</td>
<td>10.6</td>
<td>3.6</td>
<td>11.4</td>
<td>2.1</td>
<td>2.7</td>
<td>7.4</td>
<td>3.4</td>
<td>2.3</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>H-index Pesquisador</td>
<td>20</td>
<td>13</td>
<td>13</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>H-index citing articles</td>
<td>30</td>
<td>18</td>
<td>32</td>
<td>26</td>
<td>12</td>
<td>31</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Articles in 2010</td>
<td>16</td>
<td>28</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Citations in 2010</td>
<td>208</td>
<td>313</td>
<td>67</td>
<td>72</td>
<td>41</td>
<td>55</td>
<td>33</td>
<td>25</td>
<td>16</td>
<td>13</td>
<td>15</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Articles 2006-2010</td>
<td>81</td>
<td>73</td>
<td>21</td>
<td>18</td>
<td>28</td>
<td>6</td>
<td>25</td>
<td>19</td>
<td>6</td>
<td>7</td>
<td>9</td>
<td>19</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2. Correlations between the H-index of citing articles and other scientific metrics (N = 13).

<table>
<thead>
<tr>
<th>H-index of the citing articles</th>
<th>Correlation Coefficient</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articles indexed in the ISI</td>
<td>0.685</td>
<td></td>
</tr>
<tr>
<td>Citations ISI</td>
<td>0.922</td>
<td></td>
</tr>
<tr>
<td>Most cited article</td>
<td>0.893</td>
<td></td>
</tr>
<tr>
<td>Mean of citations per article</td>
<td>0.850</td>
<td></td>
</tr>
<tr>
<td>H index researcher</td>
<td>0.924</td>
<td></td>
</tr>
<tr>
<td>Articles in 2010</td>
<td>0.135</td>
<td></td>
</tr>
<tr>
<td>Citations in 2010</td>
<td>0.499</td>
<td></td>
</tr>
<tr>
<td>Articles 2006-2010</td>
<td>0.265</td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION

It seems increasingly evident that the scientific production metrics is necessary and relevant [8,21]. However, it is still important to consider the many limitations of the currently existing indicators as well as the broadening of horizons in the search for new information or indicators which are able to be obtained from the available data, especially within the Brazilian reality [21,42,52]. The H-index of the citing articles proposal is inserted in this context.

It is relevant and interesting to observe that the stratification of the scientific journals by the CAPES Medicine field in the system known as Qualis, primarily considers the impact factor of the journal. In the area 21, representing Physical Education, Physiotherapy, Phonoaudiology and Occupational Therapy, in a broader and probably suitable position, the journals are stratified both by the impact factor and by the H-index, but keeping a fairly inaccurate differentiation concerning whether or not a journal is from the field.

From the professional point of view, more specifically in the health field, the true impact of an article can be analyzed by the effective incorporation of its proposals, methods, results and conclusions to practice. The biggest example of this incorporation to the professional practice in Exercise and Sports Medicine may be observed in the protocols and nominated scales, such as the Cooper test, the Wingate test, the Astrandnomogram, the Bruce protocol, the Faulkner equation, the Borg scale, the Tanner stage. These examples and many others may have been originally published in original articles which did not necessarily reach a high number of citations, but which the subsequent articles that used them were widely cited. More over, it should be highlighted that these tests were progressively incorporated in the main textbooks of the knowledge field and, more importantly, inserted in the professional practice.

In practice, the logic of the process works in the following way: the journals tend to positively restrict the number of references an original article can present (the rule is less restrictive with review articles). Thus, the authors have to carefully select which articles should be cited to support his/her studies and data analyses and naturally tend to choose those which present higher quality and scientific importance. This is especially valid in the articles submitted to journals with stricter editorial policy, in which there is very critical evaluation of the literature used to support the article by the revisers and associated editors. As a result of this careful and selective process, the main articles with the best possible citations are published in the best journals and will be hence read by the main researchers and professionals around the world. These subjects on their turn, when devising their future studies or speak about their area of expertise, will include data from these new articles and from those which have been used in the citations and so forth, resulting in a progressive and differentiated exposition and high potential of influence from the original articles.

In other words, the true potential which a scientific article presents to increment the knowledge on the field is determined by how important who cites it, and where it is cited. Naturally, the only metrics which does not significantly associate with the H-index of the citing articles is the number of published articles in 2010, since there was not sufficient time for the articles published in 2010 to be cited in articles and these articles would be cited. Therefore, it is natural that the articles published in the last quinquennium also present lower potential to be reflected in the H-index of the citing articles. Conversely, the correlation with the H-index of the researchers is excellent and it is very good with the mean of citations per article and with the number of citations in the ISI, reflecting the tendency of these measurements to be very similar when a broader group of researchers, mixing higher and lower experience or production, is analyzed.

Considering the scientific articles themselves, the simple quantification of the number of citations which an article receives does not allow evaluating or discriminating the multiplying or impact potential in the knowledge status on a given theme or topic. Therefore, the metrics with the H-index presents many advantages, beginning with the best characterization of the central tendency, not biased by the extreme values [16,29,44,53,54], but also for better reflecting the collection of the scientific production of a specific researcher. In our proposal, the H-index of the citing articles, the valuing of the scientific production and/or its potential impact can be even much better evaluated and discriminated, when more experienced and productive researchers are considered. The results presented here indicate that when only the data of more experienced and productive researchers were analyzed, that is, with H-index above 7, the association between the two metrics – H-indices and publishing of the researchers and the H-index of the citing articles of the publishing – was not significant anymore. As occurs with many other bibliography indicators, the metrics proposed here – the H-index of the citing articles – should be contextualized to the knowledge field which is being analyzed, since apparently, distinct fields tend to present different result profiles [48]. Comparative analysis with researchers from other fields or foreign ones with high scientific production also let us identify that, in rare situations, the number of citations may be so high that it makes the determination of the H-index of citing articles impossible. However, at least concerning researchers with H-indices up to 30, with at least 300 published articles and a number of citations lower than three thousand, it still seems possible to evaluate this production by the H-index of citing articles. Thus, almost all the Brazilian researchers in the health field will be able to be assessed by this metrics.

To sum up, the present article, when demonstrating the discriminating potential of the H-index of the citing articles in the differentiation of the scientific production of senior researchers corroborates its use as an additional indicator within the evaluation processes. Finally, it is suggested that this metrics proposed in the present study – H-index of the citing articles – is incorporated in the analysis of the scientific production of the researchers by the main providing agencies as well as higher education institutions and made available by the specific systems of the Web of Science-ISI, the Scopus, the Academic Google and the Lattes Platform.

ACKNOWLEDGEMENTS

The authors thank the financial support by CNPq and FAPERJ through research productivity and PhD scholarships.

All authors have declared there is not any potential conflict of interests concerning this article.
REFERENCES

5. Garfield E. Citation indexes for science; a new dimension in documentation through association of ideas. Science 1955;122:108-11.
14. Dempsey JA. Impact factor and its role in academic promotion: a statement adopted by the Inter-
27. Bressan RA, Miguel EC, Man JJ, Rohde LA, Mercadante MT. We have reached the SJR Rev Bras Pneumol 2005;27:170-1.