An index that rates relevance of scientific work in bio-sciences: The scientific relevance-index

Demosthenes B. Panagiotakos, Fragiskos G. Bersimis[1]

Department of Nutrition and Dietetics, School of Health Science and Education, Athens 17671, [1]Department of Informatics and Telematics, School of Digital Science, Harokopio University, Athens 17778, Greece

ABSTRACT

This work aims to present an objective index that rates scientific relevance (SR) of scientists' published work, the SR index, in a specific thematic field. The proposed index is calculated based on equally or weighted individual parameters that measure scientist's publication record and recognition, using easily accessible and unbiased data from existing bibliometric databases. The application of the SR-index could be in any scientific field; an example is given here on cell and molecular biology field.

Keywords: Scientific relevance, index, bibliometry

INTRODUCTION

To find the most relevant, but also the best, scientists in order to develop a scientific committee for the evaluation, and potential hiring, new faculty members for an Academic Institution or Organization is a very difficult task. Similar to the aforementioned mission, to rate and select the best new faculty member among a group of potential candidates, based on objective, quantitative criteria, is also very difficult.

During the past years, several academic measures have been proposed, with the oldest ones being the number of publications of an individual in a specific area and the total number of citations of the published work that highlights the recognition of the work. More recently, the conditional number of citations after excluding self-citations that highlight the recognition of the work irrespective of the individual's publication record (since it has been observed that scientists with higher number of publications have higher number of citations, too, because they usually cite their own work) or focusing on a specific time period, e.g., the past 5 years, have also been proposed and used. There are also some more complicated indexes,[1,2] such as the *h*-index which measures both the number of publications and the number of citations per publication, as well as the contemporary *h*-index which measures the number of publications and the number of citations per publication within a short-time window,[3] the individual *h*-index normalized by the average number of co-authors,[4] the *m*-index (or *m*-quotient) which has been defined as *h*-index/the number of years since the first published paper of the scientist,[5] the *c*-index that takes into account not only the citations, but also the quality of the citations in terms of the collaboration distance between citing and cited authors, as well as many others with less use. Despite the criticism the aforementioned indices have received, ratings like these have been extensively used to measure the impact of research and to justify tenure and in some cases, funding decisions. At this point, it should be clearly defined that bibliometric measures do not necessarily reflect the quality of a published work. It could be argued that a high-quality paper, by not "well-known" authors or/and Institutions, published in a journal with limited promotion resources, may receive less recognition than another work of similar quality, but written by "well-known" authors

***Address for correspondence:** Prof. Demosthenes B. Panagiotakos
E-mail: dbpanag@hua.gr

Access this article online	
Quick Response Code:	Website: www.jscires.org
	DOI: 10.4103/2320-0057.143688

and published in widely distributed or open-access journals. However, how quality of the publication can be measured is a very difficult question, and it is not going to be answered here.

Despite all the aforementioned considerations, the single use of these measures to rate the most relevant, among a set of scientists who have the "apparent" relevance in a specific scientific area, shares several limitations and biases. Thus, the purpose of this work was to propose a multi-dimensional index (the scientific relevance [SR] index) that rates the most relevant members among a set of *k*-scientists, based on their publication record, as well as the international recognition of their scientific work. The suggested index may be found useful for the selection of optimal election committees in Academic Institutions or Organizations, as well as for rating candidates for faculty vacant positions.

METHODOLOGY

The SR-index is defined based on the following parameters: (a) The total number of publications of the *i*-individual, let N^i, (b) the number of the most relevant publications to the specific scientific field, let v_1^i,[6,7] as defined according to the exact match of the keywords that characterizes the *i*-individual, using a keywords set of an archiving database [Appendix], e.g., Thompson Reuters - Institute of Scientific Information, Scopus, (c) the number of the less relevant publications, let v_2^i, as defined according to the keywords that characterizes the *i*-individual, using a keywords set of the aforementioned archiving databases and (d) the *i*-individuals' h_c^i-index.

Let *k*-be the size of the set of all scientists according to their "apparent" relevance in a scientific field. The maximum value $\max_{i=1,2,..,k}\{*\}$ of scientists is calculated for each of the following elements:

- $\frac{v_1^i}{N^i}$, which denotes the percentage of published work that match in the exact scientific field (i.e., all keywords of individual's published papers are a super-set of the scientific topic)
- $\frac{v_2^i}{N^i}$, which denotes the percentage of scientific work that belongs to a broader scientific field (i.e., some of the keywords of individual's published papers are a super-set of the scientific topic)
- h^i-index of individual's published papers
- h_c^i-index, which is the contemporary h^i-index for the past *m* years.

The SR-index is then defined as follows:

$$\text{SR - index} = \frac{\frac{v_1^i}{N^i}}{\max_{i=1,2,..,k}\left\{\frac{v_1^i}{N^i}\right\}} + \frac{\frac{v_2^i}{N^i}}{\max_{i=1,2,..,k}\left\{\frac{v_2^i}{N^i}\right\}} + \frac{h^i}{\max_{i=1,2,..,k}\left\{h^i\right\}} + \frac{h_c^i}{\max_{i=1,2,..,k}\left\{h_c^i\right\}}$$

The SR-index is a continuous variable that takes values in the interval [0, 4]. Greater values indicate higher SR.

An issue that may arise is that SR-index does not fully incorporate the true SR, neither the international recognition nor the four parameters have the same weighting. Therefore, a weighted SR-index is also proposed here, where the weights b_1,b_2,b_3, and b_4 can incorporate, depending on the values assigned, the true relevance and recognition of an individual's scientific work. Weights have to satisfy the equation: $b_1+b_2+b_3+b_4=1$ and should be $b_j \geq 0, j = 1,2,3,4$. The weighted SR-index is then defined as follows:

$$\text{Weighted SR - index} = b_1 \cdot \frac{\frac{v_1^i}{N^i}}{\max_{i=1,2,..,k}\left\{\frac{v_1^i}{N^i}\right\}} + b_2 \cdot \frac{\frac{v_2^i}{N^i}}{\max_{i=1,2,..,k}\left\{\frac{v_2^i}{N^i}\right\}} + b_3 \cdot \frac{h^i}{\max_{i=1,2,..,k}\left\{h^i\right\}} + b_4 \cdot \frac{h_c^i}{\max_{i=1,2,..,k}\left\{h_c^i\right\}}$$

The Weighted SR-index takes values in the interval [0, 1]. Higher values indicate greater SR among the members of the potential candidates.

Both SR-and weighted SR-index are increasing functions of their components.

An Example

In this section, and as an example, it is attempted to rate the SR of scientists' published work using a data sample from Greek academic registry in the scientific field "Molecular and cell biology." For this purpose, the SR-index was calculated for each member of the aforementioned sample, based on the number of scientists' published papers as appeared in Scopus until June 30, 2013, scientific relevance was evaluated using keywords matching as described in Table 1, the full data are presented in Table 2. As mentioned, greater values of the SR-index rate the most relevant scientists among the registry's members that are initially considered as having the "apparent" relevance. For example, the third member's SR-index value was given by the following formula, in which the publications' number was equal to 83, the most relevant publications' number was equal to 80, the wide relevant publications' number was equal to 81, the h^3-index was equal to16 and the h_c^3-index (for the previous 5 years) was equal to 8.

$$\text{SR - index}^3 = \frac{\frac{v_1^3}{N^3}}{\max\limits_{i=1,2,..,20}\left\{\frac{v_1^i}{N^i}\right\}} + \frac{\frac{v_2^3}{N^3}}{\max\limits_{i=1,2,..,20}\left\{\frac{v_2^i}{N^i}\right\}} + \frac{h^3}{\max\limits_{i=1,2,..,20}\left\{h^i\right\}} + \frac{h_c^3}{\max\limits_{i=1,2,..,20}\left\{h_c^i\right\}} = \frac{0.96}{0.98} + \frac{0.98}{0.99} + \frac{16}{36} + \frac{8}{10} = 3.21 \in [0,4]$$

The third member's SR-index value was relatively high; therefore, this scientist was classified in a high position in the list. In addition, the weighted SR-index could be calculated using weightings (subjective or objective). For example, the value of the weighted SR-index for the third member was given by the following formula, in which the suggested weights b_1,b_2,b_3, and b_4 take the values: 0.35, 0.25, 0.20, and 0.20, respectively.

$$\text{Weighted SR}^3\text{ - index} = 0.35\ldots\frac{\frac{v_1^3}{N^3}}{\max\limits_{i=1,2,..,k}\left\{\frac{v_1^i}{N^i}\right\}} + 0.25\cdot\frac{\frac{v_2^3}{N^3}}{\max\limits_{i=1,2,..,k}\left\{\frac{v_2^i}{N^i}\right\}} + 0.20\ldots\frac{h^3}{\max\limits_{i=1,2,..,k}\left\{h^i\right\}} + 0.20\cdot\frac{h_c^3}{\max\limits_{i=1,2,..,k}\left\{h_c^i\right\}} = 0.35\cdot\frac{0.96}{0.98} + 0.25\cdot\frac{0.98}{0.99} + 0.20\cdot\frac{16}{36} + 0.20\cdot\frac{8}{10} = 0.84 \in [0,1]$$

The weighted SR-index value was calculated for each member of the data sample and is presented in the last column of Table 2. Depending on the values assigned to the weights, evidently weighted SR-index value could be altered in order to highlight the SR based on another combination of weight values that better represent Academic Institution's or Organization needs.

Regarding the mathematical properties of the SR-index, it was an increasing function of all four components; with more influential components being the h_c^i-index/max (h_c-index), followed by the percentage of published work that fully match in the scientific field and the h-index/max (h-index).

Table 1: Parameters used to calculate the index of SR, in the example presented in this work

Total number of publications of the *i*-th individual	N^i	Relevant topics
Number of publications of the *i*-th individual based on relevant scientific work	v_1^i	The following modules of the archiving database Scopus, define the "relevance" in this work (keywords based on Scopus) Agricultural and Biological Sciences Biochemistry, Genetics and Molecular Biology Chemistry Environmental Science Health Professions Immunology and Microbiology Medicine Neuroscience Nursing Pharmacology, Toxicology and Pharmaceutics Psychology Social Sciences
Number of publications of the *i*-th individual based on wider relevant scientific work	v_2^i	The following modules of the archiving database Scopus, define the "wide relevance" in this work Agricultural and Biological Sciences Chemistry Environmental Science Biochemistry, Genetics and Molecular Biology, Neuroscience Medicine, Health Professions Nursing, Pharmacology Toxicology and Pharmaceutics Immunology and Microbiology Psychology, Social Sciences
h-index of the *i*-th individual	h^i	Number of papers with at least h citations in the literature (including self-citations)
Contemporary h-index of the *i*-th individual (5 years)	h_c^i	Number of papers with at least h citations in the literature (including self-citations) within the past 5 years

SR=Scientific relevance

DISCUSSION

In this paper, an index was proposed, which aimed to rate most relevant scientists among a set of individuals in a specific thematic module. This index may be a useful tool for the selection of members, e.g., of faculty election committees for vacant positions in academic institutions, or in various other research institutions, or for funding purposes, in order to facilitate the work in finding the most appropriate members for the evaluation committees. This index may be also useful for the selection of the

Table 2: Data used to calculate the SR-index in the example

Order	Number of publications (*N*)	Full relevance (v_1)	$\frac{v_1^i}{N}$	Wider relevance (v_2)	$\frac{v_2^i}{N}$	h-index	h_c-index	SR-index	Weighted SR-index
1	91	87	0.96	90	0.99	36	8	3.77	0.95
2	214	205	0.96	210	0.98	26	10	3.69	0.93
3	83	80	0.96	81	0.98	16	8	3.21	0.84
4	73	71	0.97	72	0.99	20	6	3.14	0.83
5	41	40	0.98	40	0.98	23	3	2.92	0.78
6	110	55	0.50	98	0.89	18	10	2.91	0.70
7	53	31	0.58	51	0.96	30	5	2.90	0.72
8	60	59	0.98	59	0.98	18	3	2.79	0.76
9	79	76	0.96	77	0.97	15	4	2.78	0.75
11	54	42	0.78	53	0.98	18	4	2.68	0.70
12	63	40	0.63	60	0.95	28	3	2.69	0.68
13	38	36	0.95	37	0.97	15	3	2.66	0.73
14	26	24	0.92	25	0.96	8	5	2.63	0.72
15	74	37	0.50	70	0.95	13	8	2.63	0.65
16	28	21	0.75	26	0.93	17	4	2.57	0.68
17	69	30	0.43	67	0.97	27	4	2.57	0.63
18	106	37	0.35	95	0.90	25	6	2.56	0.61
19	61	30	0.49	55	0.90	17	5	2.38	0.60
20	59	45	0.76	55	0.93	14	2	2.31	0.62

SR=Scientific relevance

best candidates to cover a faculty or research position. However, at this point, it should be strongly underlined that the quality of a candidate, and not only the quantity, even in terms of academic recognition, is what it matters. Although it has been suggested that academic recognition is a cornerstone of academic quality, this is not always true, and various other characteristics also exist to define quality. The advantage of the suggested index is that it is not only unbiased, since it uses pure science metrics, but it is also easily interpretable and applicable to a broad range of scientific fields. The use of weights gives another interpretation of SR values, based on what it is considered more important, that is, the number of relevant publications or the level of recognition or the contemporariness of the published work. However, weights' assignment should be done with great care.

The proposed index has many advantages as well as disadvantages. A major advantage is the fact that only easily achievable through various scientific search machines,[6,7] such as the Scopus and the ISI Thompson Reuters archiving databases, as well as objective criteria are used, whereas subjective assessment of the scientific work's relevance of the nominees is not involved, since the keywords used may be a best approach to define what is relevant and what is not. However, assessment of the relevance may lead to distortions of the final outcome, regarding the selection of the most appropriate members, due to methodological, conceptual, and sometimes emotional reasons. As regards the components of the SR-index, the first two components (i.e., $\frac{v_1^i}{N}$ and $\frac{v_2^i}{N}$) rate the publication impact of the individual, and are only subject to the subjective determination of relevance, as well as wider relevance. The next two components (i.e., h^i and h_c^i) are considered as the most reputable evaluation tools for the recognition of a scientist's work. The *h*-index attempts to measure both productivity and readability of a scientific work of either a scientist or a group of scientists or an academic department or university or even a country.[1] The *h*-index was suggested by Jorge E. Hirsch, physicist at University of California (San Diego, US), as a tool for the determination of the relative quality among theoretical physicists. Although the *h*-(Hirsch) index is generally accepted, there is a number of situations, in which may provide misleading information. Specifically, the index *h* does not take into account the number of authors who participate in a paper. For solving this problem, a corrected index has been proposed; the individual *h*-index that normalizes, regarding the number of co-authors, by dividing the *h*-index by the average number of authors, who participate in a scientist's work. However, the position and the "true" contribution of an individual in the writing group are disregarded and should be further considered. It is also a fact that the *h*-index occurs with large variability among different scientific disciplines and sectors. For that reason, this index is not recommended for readability comparisons among scientists of different disciplines, such as medicine and anthropology, as well as in sectors of the same discipline, such as basic research and public health. In the case of SR-index, this problem is, even partially, eliminated since the sample space is created by scientists with relevant scientific work (i.e., keywords matching).

Besides, *h*-index is calculated from the total number of publications of a scientist. This means that scientists having a short career, regardless of the importance of their scientific discoveries, are at an inherent disadvantage, because they have not been exposed to the scientific public sufficient time to receive citations of their work.[1] However, as it has been noted by Hirsch, the index proposed should be treated as a tool for the evaluation of researchers at the same stage of their careers, and not for historical comparisons. In this work, the h_c-index, proposed by Sidiropoulos *et al.*,[8] was used in order to smooth the previous problem. For this reason, the "age" of the article, that is, the time that the article has been exposed in the literature is taken into account. Finally, it is generally accepted that *h*-index can be manipulated by self-citations. There are several academic debates being done about whether self-citations of a scientific paper have to be taken into account or not. The self-citations' inclusion may lead to the "Matthew Effect"[9] where the "rich" in published work become "richer" in references and the "poor" ones become "poorer" in citations of their work. However, it may also be argued that large published work is a denotation of scientists' productivity, as well as their continuity in time.

Despite the majority of other scientometric indices presented in the literature, the aim of the proposed index was to become a tool, easy to use, and calculate that shows the SR in a scientific area via an objective, as well as a fair approach. The involvement of weights is an attempt to highlight the "absolute" relevance and timeless readability even more, but the inability to establish the weights objectively may convert the index to a discriminatory tool. Finally, the proposed index may hide some potential interpretation problems in practice, but only its implementation will highlight these methodological issues.

REFERENCES

1. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005;102:16569-72.
2. Wendl MC. H-index: However ranked, citations need context. Nature 2007;449:403.
3. Batista PD. Is it possible to compare researchers with different scientific interests? Scientometrics 2006;68:179-89.
4. von Bohlen Und Halbach O. How to judge a book by its cover? How useful are bibliometric indices for the evaluation of "scientific quality" or "scientific productivity"? Ann Anat 2011;193:191-6.
5. Amorós MB, Domingo-Ferrer J, Torra V. A bibliometric index based on the collaboration distance between cited and citing authors. J Informetr 2011;5:248-64.
6. Archiving and Data Base ISI Thompson Reuters. Available from: http://www.thomsonreuters.com/. [Last accessed on 2014 Jun 30].
7. Archiving and data base Scopus. Available from: http://www.scopus.com. [Last accessed on 2014 Jun 30].
8. Sidiropoulos A, Katsaros D, Manolopoulos Y. Generalized Hirsch h-index for disclosing latent facts in citation networks. Scientometrics 2007;72:253-80.
9. Merton RK. The Matthew effect in science: The reward and communication systems of science are considered. Science 1968;159:56-63.

How to cite this article: Panagiotakos DB, Bersimis FG. An index that rates relevance of scientific work in bio-sciences: The scientific relevance-index. J Sci Res 2014;3:15-21.

Source of Support: Nil, **Conflict of Interest:** None declared

APPENDIX

Keywords based on Scopus archiving database for various scientific fields.

AGRICULTURAL SCIENCES

The "Agricultural Sciences" category covers journals in general agriculture, agricultural chemistry, and agronomy:

- Agricultural engineering
- Agronomy
- Tillage research
- Agroforestry
- Horticulture
- Crop protection and science
- Agrochemistry
- Phytochemistry
- Agricultural biochemistry
- Food chemistry
- Cereal chemistry
- Carbohydrate and lipid research.
  - Food science and nutrition
  - Composition, additives, and contaminants
  - Microbiology and technology

- Engineering and processing
- Meat and dairy science
- Nutrition science
- Nutrition and metabolism
- Nutritional biochemistry.

BIOLOGY AND BIOCHEMISTRY

- Structure and chemistry of biological molecules
- Molecular, cellular, and clinical studies of the endocrine system
- Regulation of cell, organ, and system functions by hormones
- Experimental research in general biology and biological systems
- Regulation of biological functions at the whole organism level
- Exploitation of living organisms or their components
- Industrial microbiology
- Pollution remediation
- Industrial chemicals and enzymes
- Biosensors
- Bioelectronics
- Pesticide development
- Food, flavor, and fragrance industry applications
- Waste treatment.

CHEMISTRY

- Analytical chemistry
- Spectroscopy
- Instrumentation
- Inorganic and nuclear chemistry
- Organic chemistry
- Physical chemistry
- Polymer science
- Food chemistry
- Chemical methods and structures
- Natural and laboratory syntheses
- Isolation and analysis of clinically significant molecules
- Medicinal chemistry
- Chemical engineering.

CLINICAL MEDICINE

- Anesthesia
- Cardiovascular medicine
- Dentistry
- Dermatology
- General and internal medicine
- Endocrinology
- Environmental medicine
- Gastroenterology
- Gynecology
- Hepatology
- Hematology
- Nephrology
- Nuclear medicine
- Obstetrics
- Oncology
- Ophthalmology
- Otolaryngology
- Pediatrics
- Pharmacology
- Radiology
- Toxicology
- Respiratory medicine
- Rheumatology
- Surgery
- Urology.

IMMUNOLOGY

- Cellular and molecular studies in immunology
- Clinical research in immunopathology
- Infectious diseases
- Autoimmunity and allergy
- Host-pathogen interactions in infectious disease
- Experimental therapeutic applications of immunomodulating agents.

MICROBIOLOGY

- Biology and biochemistry of microorganisms (bacterial, viral, and parasitic)
- Medical implications of the subsets of these organisms known to cause diseases
- Biotechnology applications of microorganisms for basic science or clinical use.

MOLECULAR BIOLOGY AND GENETICS

- Biochemistry
- Molecular biology
- Biophysics
- Pharmacology
- Receptor biology
- Signal transduction
- Regulation of gene expression
- Developmental genetics and biology

- Morphogenesis
- Cell-environment interactions
- Molecular genetics
- Mechanisms of mutagenesis
- Structure, function, and regulation of genetic material
- Clinical genetics, patterns of inheritance, genetic causes, and screening and
- Treatment of diseases.

NEUROSCIENCE AND BEHAVIOR

- Cellular and molecular neuroscience
- Neuronal development
- Basic and clinical neurology
- Psychopharmacology biobehavioral psychology
- Molecular psychology
- Neuronal function underlying higher cognitive processes.

PHARMACOLOGY

- Pharmacology
- Pharmaceutics
- Cellular and molecular pharmacology
- Drug design and metabolism
- Mechanisms of drug action
- Drug delivery
- Natural products
- Xenobiotics
- Mechanisms of action for clinical therapeutics.
- Toxicology
    - Molecular and cellular effects of harmful substances
    - Environmental toxicology
    - Occupational exposure
    - Clinical toxicology.

PSYCHIATRY/PSYCHOLOGY

- Biological
- Clinical
- Developmental
- Educational
- Mathematical
- Organizational
- Personal
- Social
- Diagnosis and treatment.