

---

## Original articles

---

### The Subject Sameness Index: a new scientometric indicator

**Aram Tirgar**

*Social Medicine Department, Babol University of Medical Sciences, Babol, Iran; a\_tirgar@yahoo.com*

**Mousa Yaminfirooz**

*Department of General Education, Babol University of Medical Sciences, Babol, Iran; yaminfirooz@yahoo.com*

**Hemmat Gholinia Ahangar**

*Department of Statistics, Babol Noshirvani University of Technology, Babol, Iran; ahangar1353@yahoo.com*

#### Abstract: Background

A variety of scientometric indicators exists for quantitative and qualitative evaluation of scientific output. Each of these indicators has its own strengths and limitations, and the search for more correct ones is still ongoing. This study proposes and evaluates a new scientometric indicator, Subject Sameness Index (SSI).

#### Method

This cross-sectional study was conducted over a six-year period (2005-2010). We analysed all papers published by researchers of Babol University of Medical Science (BUMS), who were first authors with at least two papers listed in Web of Science (WoS) during the six-year period. A special tool was developed for data collection. The ratio of the same (repeated) or synonym keywords to all keywords listed by researchers in their papers was analysed for estimating 'subject sameness'.

#### Results

We retrieved 90 papers of BUMS researchers. These researchers were first authors, with 25 of them having at least 2 papers (67 in total). Of these, 11 papers had at least one repeated or synonym keyword. In these papers there were 300 keywords in total, with 21 repeated or synonymous keywords (7% of all the keywords were repeated). Regression analysis showed a reverse linear relationship between the number of published papers and the number of repeated or synonym keywords ( $R = -0.963$ ,  $P = 0.037$ ).

#### Conclusion

Quantity of publications, citations and their combinations are not always appropriate indicators of the quality of scientific work. Low level of the subject sameness may indicate a researcher's dispersed scientific activity. The new indicator, SSI, may add important information of interest to research administrators and science editors.

**Keywords** Scientometric indicators; research performance; Subject Sameness Index; science editing.

#### Introduction

Scientometrics provides quantitative characteristics of scientific productivity at the individual researcher and institutional levels. It also directs scientific policy-making,

facilitates science communication, and draws a global scientific map.<sup>1</sup> Scientometric performance of individuals, institutions and journals is assessed by citation analyses.<sup>2</sup>

Several research performance indicators have been proposed to comprehensively characterise the influence of individual authors. Of these, the *h* index and its variants are widely accepted and used to judge the quantity and influence of a set of publications.<sup>3,4</sup> Each scientometric indicator has its strengths and limitations, driving the search for new and more comprehensive metrics.<sup>5</sup> In fact, currently available indicators are inappropriate for measuring some important features of scientific output, and particularly the researcher's focus on a certain subject category.

The widely used *h* index and some other individual metrics are not corrected for 'subject sameness'. We propose to complement the *h* index with a Subject Sameness Index (SSI). This new indicator tracks subject sameness in papers by an individual author over a certain period of time. It indicates how focused are a researcher's scientific activities. The SSI is a ratio of papers with subject sameness to all papers published by the same author(s) over a certain period of time. Subject sameness is traced by subject-related keywords. The index indicates how focused a researcher's activities are: the higher the SSI, the greater focus on a certain subject.

The aim of this study was to assess the subject sameness in journal papers of researchers from Babol University of Medical Science (BUMS), Babol, Iran listed in the Web of Science (WoS) between 2005 and 2010. The SSI of the first authors was calculated using papers with the same subject keywords appearing in the papers.

#### Methods

We retrieved all WoS-listed papers of first authors affiliated to BUMS (2005-2010). First authors were chosen on the assumption that they had contributed most to the study design, execution and writing.

We collected data on the number of papers, keywords and repeated or synonym keywords. To trace the subject sameness, the first authors had to have at least two papers. The ratio of the same (repeated) or synonym subject keywords to all keywords used by a researcher was calculated.

Data were analysed by SPSS packages (version 16). We employed descriptive statistics methods to calculate frequencies, percentages and means. Regression analysis was also performed.  $P < 0.05$  was considered significant.

## Results

Overall, 127 WoS-indexed papers published during the six-year period were retrieved and analysed. The first authors affiliated to BUMS were in 90 papers. Of them, 25 had more than two papers (67 in total) (Table 1). Of the 67 papers, 30 had a repeated or synonym subject keyword. The total number of keywords in these papers was 300 (average 4.5 keywords per paper), with 21 repeated or synonym keywords. These same or synonym keywords were repeated 46 times in the papers (with a mean of 2.2 repeated keywords per paper). The ratio of repeated or synonym subject keywords to all keywords was 0.07. Hence, only 7% of all the keywords were repeated in the papers. Based on Table 1, a researcher's use of repeated or synonym subject keywords decreases with an increase of the number of his/her papers.

The SSI for a given period is calculated based on the following formula:  $SSI = SS_p / T_p$  ( $0 \leq SSI \leq 1$ ).  $SS_p$  is the number of a researcher's papers with subject sameness that includes all papers with at least one repeated or synonym keyword, and  $T_p$  is the total number of papers over the same period.

A value of one for the SSI indicates that the researcher has a strong focus on a specific subject over a certain period of his/her scientific career, whereas values close to null suggest weak, if any, focus on a subject category. As an example, if a researcher published 20 papers with a total of 80 keywords, and eight papers have at least one repeated keyword, his/her SSI will be 0.4.

The index has some inherent limitations. First of all, it is not useful for assessing the performance of researchers

**Table 1. Indicators relating to the subject sameness in papers authored by BUMS researchers (2005-2010)**

Repeated/synonym keywords per paper	Sum of repeated/synonym subject keywords	N of authors with repeated/synonym keywords	N of keywords per paper	N of keywords	N of authors with repeated/synonym keywords	N of authors	N of each author's papers ( $\geq 2$ )
0.36	20	9	4.4	105	4	12	2
0.37	24	11	4.3	132	6	10	3
0.12	2	1	4.7	38	1	2	4
0	0	0	5.0	25	0	1	5
0.07	46	21	4.5	300	11	25	Total

Regression analysis indicated a reverse linear relationship between the number of papers and repeated or synonym subject keywords ( $R = -0.963$ ,  $P = 0.037$ ) (Table 2).

**Table 2. Results of regression analysis of the number of papers and repeated or synonym subject keywords**

R	R2	Adjusted R2	Standard error of the estimate
0.963	0.927	0.891	0.427
Model goodness of fit: $F(1, 2) = 25.40$ ; $P = 0.037$			

## Discussion

Our study showed that the number of papers and repeated or synonym keywords are inversely associated. It is likely that experienced and actively publishing authors do not focus on the same subject or research field, and their scientific work is increasingly dedicated to dispersed subjects.

Given the debates on the importance of specialisation and focusing on a specific topic in scientific works,<sup>6</sup> we propose use of the SSI in combination with other research performance indicators, such as the *h* index and its variants. The use of this index is especially important in our times. The 'Publish or Perish' mantra is now the main driver of the growth of scholarly publications, which do not always address scientific needs, but rather may be just aimed at inflating the number of an author's papers. This trend also results in dispersing the efforts of researchers, increasing the number of publications by switching between different subject categories.

working on the interface of two or more specialities. The correctness of the index depends on which databases are used to retrieve papers. Appropriate retrieval of the papers and recording of their keywords may help avoid inaccuracies. In our study we focused on the WoS database. Future studies are warranted to find out whether one and several databases should be searched.

## Acknowledgements

The authors wish to thank Frank-Thorsten Krell, Werner Marx and Armen Yuri Gasparyan for critical reading of the paper and valuable comments.

## References

- Hood WW, Wilson CS. The literature of bibliometrics, scientometrics, and informatics. *Scientometrics* 2001;52(2):291-314.
- Bornmann L, Marx W, Gasparyan AY, Kitas GD. Diversity, value and limitations of the journal impact factor and alternative metrics. *Rheumatology International* 2012;32(7):1861-1867. doi: 10.1007/s00296-011-2276-1
- Bornmann L, Marx W. The h index as a research performance indicator. *European Science Editing* 2011; 37(3):77-80.
- Krell F-T. The Journal Impact Factor as a performance indicator. *European Science Editing* 2012;38(1):3-5.
- Abbas AM. Bounds and inequalities relating h-index, g-index, e-index and generalized impact factor: an improvement over existing models. *PLoS One* 2012; 7(4):e33699. doi: 10.1371/journal.pone.0033699
- Schoepflin U, Glänzel W. Two decades of "scientometrics": an interdisciplinary field represented by its leading journals. *Scientometrics* 2001;50(2):301-312.