Journal of Informetrics 8 (2014) 503–507

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

On standardization of the Activity Index

Janez Stare, Nataša Kejžar*

Institute for Biostatistics and Medical Informatics, Faculty of Medicine, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 17 February 2014
Received in revised form 28 March 2014
Accepted 2 April 2014
Available online 4 May 2014

Keywords:
Relative Specialization Index
Activity Index

ABSTRACT

Relative Specialization Index (*RSI*) was introduced as a simple transformation of the Activity Index (*AI*), the aim of this transformation being standardization of *AI*, and therefore more straightforward interpretation. *RSI* is believed to have values between −1 and 1, with −1 meaning no activity of the country (institution) in a certain scientific field, and 1 meaning that the country is only active in the given field. While it is obvious from the definition of *RSI* that it can never be 1, it is less obvious, and essentially unknown, that its upper limit can be quite far from 1, depending on the scientific field. This is a consequence of the fact that *AI* has different upper limits for different scientific fields. This means that comparisons of *RSI*s, or *AI*s, across fields can be misleading. We therefore believe that *RSI* should not be used at all. We also show how an appropriate standardization of *AI* can be achieved.

1. Activity Index

The Activity Index (*AI*) was introduced by Frame (1977) and defined as

$$AI(C,F)=\frac{\text{the country's } C \text{ share in the world's publication output in the field } F}{\text{the country's } C \text{ share in the world's publication output in all science fields}} \tag{1}$$

or, equivalently,

$$AI(C,F)=\frac{\text{share of a country's publications in a given field in all publications of the country}}{\text{share of all publications in a given field in the world's total publications}}$$

The index was obviously introduced with the aim to identify scientific areas in which a country is more/less productive than its overall production. The neutral value is 1, values greater/less than 1 indicating a higher/lower country's production in a given field as compared to its overall production. If we are only interested in identifying areas in which a country is more/less productive, then looking at *AI* is (more or less) ok, but if we are paying attention to the actual values of *AI*, then one should be aware that the maximum value of *AI* depends on the share of the scientific field in the world's production. We explain this in detail in Section 3.

* Corresponding author. Tel.: +386 1 543 77 85; fax: +386 1 543 77 71.
E-mail addresses: janez.stare@mf.uni-lj.si (J. Stare), natasa.kejzar@mf.uni-lj.si (N. Kejžar).

http://dx.doi.org/10.1016/j.joi.2014.04.004
1751-1577/

Table 1
Number and proportion of publications per scientific field in the period 2005–2009.

	Frequency	Proportion (in %)
Agricultural Sciences	263,427	4.97
Engineering and Technology	1,136,921	21.46
Humanities	123,528	2.33
Medical and Health Sciences	1,722,074	32.51
Natural Sciences	2,755,311	52.01
Social Sciences	407,290	7.69

2. Relative Specialization Index

Glänzel (2000) recalls the definition and interpretation of the Relative Specialization Index and gives (REIST-2, 1997) as the original reference. The definition of *RSI* is

$$RSI = \frac{AI - 1}{AI + 1} \tag{2}$$

and the supposed interpretation the following:

a) $RSI = -1$; when a country C is not active in the field F (has no publications there)
b) $RSI = 0$; when a country C's share in the field F is the same as its overall share in the world
c) $RSI = 1$ when the country is active in no other than the given field.

Obviously, *RSI* brings no new information, it simply aims to somehow standardize *AI*. And it is of course also obvious from the definition that *RSI* can never be 1. One could accept the interpretation c) if values of *RSI* would be very close to 1. Unfortunately, this is far from being true.

3. Maximum values of *AI* and *RSI*

Let

n_{ij} denote the number of publications in the field i in the country j
$n_{i.}$ denote $\sum_j n_{ij}$ (number of all publications in the world in the field i)
$n_{.j}$ denote $\sum_i n_{ij}$ (number of all publications in the country j)
$n_{..}$ denote $\sum_i \sum_j n_{ij}$ (the number of all publications in the world)

Then Eq. (1) can be rewritten as

$$AI = \frac{n_{ij}/n_{i.}}{n_{.j}/n_{..}} = \frac{n_{ij}\, n_{..}}{n_{i.}\, n_{.j}} \tag{3}$$

and Eq. (2) as

$$RSI = \frac{(n_{ij}n_{..}/n_{i.}n_{.j}) - 1}{(n_{ij}n_{..}/n_{i.}n_{.j}) + 1} = \frac{n_{ij}n_{..} - n_{i.}n_{.j}}{n_{ij}n_{..} + n_{i.}n_{.j}}.$$

When a country is active in only one field of research (note here that country can be substituted by an institution or similar), we have $n_{ij} = n_{.j}$ and therefore

$$AI = \frac{n_{..}}{n_{i.}} \tag{4}$$

or

$$RSI = \frac{n_{..} - n_{i.}}{n_{..} + n_{i.}}. \tag{5}$$

If the field of research is small, then *AI* will be very high and *RSI* close to 1, but if that is not the case, and it often is not, the differences can be substantial. We illustrate this in Tables 1 and 2. Table 1 gives the number of publications per scientific field in the years 2005–2009 (InCites™, 2011), which are the basis of calculation of maximum values of *AI* and *RSI* in Table 2. Note that the frequencies in the table do not sum up to 5,297,653, the actual number of papers published in the world in the given period, and that the percentages are calculated using this total value, so they do not sum up to 100. The differences in maximum values of *AI* and *RSI* are striking.

Table 2
Maximum values of *AI* (Eq. (4)) and *RSI* (Eq. (5)) per scientific field.

	max(*AI*)	max(*RSI*)
Agricultural Sciences	20.11	0.91
Engineering and Technology	4.66	0.65
Humanities	42.89	0.95
Medical and Health Sciences	3.08	0.51
Natural Sciences	1.92	0.32
Social Sciences	13.01	0.86

4. Standardization of *AI*

Standardized values convey information more rapidly and are, usually, directly comparable. Further reason to standardize *AI* are its different maximum values for different disciplines. And while *RSI* was an attempt to do so, it unfortunately also inherits this unwelcome property from *AI*.

An obvious possibility to standardize *AI* is to calculate $AI/\max(AI)$. This will be 0 if *AI* is 0, and 1 if *AI* is at its maximum. However, the neutral value, 1 for *AI*, will be $1/\max(AI)$ and will therefore depend on the maximum value of *AI*. This is certainly not desirable, so one needs to transform $AI/\max(AI)$ in such a way that a chosen value, say 1/2, will always be the neutral value, while 0 and 1 would still be as before.

Denoting $u = 1/\max(AI)$ we need to transform values in $[0, u]$ into $[0, 1/2]$ and values in $(u, 1]$ into $(1/2, 1]$. This is easily achieved using the following linear functions

$$y = \begin{cases} \dfrac{x}{2u} & x \in [0, u] \\ \dfrac{x + 1 - 2u}{2 - 2u} & x \in (u, 1] \end{cases}$$

Remembering what u stands for, and that $x = AI/\max(AI)$ we get for the Standardized *AI*

$$SAI = \begin{cases} \dfrac{AI}{2} & AI \le 1 \\ 1 - \dfrac{1}{2} \cdot \dfrac{\max(AI) - AI}{\max(AI) - 1} & AI > 1 \end{cases}$$

This transformation guarantees that *SAI* is 0 when there is no activity in the given field, 1/2 when a country's production in a given field is the same as its overall production, and 1 when a country is only active in the given field. If one would prefer values −1, 0, and 1 to express this, the change of our transformation is of course trivial. The transformation is illustrated in Fig. 1 for the case where $\max(AI) = 10$.

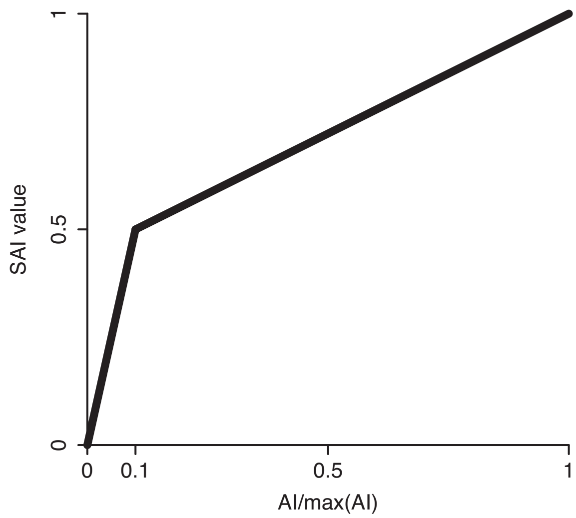

Fig. 1. A way of standardizing *AI*.

Table 3
Example: indicators in Geosciences for Nordic countries.

	Geosciences	
	RSI	*SAI*
Denmark	0.01	0.50
Finland	−0.08	0.43
Iceland	0.67	0.56
Norway	0.40	0.52
Sweden	−0.09	0.42

Table 4
Example: indicators in Geosciences and Agriculture.

	Geosciences		Agriculture	
	RSI	*SAI*	*RSI*	*SAI*
Denmark			0.19	0.52
Norway	0.40	0.52		

One might be tempted to use a nonlinear transformation, one that would be smooth (i.e. derivative existed) at the point $(u, 1/2)$, like

$$y = \frac{u-1/2}{u(1-u)}x^2 + \frac{1/2-u^2}{u(1-u)}x.$$

But, this is not necessarily a monotonically increasing function, and can, for small u, take values greater than 1 somewhere on the interval $(u, 1)$ and then decline to 1.

Linear functions could of course be replaced by nonlinear, monotonically increasing, functions on $[0, u]$ and $(u, 1]$, but it is hard to see a reason for such complications at this point.

SAI is quite different from RSI. The problem with RSI is that the same values for two different disciplines are not really the same, or that one value being less than the other does not necessarily mean it is so. The order might actually be reversed. For example, a value 19 with a maximum 20, is much more than a value 25 with a maximum 40. The two slopes in Fig. 1 are different (we would only have one slope if max(*AI*) was equal to 2), and slopes across disciplines will also be different, but that only means that the speeds of getting to a certain value are different, while the final values are directly comparable and have the same meaning for every situation. The two speeds are of course inherent to the property of the AI. Its minimum is 0, its neutral value is 1, and the maximum depends on the discipline, but can be very high. So, small improvements in AI are usually needed to get to 1, but big to get closer to maximum. This is reflected in the slopes of the two lines. The values of SAI less than 1/2 are simply AI/2, so directly comparable as AIs are directly comparable for values of AI less than 1 (they simply measure how far we are on the path from zero productivity to the average productivity). For values of $AI > 1$, SAI measures the distance covered from 1 to max(*AI*), and these values are of course directly comparable across disciplines.

5. Example

An example from the report "Bibliometric Performance Indicators for the Nordic Countries" (Schneider et al., 2010) shows, how the indicators change if *SAI* is used instead of *RSI*. In Table 3 *RSI* and *SAI* are given for Geosciences. Based on *RSI*, the authors conclude that Iceland is very different, followed by Norway, while the other three countries are similar among them and far from these two. Recalculation with *SAI* shows the same order (of course), but much less drama.

Such comparisons can become even more deceiving when one compares different fields between countries. Using the same source, we see in Table 4, that Norway has an *RSI* of 0.40 for Geosciences and Denmark 0.19 for Agriculture, which would make one conclude that Norway is somehow more concentrated on Geosciences than Denmark is on Agriculture. When we calculate *SAI*, we get exactly the same value, thus telling a different story.

Such differences in interpretation, we expect, would be even more pronounced, if one compared countries which are much more different than the five Nordic countries.

6. Conclusion

The differences in maximum values of *AI* and *RSI* between scientific fields are so big, that any conclusions based on analyzes of these indices seem questionable. Maybe some would not essentially change, but, as our example shows, some might. We are not judging the importance of the case of *AI* and *RSI*, but if it reflects a certain ease with which new indices are introduced, then a warning is in order. *RSI* is not the only transformation of the type $(x-1)/(x+1)$ used in bibliometry, *NMCR* is another (that we know of) (see Hu & Rousseau, 2009). To repeat, such 'transformations' bring no new information, and should be avoided.

References

Frame, J. D. (1977). Mainstream research in Latin America and the Caribbean. *Interciencia*, *2*, 143–148.

Glänzel, W. (2000). Science in Scandinavia: A bibliometric approach. *Scientometr.*, *48*(2), 121–150 (Correction: Scientometr., 49 (2) (2000) 357).

InCites™. (2011). Essential science indicators, Thomson Reuters. Web based platform.

Hu, X., & Rousseau, R. (2009). A comparative study of the difference in research performance in biomedical fields among selected Western and Asian countries. *Scientometrics*, *81*(2), 475–491.

REIST-2. (1997). *The European report on science and technology indicators 1997. EUR 17639*. Brussels: European Commission.

Schneider, J. W., Aksnes, D. W., Faurbæk, L., Finnbjörnsson, þ, Fröberg, J., Gunnarsson, M., Karlsson, S., Kronman, U., Lehvo, A., Nuutinen, A., Sivertsen, G., & Sveinsdóttir Morthens, S. G. (2010). *Bibliometric research performance indicators for the Nordic Countries. A publication from the NORIA-net "The use of bibliometrics in research policy and evaluation activities". NORIA-net Report 3*. Oslo: NordForsk.