JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY (ISSN: 1532-2882) 64: (5) pp. 1084-1085. (2013)

DOI: 10.1002/asi.22833

11, 09, 2012, 2. corr. JASIST

Letter to the Editor

Quantity and Impact through a Single Indicator

Dear Sir:

There are three main aspects of the assessment of publication activity of scientists or teams in evaluative scientometrics: measuring the total quantity and impact of information published and, measuring the impact of the most influential part (i.e. core publications or elite set) of the publications. In natural science the quantity of information can be approximated by the number of journal publications and the total impact can be measured by the total number of citations, whereas the most frequently cited papers may be regarded as most influential.

Characterizing quantity together with impact by a single index, it may be preferably made through a composite index (Vinkler, 2006). Van Raan (2008, p. 474) e. g. suggested the "brute force impact indicator" (*BFII*) consisting of a quality (impact) and a quantity part (Eq. 1).

$$BFII = \frac{CPP}{FCS_m} P \qquad (1)$$

where *CPP* is the citation rate i.e. *C/P* , where *C* is the number of citations obtained to the publications (*P*) assessed, and FCS_m is the average citation rate of publications in the field worldwide.

Earlier, I suggested Eq. 2 for calculating the Relative Publication Potential (*RPP*) of a team (RPP_t) for representing quality and quantity together (Vinkler, 2000).

$$RPP_t = \frac{\sum_{p=1}^{P_t} c_p}{\sum_{p=1}^{P_t} GF_p} P_t = \frac{C_t}{C_p} P_t \qquad (2)$$

where c_p is the number of citations obtained to the p-th publication of the team, P_t is the total number of publications, and GF_p is the Garfield (impact) factor of the journal where the p-th publication was published. Accordingly, C_t is the total number of citations received by the team and C_p is the total number of citations obtained by P_t number of articles in the publishing journals. Note the sum of the GF_p values (C_p) represents the standard applied (i.e. "required" number of citations). (For the sake of simplicity, the time factor is neglected here.) If the C_t/C_p ratio is equal to unity, the impact of the publications of the team will correspond to that of the publishing journals. According to Eq. 2 the RPP_t index expresses the impact of the publications of the team assessed by the relative impact index (C_t/C_p), whereas the quantity of the information published is represented by the number of publications (P_t). In contrast to the above, Eq. 1 referring to the "brute force impact indicator" of a team *($BFII_t$)* (Van Raan, 2008) may be transformed to Eq. 3.

$$BFII_t = \frac{CPP_t}{FCS_m} P_t = \frac{\frac{C_t}{P_t}}{\frac{C_F}{P_F}} P_t = \frac{C_t}{C_F} P_F \qquad (3)$$

where the CPP_t index is the mean citation rate of the publications (P_t) assessed, whereas FCS_m is the mean citation rate of the publications in the field where the team is active. Accordingly, the former index equals to C_t/P_t where C_t is the number of citations to P_t papers published by the team, whereas FCS_m equals to C_F/P_F where C_F is the total number of citations obtained by all papers in the field used as a standard, and P_F is the number of publications in the respective field. According to Van Raan (2008) the "size dependent *BFII* is calculated by the multiplication of *P* with the university's field-normalized average impact, *P(CPP/FCS$_m$)*." Eq. 3 clearly indicates however, that *BFII* should be regarded as the product of the contribution index, C_t/C_F with a growth index, P_F, and should not be assumed as the product of a relative impact index with a growth index. As to the quantitative part (P_F) of *BFII* is concerned, it refers to the production of the corresponding *field* and not to the production of the team. Consequently, *BFII* depends linearly on the *size of the field*. Nevertheless, the contribution index, C_t/C_F may be an important indicator reflecting the share of the team studied within the total impact of the field. Note the *100(C_t/C_F)* index of the team related to its share in publications, *100(P_t/P_F)* in the field yields the *Specific Impact Contribution (SIC)* index, *C%/P%* which is comparable across fields (Vinkler, 2009a). According to the definition of BFII Eq. 3, a team working in a greater field could attain a higher indicator than a team active in a smaller field. Team *X* e.g., working in field *A* may obtain 10% of total citations in the field (C_t/C_F=0.1). The total number of papers in the field is, say 1000, accordingly: *BFII*(*A*)=100. Another team receiving also 10% of total citations but, active in field *B* with significantly more papers (P_F=10000), may attain a higher index: *BFII*(*B*)=1000. According to the above, the team active in the greater field would enjoy an undeserved advantage.

For publications of universities working in several fields, an average FCS_m index can be calculated (Van Raan, 2008). Accordingly, for the publications, $P(A)$ and $P(B)$ of a university active in two different fields (A,B), the following average standard, $FCS_m(A,B)$ can be calculated (Eq. 4).

$$FCS_m(A,B) = 1/(P(A)+P(B))\,(FCS_m(A)\cdot P(A)) + FCS_m(B)\cdot P(B)) \quad (4)$$

The number of publications in each field is used as a weight. Let us suppose that University I published 40 papers in field A (FCS_m=5.00), whereas 10 papers in field B (FCS_m=1.00). Accordingly, the compound FCS_m index is calculated as: 1/50 ((40 · 5.00)+(10 · 1.00)) = 4.20. In contrast, University II published 10 papers in field A (FCS_m=5.00) and 40 papers in field B (FCS_m=1.00), accordingly the compound FCS_m=1/50((10 · 5.00) + (40 · 1.00)) = 1.80. The difference between the two indices clearly shows that the compound reference standard is highly influenced by the share of activities in different fields, which is, however mostly beyond the control of the universities or teams studied. And, it is obvious: there are both excellent and mediocre scientists in each field independent of the size of the field.

Several methods and indicators suggested recently may involve both qualitative and quantitative aspects. Especially those indicators seem to be important which are derived from the "elite set" (Vinkler, 2009b) of publications. As the elite set, the most influential part of a publication set may be regarded. In the first approximation the *h*-index (Hirsch, 2005), i.e. the number of papers in the *h*-core may be assumed as a quantity and impact indicator of the elite set. According to Iglesias & Pecharroman (2007) the h-index may be related also to the number of journal papers (P) (Eq. 5).

$$h\text{-}index=(P/4)^{1/3}\,(C/P)^{2/3} \quad (5)$$

where C is the number of citations. A similar relation was found also by Schubert & Glänzel (2007). It follows: the *h*-index may be used for characterizing the publication production both qualitatively and quantitatively. For the inconsistencies of the *h*-index, see however Waltham & Van Eck (2012).

The π-index (Vinkler, 2009b) refers also to the elite set. It can be calculated as 1% of the citations to the top square root of the papers in the total set. (The papers are ranked by citation frequency.)

The simplest method for obtaining the number of publications in the possible elite set is to calculate 0.1%, 1.0%, or 10.0% of the total. The total number of citations to these papers may be regarded as a quantity and impact index.

For characterizing impact and quantity together through a single index, the Citation Distribution Score (*CDS*) was suggested recently (Vinkler, 2011), which takes into account the number of journal papers (P) according to their citation rate. The first citation category of a set of publications refers to the papers with zero *or* a single citation, the second refers to the articles with 2-4 citations. From the third category on, the lower limit of the categories is calculated as 2^n+1, where n=2, 3, 4, etc., and the upper limit as 2^n with n=3, 4, etc., respectively. Accordingly, the lower and upper limit of citations of

the categories is the following: 0-1; 2-4; 5-8; 9-16… 4097-8192; >8192. The number of publications in the individual categories may be multiplied with optional *weights* stressing the importance of highly cited articles. The sum of the weighted number of publications over the categories yields the *CDS* indicator of the publication set assessed. Surveying the recent literature however, I came to the conclusion that a substantial theoretical work and several case studies are needed yet to arrive at a widely acceptable solution concerning the characterization of the eminence of publications of scientists and teams both qualitatively and quantitatively by a single indicator.

References

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569-16572.

Iglesias, J. E. & Pecharroman, C. (2007). Scaling the h-index for different scientific ISI fields. Scientometrics, 73(3), 303-320.

Schubert, A. & Glänzel, W. (2007). A systematic analysis of Hirsch-type indices for journals. Journal of Informetrics, 1(3), 179-184.

Van Raan, A. F. J. (2008). Bibliometric statistical properties of the 100 largest European research universities: Prevalent scaling rules in the science system. Journal of the American Society for Information Science and Technology, 59(3), 461-475.

Vinkler, P. (2000). Evaluation of the publication activity of research teams by means of scientometric indicators. Current Science, 79, 602-612.

Vinkler, P. (2006). Composite scientometric indicators for evaluating publications of research institutes. Scientometrics, 68, 629-642.

Vinkler, P. (2009a). Introducing the Current Contribution Index for characterizing the recent, relevant impact of journals. Scientometrics, 79, 409-420.

Vinkler, P. (2009b). The pi-index: A new indicator to characterize the impact of journals, Scientometrics, 82, 461-475.

Vinkler, P. (2011). Application of the distribution of citations among publications in scientometric evaluations. Journal of the American Society for Information Science and Technology, 62, 1963-1978.

Waltham, L. & Van-Eck, N. J. (2012). The inconsistency of the h-index. Journal of the American Society for Information Science and Technology, 63 (2), 406-415.

Peter Vinkler

Research Centre for Natural Sciences, Hungarian Academy of Sciences, 1025 Budapest, Pusztaszeri út 59-67, Hungary. E-mail: vinkler.peter@ttk.mta.hu