Journal of Informetrics 4 (2010) 407–414

Contents lists available at ScienceDirect

Journal of Informetrics

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

The *h* index research output measurement: Two approaches to enhance its accuracy

Lutz Bornmann[a,*], Rüdiger Mutz[a], Hans-Dieter Daniel[a,b]

[a] Professorship for Social Psychology and Research on Higher Education, ETH Zurich, Zähringerstr. 24, 8092 Zurich, Switzerland
[b] Evaluation Office, University of Zurich, Mühlegasse 21, 8001 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 4 November 2009
Received in revised form 5 March 2010
Accepted 15 March 2010

Keywords:
h index
Research output
Accuracy
h^2 lower
h^2 center
h^2 upper
Segmented regression model
sRM value

ABSTRACT

The *h* index is a widely used indicator to quantify an individual's scientific research output. But it has been criticized for its insufficient accuracy—the ability to discriminate reliably between meaningful amounts of research output. As a single measure it cannot capture the complete information on the citation distribution over a scientist's publication list. An extensive data set with bibliometric data on scientists working in the field of molecular biology is taken as an example to introduce two approaches providing additional information to the *h* index: (1) h^2 lower, h^2 center, and h^2 upper are proposed, which allow quantification of three areas within a scientist's citation distribution: the low impact area (h^2 lower), the area captured by the *h* index (h^2 center), and the area of publications with the highest visibility (h^2 upper). (2) Given the existence of different areas in the citation distribution, the segmented regression model (sRM) is proposed as a method to statistically estimate the number of papers in a scientist's publication list with the highest visibility. However, such sRM values should be compared across individuals with great care.
© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. The h index as an indicator to quantify an individual's scientific research output

Physicist Hirsch (2005) introduced an indicator for quantifying the research output of scientists that has ever since been discussed and studied theoretically and empirically in a number of disciplines (Bornmann & Daniel, 2007; Bornmann & Daniel, 2009). Hirsch's *h* index was proposed as a better alternative to other bibliometric indicators (such as number of publications, average number of citations, and sum of all citations) (Hirsch, 2007). It is based on a scientist's lifetime citedness (Seglen, 1992), which incorporates productivity as well as citation impact: "A scientist has index *h* if *h* of his or her N_p papers have at least *h* citations each and the other $(N_p - h)$ papers have $\leq h$ citations each" (Hirsch, 2005, p. 16569). All works by a scientist having at least *h* citations are called the 'Hirsch core' (Rousseau, 2006); these are the publications within a scientist's publication list that have the greatest visibility (or greatest impact) (Burrell, 2007).

Today, the *h* index is a widely used indicator of research output; the *h* index is computed automatically in the Web of Science (WoS, provided by Thomson Reuters, Philadelphia, PA, USA) and other literature databases (van Eck & Waltman, 2008). A number of studies showed that a scientist's *h* index corresponds to peer judgments (Bornmann & Daniel, 2007; Bornmann & Daniel, 2009) and thus has convergent validity. The most-often expressed criticism of the *h* index (and which led to the development of numerous variants of the *h* index) is applicable not only to the *h* index itself but also to bibliometric

* Corresponding author. Tel.: +41 044 632 48 25; fax: +41 044 632 12 83.
E-mail address: bornmann@gess.ethz.ch (L. Bornmann).

1751-1577/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2010.03.005

indicators generally (the problems of field dependency, self-citations, and multi-authorship). Criticism specific to the *h* index is much more rarely to be found in the literature. But as it concerns its insufficient accuracy (Kim, 2009; Lehmann, Jackson, & Lautrup, 2006), or the ability to discriminate reliably between meaningful amounts of research output, it is fundamental and noteworthy: for one, it is said that a single *h* index value does not yield a reliable picture of the research output of a scientist; additional data are necessary. For another, a scientist's *h* index value is said to differ more or less from the 'true' value of the number of his/her core publications (that is, those publications with the greatest visibility, see above).

1.2. Criticism of the accuracy of the h index

The *h* index value results from the distribution of citations (Adler, Ewing, Taylor, & Hall, 2009; Joint Committee on Quantitative Assessment of Research, 2008) over a scientist's rank-ordered publications. This distribution contains the complete information on the productivity and the citation impact of a scientist. However, the *h* index captures only "a small amount of information about the distribution of a scientist's citations" (Joint Committee on Quantitative Assessment of Research, 2008, p. 2) and discards "almost all the detail of citation records" (Joint Committee on Quantitative Assessment of Research, 2008, p. 14), for which reason it is "not applicable to the general body of researchers" (Evidence Ltd., 2007, p. 16). The *h* index acts as a "double filter," as (1) once a paper belongs to the *h* core it is not used any more in the calculation of the *h* index value and (2) the *h* index does not take into account the citation counts of papers with fewer than *h* citations (Franceschini & Maisano, 2010).

According to Zhang (2009) "a group of scientists having an identical *h* index is said to be within an isohindex group." However, scientists within one group can be of very different research output types (García-Pérez, 2009) "described in terms of the production of scientific papers and their quality (as assessed by citations)" (Cole & Cole, 1967, p. 382). "Think of two scientists, each with 10 papers with 10 citations, but one with an additional 90 papers with 9 citations each; or suppose one has exactly 10 papers of 10 citations and the other exactly 10 papers of 100 each. Would anyone think them equivalent?" (Joint Committee on Quantitative Assessment of Research, 2008, p. 13). In Section 3 ("calculation of h^2 lower, h^2 center, and h^2 upper"), we will present h^2 lower, h^2 center, and h^2 upper, which shed light on the amount of information about the distribution of a scientist's citations not captured by the *h* index. Although there are other measures that can describe a citation distribution, such as the Lorenz curve or the Herfindahl Index (see here Bornmann, Mutz, Neuhaus, & Daniel, 2008), they do not compensate for the specific limitations of the *h* index to the same extent as our approaches do.

The synthesis of publication and citation numbers in one *h* index value is seen as making sense "because the two... parameters have roughly the same order of magnitude" (Franceschini & Maisano, 2010, p. 495). The problem with the way in which the *h* index combines publication and citation numbers has been described as follows: "The problem is that Hirsch assumes an equality between incommensurable quantities. An author's papers are listed in order of decreasing citations with paper i having $C(i)$ citations. Hirsch's index is determined by the equality, $h = C(h)$, which posits an equality between two quantities with no evident logical connection" (Lehmann, Jackson, & Lautrup, 2008, p. 377). The equality $h = C(h)$ is viewed as an oversimplification (Leydesdorff, 2009) and as arbitrary: "Hirsch could equally well have defined the *h* index as follows: a scientist has index h if h of his n papers have at least $2h$ citations each and the other $n - $ h papers have fewer than $2(h+1)$ citations each. Or he could have used the following definition: a scientist has index h if h of his n papers have at least $h/2$ citations each and the other $n - h$ papers have fewer than $(h+1)/2$ citations each. A priori, there is no good reason why the original definition of the *h* index would be better than these two alternative definitions and other similar definitions. Hence, the *h* index can be seen as a special case of a more general research performance measure. The *h* index is obtained from this more general measure by setting a parameter to an arbitrarily chosen value" (van Eck & Waltman, 2008, pp. 263–264). In Section 3 ("calculation of sRM values"), we will present a simple approach whereby this parameter is not set arbitrarily but is instead estimated based on the distribution of a scientist's publication and citation data. This estimated value, which we call the sRM value, sheds light on the 'true' value of the number of a scientist's publications with the greatest visibility (that is, the scientist's 'true core'), and the difference between the *h* index and this estimated, 'overarching' research output measure can be determined.

Both approaches presented in this study (calculation of h^2 lower, h^2 center, and h^2 upper as well as the sRM value) are related to suggestions for *h* index variants published recently against the backdrop of the points of criticism mentioned above. These variants aim to represent the citation distribution better than the *h* index does: the tapered *h* index (Anderson, Hankin, & Killworth, 2008), the *e* index (Zhang, 2009), the R_{m-cv} index (Panaretos & Malesios, 2009), and the *w* index (Wohlin, 2009). However, with all of these variants, a scientist's research performance is expressed in one single number. According to Gągolewski and Grzegorzewski (2009) a "drawback of all one-dimensional indices is that they do not tell anything about the shape of the original citation function. They cannot answer such natural questions like: has the citation function a long tail? Is it flat or peaked?" (p. 626). Because the *h* index has by now become established as a single measure of research performance, we present in this study two approaches providing additional information to the *h* index. These approaches can increase the accuracy of the *h* index research output measurement by giving information about a scientist's citation distribution.

2. Methods

We investigate the *h* index and present two approaches that increase the accuracy of the *h* index research output measurement using a data set containing the publications of applicants to the Young Investigator Programme (Bornmann, Wallon, &

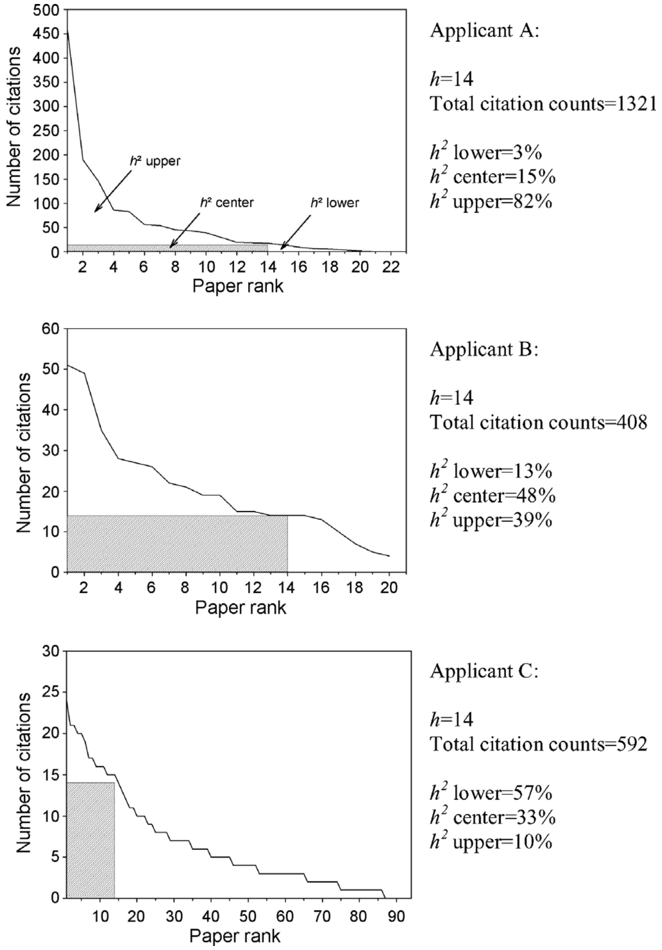

Fig. 1. Citation distributions of three applicants (A, B, and C) belonging to one isohindex group (having the same *h* index).

Ledin, 2008a, 2008b; Ledin, Bornmann, Gannon, & Wallon, 2007). This program of the European Molecular Biology Organization (EMBO) in Heidelberg, Germany, has been supporting outstanding young group leaders in the life sciences in Europe since 2000 (see http://www.embo.org/yip/index.html; accessed: July 10, 2009). The program targets researchers who have been leading their first independent laboratory in a European Molecular Biology Conference (EMBC) Member State (see http://embc.embo.org/, accessed: July 10, 2009) normally not more than four years before applying to the program. This study examined publication and citation data for 297 applicants to the EMBO Young Investigator Programme from the years 2001 and 2002. These applicants published a total of 6087 papers (articles, letters, notes, and reviews) prior to submitting their applications (publication window: from 1984 to the application year in 2001 or 2002). These papers received an average of 46.56 citations (median = 23) (citation window: from publication year to the beginning of 2007). The applicants' *h* index values were on average 13.13 (arithmetic average, median = 13) and range from 1 (minimum) to 34 (maximum). The citation analyses for the present study were conducted based on WoS.

The statistical analyses were calculated using SAS 9.2 (SAS Institute Inc., 2009).

3. Results

3.1. Calculation of h^2 lower, h^2 center, and h^2 upper

Fig. 1 shows for three applicants the distribution of citations over each applicant's publication set. The area of a distribution tallies with the applicant's total citation counts. As a rule, the citation distribution for a larger number of publications is right-skewed, distributed according to a power law (Egghe, 2005). In a publication set there are mostly a few highly cited papers and many hardly cited papers. As the distribution for scientist A shows, the *h* index captures only a small part of the publication and citation data, if the distribution is right-skewed. The *h* index refers to the area $h \times h$ and does not take into consideration the areas starting at *h* citations (we will call this h^2 upper) or starting at *h* papers (we will call this h^2 lower). For this reason, different scientists for whom the citation frequencies are distributed very differently right-skewed to their publications can be within the same isohindex group (see scientists A, B, and C in Fig. 1). The area proportions h^2 lower, h^2 center, and h^2 upper are defined as follows:

$$h^2 \text{ upper} = \frac{\sum_{j=1}^{h}(cit_j - h)}{\sum_{j=1}^{n}(cit_j)} \times 100 \tag{1}$$

$$h^2 \text{ center} = \frac{h \times h}{\sum_{j=1}^{n}(cit_j)} \times 100 \tag{2}$$

$$h^2 \text{ lower} = \frac{\sum_{j=h+1}^{n} cit_j}{\sum_{j=1}^{n} cit_j} \times 100 \tag{3}$$

As the equations show, the research output of two or more scientists could be compared by examining h^2 lower, h^2 center, and h^2 upper in percent of total citation counts. To determine h^2 upper in WoS, publications in a scientist's publication list (sorted by 'times cited') that have citation counts greater than the scientist's h index value must be marked and for these publications a citation report produced to obtain the sum of citations for these publications (see 'sum of the times cited' in the report). If $h \times h$ is subtracted from this sum, the result – given in percent of the total citation counts – is h^2 upper. To obtain h^2 lower, h^2 upper and h^2 center must be subtracted from 100.

As Fig. 1 shows, the three applicants A, B, and C belonging to one isohindex group (here $h = 14$), have very different values for h^2 lower, h^2 center, and h^2 upper. This indicates very different research output types. Whereas h^2 lower for applicant A makes up about 3% of the entire area of the citation distribution, for applicant C this is 57%. We find the opposite for h^2 upper, which makes up 82% of the entire area of the citation distribution for applicant A but only 10% for applicant C. Applicant A is a scientist who has rather few but very highly cited publications. Cole and Cole (1967) call this type of scientist *perfectionists*, researchers who publish "comparatively little but what they do publish has a considerable impact on the field" (p. 382). Applicant B can be called a *prolific scientist* following Cole and Cole (1967), "in the dual sense of producing an abundance of papers which tend also be fruitful" (p. 382), and Applicant C is a *mass producer*, a scientist who publishes "a relatively large number of papers of little consequence" (p. 382).

Table 1 shows the area proportions h^2 lower, h^2 center, and h^2 upper for applicants that have similar h index values (they belong to very similar isohindex groups). For example, for applicants who have an h index value of 10 or 11 ($n = 40$), h^2 center covers on average 20% of the area of the distribution of citations. The minimum value for the applicants is on average 6% and the maximum value 44%. This means, for one, that the variability in the covering of the citation distribution by h^2 center is very high, and it shows, for another, that in this h index subgroup there is no applicant for whom h^2 center makes up at least one-half of the entire area. As the percentages in the column total show, h^2 lower generally makes up 7% of the area. Across all h index subgroups, h^2 center refers to only about one-fourth of the entire area. The by far greatest share of the area (70%) generally traces back to h^2 upper. Hence, especially the area of highly cited papers in a distribution will be depicted only insufficiently by the h index value.

Altogether, on the basis of the three graphs in Fig. 1 and the minimum and maximum values in Table 1, it is clearly visible that for applicants *belonging to very similar isohindex groups*, h^2 center covers a very different proportion of the area of the individual citation distribution. Among the applicants having similar h index values, there are therefore very different research output types, which can be identified through the additional values h^2 lower and h^2 upper, however.

3.2. Calculation of sRM values

According to Seglen (1992) "each individual scientist may constitute a 'microfield' with a characteristic citation probability determined by that individual's research profile" (p. 637). Accordingly, the 'true' value of the number of publications in the core (that is, the scientist's most visible publications) can result only from the citation distribution over the scientist's publications. As we showed in the section above, however, with very different citation distributions for scientists the h index values can still be very similar. Thus, the h index only insufficiently captures the complete distribution in the calculation. But with the use of the segmented regression model (sRM) there is a way to determine the 'true' value (we will call it the sRM value) of the number of publications in the core based on an *individual* citation distribution. The sRM is a common statistical optimization method that is used mainly in medicine (e.g., Kim, Yu, & Feuer, 2008) and forest science (e.g., Sauter, Mutz, & Munro, 1999) and that we here propose for use in evaluative bibliometry. Seen statistically, bibliometric data have a certain degree of uncertainty (white noise, sampling variability, measurement error). Statistical methods, such as the sRM, therefore seek to estimate functional correlations in such a way that this uncertainty is minimized. Thus, in sRM the curves are estimated such that the squared difference between the predicted value (the sRM value) and the real cumulative citation value will be on average as small as possible (method of least squares of residuals).

Usually, the cumulated citation counts increase from a high rank publication to a low rank publication, with a steep slope in the first part ('core' publications with high visibility) and a flat slope in a second part (publications with low visibility) (see Fig. 2). To statistically model the citation transition zone between the first and the second part there is a need for two separate regressions to obtain a reasonable fit for the whole citation distribution (Mutz, Guilley, Sauter, & Nepveu, 2004;

Table 1
Arithmetic mean, standard deviation, minimum and maximum of h^2 lower, h^2 center, and h^2 upper by different h index values of the applicants (percentages).

h index value	Number of applicants	Arithmetic mean	Standard deviation	Minimum	Maximum
h^2 lower					
≤7	32	4	5	0	18
8–9	47	6	10	0	50
10–11	40	5	6	0	27
12–13	50	6	7	0	39
14–15	43	8	11	0	57
16–17	30	9	9	0	43
18–19	22	6	6	0	21
≥20	33	9	7	1	25
Total	297	7	8	0	57
h^2 center					
≤7	32	17	14	4	64
8–9	47	22	12	3	60
10–11	40	20	8	6	44
12–13	50	25	10	11	47
14–15	43	25	9	11	48
16–17	30	25	8	9	45
18–19	22	24	8	10	39
≥20	33	26	9	8	45
Total	297	23	10	3	64
h^2 upper					
≤7	32	79	17	29	96
8–9	47	72	19	10	97
10–11	40	76	13	29	94
12–13	50	69	15	34	88
14–15	43	66	17	9	88
16–17	30	67	15	32	91
18–19	22	70	12	42	90
≥20	33	64	14	41	91
Total	297	70	16	9	97

Sauter et al., 1999). The first part can be best described by a quadratic curve, the second part by a linear curve. Segmented regression is a statistical model that is able to simultaneously estimate the parameters of the two curves and the joint point between publications with high impact (the 'true core') and publications with low impact (Gallant & Fuller, 1973; Zhan et al., 1996).

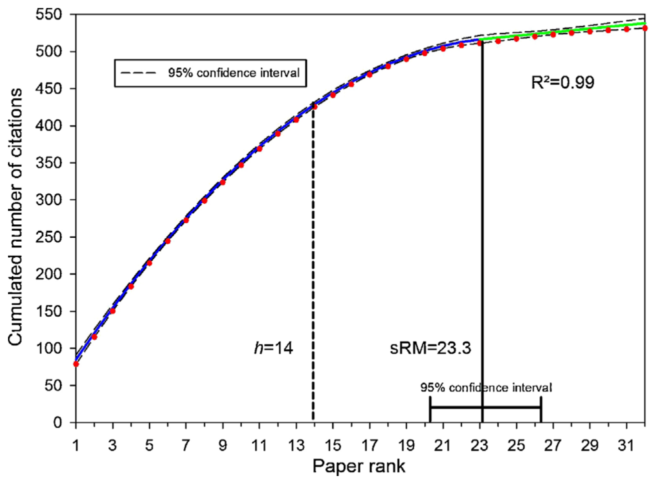

Fig. 2. sRM value and h index value for one applicant with $h = 14$.
Note. The red dots are the applicant's cumulated number of citations. The blue line is the fitted quadratic curve (with 95% confidence interval), and the green line is the fitted linear curve (with 95% confidence interval). The R^2 of 99% indicates that the fitted values explain the applicant's cumulated number of citations almost completely. The h index value (14), which is clearly lower than the sRM value (23.3) with 95% confidence interval (|-|), shows that the h index underestimates the number of highly cited papers of the applicant by about 11 publications. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of the article.)

Table 2
Arithmetic mean, standard deviation, minimum and maximum of sRM values by different *h* index values of the applicants.

h index value	Number of applicants	sRM values			
		Arithmetic mean	Standard deviation	Minimum	Maximum
≤7	20	5.94	1.87	2.83	9.41
8–9	37	10.19	4.83	2.95	27.59
10–11	35	9.93	2.69	4.41	14.72
12–13	35	13.86	6.29	3.53	41.90
14–15	36	15.95	8.75	7.95	60.73
16–17	28	18.73	6.87	9.94	43.77
18–19	19	18.79	5.24	9.05	31.67
≥20	31	24.83	10.13	8.20	57.39
Total	241	14.75	8.45	2.83	60.73

Note: sRM values could not be computed for 56 of the total of 297 applicants, because one or more requirements for calculation of the sRM were not met.

The following sRM for y_j was assumed, whereby z_0 is the break point between publications in the first part and publications in the second part of the distribution: if $x_j < z_0$

$$y_j = b_0 + b_1 x_j + b_2 x_j^2 + e_j, \quad e_j \sim N(0, \sigma_e^2)$$

otherwise

$$y_j = b_0 + b_1 z_0 + b_2 z_0^2 + b_3(x_j - z_0) + e_j, \quad e_j \sim N(0, \sigma_e^2) \tag{4}$$

The x values (ranked publications) for j range from 1 to k. The z_0 value is an unknown parameter that is defined as the maximum of the quadratic function:

$$z_0 = \frac{-b_1}{2b_2} \tag{5}$$

It can be determined by estimating the parameters b_1 and b_2. Alternatively, it can be estimated as an explicit model parameter (see Bang, Mazumdar, & Spence, 2006). The latter allows computation of the confidence interval of z_0.

The sRM can be estimated using non-linear least squares. The statistical software SAS, for instance, offers a procedure called NLIN, which allows computation of the parameters with Gauss–Newton iteration (Bang et al., 2006; Berman, Wong, Bhasin, & Ipp, 1996) (see the SAS program in Appendix A). Kim et al. (2008) discuss Lerman's grid search or Hudson's continuous fitting algorithm as equivalent computational alternatives.

An important advantage of the sRM is that the size of the residual variance (σ_e^2), or the proportion of explained variance to total variance (R^2), gives some evidence for the amount of model fit. The sRM is applicable to a scientist's publication set, if (i) two different parts in the citation distribution can be clearly distinguished (this is mostly the case for scientists, as their distributions of publications' citedness are found to be very skewed, Seglen, 1992), (ii) the algorithm converges, (iii) R^2 is high (>.90), (iv) the breakpoint lies within the range of publications, and (v) there are in sum at least 15–20 publications for one scientist (Kim et al., 2008). These requirements should be seen as clear advantages of the model in application: an index (also the *h* index) should not be computed for every scientist (for instance, when comparing scientists having low publication and citation counts, it is not very meaningful). With the aid of the sRM model criteria, the user gains an indication as to whether it makes sense to compute an index for a given publication list.

For subgroups of applicants having similar *h* index values, Table 2 shows the arithmetic means, standard deviations, and minimum and maximum values of sRM values. As the note of the table points out, sRM values could not be computed for 56 of the total of 297 applicants, because one or more requirements for calculation of the sRM were not met (see above). As the means in the table show, when there is an increase in the *h* index values, there is also an increase in the sRM values. At the aggregate level, sRM and *h* correspond. However, the sRM values are very spread out around the means in Table 2. Across all applicants the standard deviation of sRM values is 8.45. Thus, among the applicants with an *h* index of 8 or 9, there are scientists with sRM values ranging from 2.95 to 27.59. The other isohindex subgroups show a similar picture. This means that scientists who based on the *h* index value should have a similar research output have a very differently sized core of most visible publications—when the *individual* citation distribution is taken as a basis.

Computing a sRM brings not only advantages to the user, however. It should therefore generally be used as a complementary measure to the *h* index. Whereas the *h* index value provides an indication of the absolute number of citations of the publications in the 'Hirsch core' (at least *h* citations each), this information is lacking in the sRM value. Scientist A can have a higher sRM value than scientist B even though scientist B has, for each of his publications, received many more citations than scientist A. Therefore, sRM values should be compared across individuals only with great care. Another disadvantage of the sRM value might be the more difficult interpretation than the *h* index due to its more complicated computation. However, visualization (see Fig. 2) makes it possible to display the different segments in the citation distribution of a scientist. This makes the results understandable also to persons who do not have a working knowledge of statistics.

4. Discussion

In addition to the advantages of the h index (such as simple calculation), a number of disadvantages have been named in recent years (Bornmann & Daniel, 2007; Jin, Liang, Rousseau, & Egghe, 2007), which has led to the development of numerous h index variants (Bornmann, Mutz, & Daniel, 2008). As we showed in several publications (Bornmann, Mutz, & Daniel, 2008, 2009; Bornmann, Mutz, Daniel, Wallon, & Ledin, 2009), only some of these variants are associated with an incremental contribution for evaluation purposes—the h index and many of the variants are highly correlated. We therefore do not think it would be wise to develop further variants of the h index in future, but it is useful to complement the h index with additional information in order to obtain a more complete and more reliable picture of the research output of a single scientist. According to Franceschini and Maisano (2010) the "h index is used to compare the scientists' research output. This is possible because the h measurement scale has the ordinal property. In other terms, if we compare two scientists – a and b – and if $h(a) > h(b)$, then a is considered better than b. However, this scale property can be in contrast to other measures of diffusion and productivity" (p. 500). In this study we presented h^2 lower, h^2 center, and h^2 upper, and the sRM value as approaches that provide high information content for the h index research output measurement. As it was the aim of this study to present these two approaches in their basic application, we did not go into other potential applications. But with the aid of h^2 lower, h^2 center, and h^2 upper, further interesting comparisons can be made, for example between the citation sum in the h core (sum of h^2 center and h^2 upper) and the number of citations to papers outside the h core (h^2 lower).

Acknowledgements

The authors would like to thank Dr. Gerlind Wallon, deputy director of the European Molecular Biology Organization (EMBO), and Dr. Anna Ledin, scientific secretary at the Royal Swedish Academy of Sciences in Stockholm (and formerly at EMBO), for providing the bibliographic data on the applicants to the EMBO Young Investigator Programme.

Appendix A.

SAS program for calculating the sRM:

```
proc nlin data=dataset;
parms a = 35 to 70 by 25
     b = 20  to 80  by 10
     c = 4 to -4 by -0.50
     d=  2 to -2 by -0.25;
x0=-0.5 * b/c;
model cumcitation=ifn(art<x0, a+b*art+c*art*art,a+b*x0+c*x0*x0+d*(art-x0));
run;
```

References

Adler, R., Ewing, J., Taylor, P., & Hall, P. G. (2009). A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Statistical Science*, *24*(1), 1–28. doi:10.1214/09-sts285

Anderson, T., Hankin, R., & Killworth, P. (2008). Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics*, *76*(3), 577–588.

Bang, H., Mazumdar, M., & Spence, J. D. (2006). Tutorial in biostatistics: Analyzing associations between total plasma homocysteine and B vitamins using optimal categorization and segmented regression. *Neuroepidemiology*, *27*(4), 188–200. doi:10.1159/000096149

Berman, N. G., Wong, W. K., Bhasin, S., & Ipp, E. (1996). Applications of segmented regression models for biomedical studies. *American Journal of Physiology–Endocrinology and Metabolism*, *33*(4), E723–E732.

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*, *58*(9), 1381–1385.

Bornmann, L., & Daniel, H.-D. (2009). The state of *h* index research. Is the h index the ideal way to measure research performance? *EMBO Reports*, *10*(1), 2–6.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the *h* index? A comparison of nine different variants of the *h* index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, *59*(5), 830–837.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2009). Do we need the h index and its variants in addition to standard bibliometric measures? *Journal of the American Society for Information Science and Technology*, *60*(6), 1286–1289.

Bornmann, L., Mutz, R., Daniel, H.-D., Wallon, G., & Ledin, A. (2009). Are there really two types of *h* index variants? A validation study by using molecular life sciences data. *Research Evaluation*, *18*(3), 185–190.

Bornmann, L., Mutz, R., Neuhaus, C., & Daniel, H.-D. (2008). Use of citation counts for research evaluation: Standards of good practice for analyzing bibliometric data and presenting and interpreting results. *Ethics in Science and Environmental Politics*, *8*, 93–102. doi:10.3354/esep00084

Bornmann, L., Wallon, G., & Ledin, A. (2008a). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European Molecular Biology Organization programmes. *PLoS One*, *3*(10), e3480.

Bornmann, L., Wallon, G., & Ledin, A. (2008b). Is the h index related to (standard) bibliometric measures and to the assessments by peers? An investigation of the h index by using molecular life sciences data. *Research Evaluation*, *17*(2), 149–156.

Burrell, Q. L. (2007). On the h-index, the size of the Hirsch core and Jin's A-index. *Journal of Informetrics*, *1*(2), 170–177.

Cole, S, & Cole, J. R. (1967). Scientific output and recognition: A study in operation of reward system in science. *American Sociological Review*, *32*(3), 377–390.

Egghe, L. (2005). Lotkaian informetrics: an introduction. In *Power laws in the information production process: Lotkaian informetrics*. Kidlington, UK: Elsevier Academic Press.

Evidence Ltd. (2007). An indicator process. In *The use of bibliometrics to measure research quality in UK higher education institutions*. London, UK: Universities UK.

Franceschini, F., & Maisano, D. A. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research*, *203*(2), 494–504. doi:10.1016/j.ejor.2009.08.001

Gągolewski, M., & Grzegorzewski, P. (2009). A geometric approach to the construction of scientific impact indices. *Scientometrics*, *81*(3), 617–634. doi:10.1007/s11192-008-2253-y

Gallant, A. R., & Fuller, W. A. (1973). Fitting segmented polynomial regression models whose join points have to be estimated. *Journal of the American Statistical Association*, *68*(341), 144–147.

García-Pérez, M. (2009). A multidimensional extension to Hirsch's h-index. *Scientometrics*, *81*(3), 779–785. doi:10.1007/s11192-009-2290-1

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.

Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19193–19198. doi:10.1073/pnas.0707962104

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, *52*(6), 855–863.

Joint Committee on Quantitative Assessment of Research. (2008). *Citation statistics. A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS)*. Berlin, Germany: International Mathematical Union (IMU).

Kim, H. J., Yu, B., & Feuer, E. J. (2008). Inference in segmented line regression: A simulation study. *Journal of Statistical Computation and Simulation*, *78*(11), 1087–1103. doi:10.1080/00949650701528461

Kim, Y.-M. (2009). Validation of psychometric research instruments: The case of information science. *Journal of the American Society for Information Science and Technology*, *60*(6), 1178–1191.

Ledin, A., Bornmann, L., Gannon, F., & Wallon, G. (2007). A persistent problem. Traditional gender roles hold back female scientists. *EMBO Reports*, *8*(11), 982–987.

Lehmann, S., Jackson, A., & Lautrup, B. (2008). A quantitative analysis of indicators of scientific performance. *Scientometrics*, *76*(2), 369–390.

Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2006). Measures for measures. *Nature*, *444*(7122), 1003–1004.

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science and Technology*, *60*(7), 1327–1336.

Mutz, R., Guilley, E., Sauter, U. H., & Nepveu, G. (2004). Modelling juvenile-mature wood transition in Scots pine (Pinus sylvestris L.) using nonlinear mixed-effects models. *Annals of Forest Science*, *61*(8), 831–841. doi:10.1051/forest:2004084

Panaretos, J., & Malesios, C. (2009). Assessing scientific research performance and impact with single indices. *Scientometrics*, *81*(3), 635–670. doi:10.1007/s11192-008-2174-9

Rousseau, R. (2006). New developments related to the Hirsch index. *Science Focus*, *1*(4), 23–25.

SAS Institute Inc. (2009). *SAS/STAT® 9.2 user's guide*. Cary, NC, USA: SAS Institute Inc.

Sauter, U. H., Mutz, R., & Munro, B. D. (1999). Determining juvenile-mature wood transition in Scots pine using latewood density. *Wood and Fiber Science*, *31*(4), 416–425.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628–638.

van Eck, N. J., & Waltman, L. (2008). Generalizing the *h*- and *g*-indices. *Journal of Informetrics*, *2*(4), 263–271.

Wohlin, C. (2009). A new index for the citation curve of researchers. *Scientometrics*, *81*(2), 521–533. doi:10.1007/s11192-008-2155-z

Zhan, M., Dean, C. B., Routledge, R., Gallaugher, P., Farrell, A. P., & Thorarensen, H. (1996). Inference on segmented polynomial models. *Biometrics*, *52*(1), 321–327.

Zhang, C. T. (2009). The e-index, complementing the h-index for excess citations. *PLoS ONE*, *4*(5).