Annals of Anatomy 193 (2011) 191–196

Contents lists available at ScienceDirect

Annals of Anatomy

journal homepage: www.elsevier.de/aanat

Education

How to judge a book by its cover? How useful are bibliometric indices for the evaluation of "scientific quality" or "scientific productivity"?

Oliver von Bohlen und Halbach*

Institute for Anatomy and Cell Biology, Ernst Moritz Arndt University of Greifswald, Friedrich Loeffler Strasse 23c, 17487 Greifswald, Germany

ARTICLE INFO

Article history:
Received 3 December 2010
Accepted 17 March 2011

Keywords:
h-Index
Bibliometrics
m-Index
Evaluation

SUMMARY

How to pre-select the most promising candidates for an open position out of several applications? One of the possibilities is to check the personal bibliometric indices of these candidates by looking into appropriate databases. In these databases the number of publications, the total number of citations, the average number of citations per paper and the *h*-index are easy to find. Thus, it is easy to use these parameters for a pre-election. First, the particular values for the several bibliometric indicators could be retrieved for scientists working in the field of Anatomy & Cell Biology. Next, an analysis of how useful and reliable these bibliometric indicators are is performed. Most of the indicators strongly depend on the seniority of a researcher. Thus, these indicators favour older scientists over younger ones. Based on that, these indicators are not well-suited to identify young and promising scientists. Therefore, it is proposed that indicators, which correct for the time a scientists spends working in the field, may be better suited for such a pre-election, such as the *hy* index (also known as *m*-index) or the *Py* index. In this context, it should be emphasized that these indicators may be useful for pre-selection. All available indicators are based on data obtained from the past achievements of the scientists and may not predict their future achievements. However, despite the availability of these indicators, the best method to gain an impression of the quality is currently still the old-fashioned method of reading the papers.

1. Introduction

A central question in evaluation processes is how to define the impact and relevance of the scientific research output of a scientist. This question is not easy to answer, since it is difficult to quantify the quality of individual scientists. A general assumption is that it is better to publish more than less and that the citation count of a paper is a useful measurement of its quality. Since publication records are increasingly used in funding or appointment decisions, it is important how the citations are counted, and how these counts are weighted and analysed. A scientist's full citation record is normally determined by (1) counting the number of published items (2) summarizing the citation count of all papers, published by an individual scientist, (3) calculating the average citations per paper, (4) using the Hirsch index (Hirsch, 2005). Especially the Hirsch index (*h*-index) is nowadays seen as an important indicator of scientific quality. The *h*-index is considered helpful because it includes both the scientific productivity and the apparent scientific impact of a scientist (Rieder et al., 2010).

Access to the full citation distribution for an entire subfield is essential for the analysis. Existing databases can therefore actively help to quantify the quality of individual scientist. Indeed, the standard criteria were presented in different databases (Web of Science (ISI); Scopus (Elsevier) or Google Scholar (Google Inc.)) that are often used for evaluation.

Here, it is not necessary to discuss which is the best model, based on theoretical statistical or probabilities models. Instead, the reliability and usefulness of the available methods should be tested – from a practical point of view, e.g. in filling an open position in the discipline of Anatomy & Cell Biology. Thus, the indicators must be useful for the analysis of a very small group of individuals, since only a small number of applications would be included in the search for a suitable candidate.

The different available parameters each have different advantages and disadvantages. These will be briefly summarized:

1. The total number of publications (*P*) measures mainly productivity; however, it does not reflect the impact (van Raan, 2006).
2. The personal impact factor (*C*) calculates how many times all the published items of one author have been cited. This factor has been introduced as a factor reflecting the quality of scientists (Bornmann and Daniel, 2009). However, the personal citation

* Tel.: +49 3834 86 5313; fax: +49 3834 86 5302.
E-mail address: oliver.vonbohlen@uni-greifswald.de

0940-9602/$ – see front matter © 2011 Elsevier GmbH. All rights reserved.
doi:10.1016/j.aanat.2011.03.011

Table 1
Raw data used for the subsequent analysis. The seven highest values as well as the three lowest values per parameter were highlighted by different colours.

Name	P	C	CPP	h	years	hy	Py	Cy	CPPy
A	97	1236	12.74	20	23	0.87	4.22	53.74	0.55
B	99	2723	27.51	30	14	2.14	7.07	194.50	1.97
C	89	1531	17.20	19	13	1.46	6.85	117.77	1.32
D	169	4149	24.53	38	19	2.00	8.89	218.37	1.29
E	79	1120	14.18	22	20	1.10	3.95	56.00	0.71
F	60	404	6.73	13	19	0.68	3.16	21.26	0.35
G	55	808	14.69	15	19	0.79	2.89	42.53	0.77
H	97	1329	13.70	20	24	0.83	4.04	55.38	0.57
I	207	4328	20.91	37	37	1.00	5.59	116.97	0.57
J	144	506	3.51	11	20	0.55	7.20	25.30	0.18
K	193	5657	29.31	39	33	1.18	5.85	171.42	0.89
L	57	951	16.68	19	18	1.06	3.17	52.83	0.93
M	39	1033	26.49	14	13	1.08	3.00	79.46	2.04
N	136	1798	13.22	26	21	1.24	6.48	85.62	0.63
O	81	1022	12.62	20	11	1.82	7.36	92.91	1.15
P	91	931	10.23	18	29	0.62	3.14	32.10	0.35
Q	134	2232	16.66	28	20	1.40	6.70	111.60	0.83
R	283	3215	11.36	31	28	1.11	10.11	114.82	0.41
S	147	1769	12.03	25	16	1.56	9.19	110.56	0.75
T	57	464	8.14	13	11	1.18	5.18	42.18	0.74
U	41	555	13.54	14	17	0.82	2.41	32.65	0.80
V	39	256	6.56	10	16	0.63	2.44	16.00	0.41
W	66	865	13.11	16	13	1.23	5.08	66.54	1.01
X	44	562	12.77	14	20	0.70	2.20	28.10	0.64
Y	43	742	17.26	16	21	0.76	2.05	35.33	0.82
Z	53	783	14.77	17	13	1.31	4.08	60.23	1.14
ZA	28	545	19.46	14	11	1.27	2.55	49.55	1.77
ZB	34	767	22.56	14	16	0.88	2.13	47.94	1.41
ZC	368	10555	28.68	51	44	1.16	8.36	239.89	0.65
ZD	118	3375	28.60	32	22	1.45	5.36	153.41	1.30
ZE	317	11198	35.32	61	35	1.74	9.06	319.94	1.01
ZF	144	2663	18.49	31	19	1.63	7.58	140.16	0.97
Mean	**112.78**	**2189.75**	**16.99**	**23.38**	**20.47**	**1.16**	**5.23**	**93.28**	**0.90**
SEM	**19.94**	**387.10**	**3.00**	**4.13**	**3.62**	**0.21**	**0.92**	**16.49**	**0.16**

Ranking 1 2 3 4 5 6 7 8 - 29 30 31 32

Ranking: 1 for the highest value in the respective category (highlighted in red).

rates are often based on the journal impact factor. In that context, it should be kept in mind that the article citation rates determine the journal impact factor, not vice versa (Seglen, 1997). Thus, the journal impact factors may not be representative of individual journal articles.

3. To overcome this problem, the average citations per paper index (*CPP*) as well as the *h*-index are often used. The *CPP* allows comparison of scientists of different ages, but is criticised for the fact that it rewards low productivity and penalizes high productivity (Hirsch, 2005).
4. The *h*-index was intended to address the main disadvantages of other bibliometric indicators. Thus, the *h*-index should be an indicator of the quality and sustainability of scientific output. Moreover, the *h*-index is thought to represent a better way to assess long-term performance of authors than using the journal's impact factors (Hunt et al., 2010). However, whether the *h*-index is a superior indicator of scientific quality in terms of both, accuracy and precision, is still a matter of debate (Balandin and Stancliffe, 2009; Hirsch, 2007; Honekopp and Kleber, 2008; Lehmann et al., 2006; Zhang, 2009).

The *h*-index is currently one of the most used and important factor used in evaluation of the personal impact. Like the other mentioned parameters (*P*, *C*, *CPP*) the *h*-index is easy to access in the different databases without the need for any off-line data processing. Thus, these factors, despite their individual drawbacks, were used in case of evaluations, since access to these data is simple. Indeed, there are other measurements available (Thompson et al., 2009), like the crown indicator (van Raan, 2006), which may help in evaluation processes; however, these measurements are not easily accessed and therefore have to be calculated in a more or less complicated manner – a major drawback for these additional indices.

All these bibliometric measures (*P*, *C*, *CPP*, and *h*-index) can be used in case of an evaluation to fill an open position in the field "Anatomy & Cell Biology", since these measurements (in this instance) would not depend on the field of study in which the papers are published and cited.

In this paper, answers to the following questions have been attempted:

What are good values for these parameters in case of "Anatomy & Cell Biology"? This question should be answered since different disciplines have different citation patterns; so each field would need different thresholds (Ball, 2005).

How useful are these parameters to identify a suitable person? Thus, are these parameters really helpful or do we need other parameters for defining the impact and relevance of the scientific research output of a scientist?

2. Materials and methods

The ISI Web of Science database was used (September, 11th, 2010). The tool "Author finder" was used to identify the records of a scientist. Using the tool "Create Citation Report", the total number of papers (*P*), total number of citations (*C*), citations per paper (*CPP*) and the *h*-index were obtained. From the graph "Published items per year", the time (*y*) from the first published item until the last published item (in this case mainly the year 2010) was calculated. Since the focus was on individuals working in the field "Anatomy & Cell Biology", with a reasonable personal record, parameters were obtained for 32 members of the "German Anatomical Society", among them 21 members of the editorial board of the journal "Annals of Anatomy". The obtained values were ranked (with rank 1 being the highest value, see Table 1). For each parameter the mean and the standard error of the mean (SEM) were calculated and correlation analysis and linear regression analysis for different parameters were performed using Prism 5.03 (GraphPad Inc., USA).

3. Results

To get an overview, the mean values for the different parameters were calculated first. This evaluation gives an impression of the publication record and helps to get an overview of the impact and relevance of the scientific research output (Table 1). The mean number of publications of the scientists was 112.8 ± 19.9 (ranging from 28 to 368 publications) and thee mean number of Citations (*C*) an individual scientists collected was 2190 ± 387.1 (ranging from 256 to 11,198). The mean *CPP* for the scientists in that field was 16.99 ± 3.0 (ranging from 3.51 to 35.3). Concerning the *h*-index, the mean was 23.4 ± 4.1 (ranging from 10 to 61, Table 1).

Curiously, it was found that the *h*-index strongly depends on *C*. Indeed, linear regression analyses shows that there is a strong relation between *C* and the *h*-index ($R^2 = 0.9104$; $p < 0.0001$; Fig. 1A). This strong relationship of the *h*-index can also be expressed as $h_{cal} = SQRT(C/3.1)$, resulting in values close to the *h*-index (Table 2). Indeed, h_{cal} nearly predicts h ($R^2 = 0.9610$; $p < 0.0001$; Fig. 1B).

Table 2
h_{cal} can be used to predict the *h*-index. The median for *h* was 19.5 and for h_{cal} 18.6.

Name	*h*	h_{cal}	Difference
A	20.00	19.97	0.03
B	30.00	29.64	0.36
C	19.00	22.22	−3.22
D	38.00	36.58	1.42
E	22.00	19.01	2.99
F	13.00	11.42	1.58
G	15.00	16.14	−1.14
H	20.00	20.71	−0.71
I	37.00	37.36	−0.36
J	11.00	12.78	−1.78
K	39.00	42.72	−3.72
L	19.00	17.51	1.49
M	14.00	18.25	−4.25
N	26.00	24.08	1.92
O	20.00	18.16	1.84
P	18.00	17.33	0.67
Q	28.00	26.83	1.17
R	31.00	32.20	−1.20
S	25.00	23.89	1.11
T	13.00	12.23	0.77
U	14.00	13.38	0.62
V	10.00	9.09	0.91
W	16.00	16.70	−0.70
X	14.00	13.46	0.54
Y	16.00	15.47	0.53
Z	17.00	15.89	1.11
ZA	14.00	13.26	0.74
ZB	14.00	15.73	−1.73
ZC	51.00	58.35	−7.35
ZD	32.00	33.00	−1.00
ZE	61.00	60.10	0.90
ZF	31.00	29.31	1.69
Mean	23.38	23.91	−0.15
SEM	4.13	4.23	−0.03

It is also of interest for such an analysis to know how long it would take for a scientist in the field of Anatomy & Cell Biology to achieve these bibliometric records. Within the selected group, the mean time-frame in which the scientists have published papers was 20.5 ± 3.6 (ranging from 11 to 44 years; Table 1). Thus, the average record cannot be achieved by a younger scientist. In case of an evaluation for an appointment (e.g. assistant professorship) or post-doc position, these data may not very helpful. It is even questionable whether these bibliometric data are helpful indicator of scientific quality, since most of these indicators strongly depend on the time a scientist worked in the field. Thus, scientists who have worked in their field for a longer period than younger scientists have a greater chance to publish more. Indeed, the time-factor (*y*) has strong influence upon: (1) the number of publications (*P*; $R^2 = 0.6487$; $p < 0.0001$; Fig. 2A), (2) the number of citations (*C*; $R^2 = 0.5951$; $p < 0.0001$, Fig. 2B), (3) the *h*-index ($R^2 = 0.5304$; $p < 0.0001$, Fig. 2C), and (4) to a lesser degree upon the numbers of average citations per paper (*CPP*; $R^2 = 0.1537$; $p < 0.0265$; Fig. 2D).

In order to obtain a suitable parameter that will be more independent of *y*, different approaches are possible:

1. A correction of *P* for time ($Py = P/y$); indeed this indicator is not strongly influenced by *y* ($R^2 = 0.0945$; $p = 0.0869$; Fig. 3A). The mean values for *Py* were 5.23 ± 0.92 (Table 1).
2. A correction of *C* for time ($Cy = C/y$); however, this indicator is still influenced by *y* ($R^2 = 0.2541$; $p = 0.0033$; Fig. 3B). The mean value for *Cy* was 93.28 ± 16.49 (Table 1).
3. A correction of *CPP* for time ($CPPy = CPP/y$); however, this indicator is also influenced by *y* ($R^2 = 0.1868$; $p = 0.0135$; Fig. 3C). The mean value for *CPPy* was 0.90 ± 0.16 (Table 1).

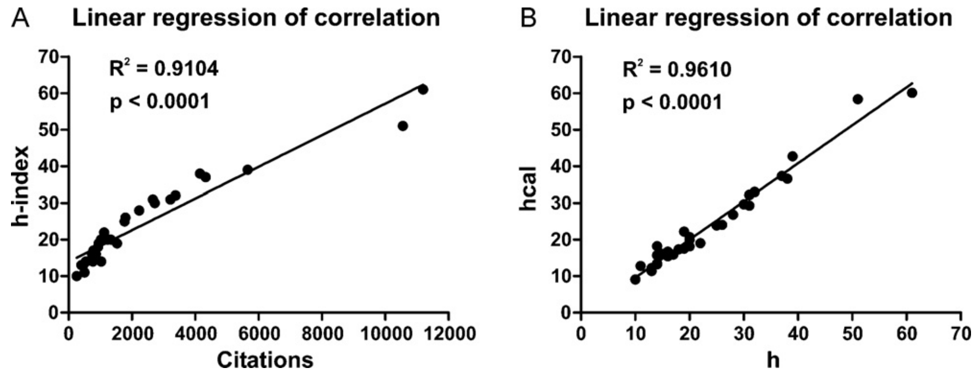

Fig. 1. Linear regression analysis revealed that h strongly correlates with the number of citations (A). This strong relationship can be described by $h_{cal} = \mathrm{SQRT}(C/3.1)$. The correlation for h_{cal} and h is shown in B.

4. A correction of the *h*-index for time (*hy*). Such a correction factor has already been introduced by Hirsch (Hirsch, 2005) and has been termed *m*-index ($m = hy = h/y$); indeed this indicator is not strongly influenced by y ($R^2 = 0.0208$; $p = 0.4309$; Fig. 3D). The mean value for *hy* was 1.16 ± 0.21 (Table 1).

4. Discussion

The analysis revealed that successful scientists in the field of Anatomy & Cell Biology have about 113 publications and have collected a *C* of about 2200, a *CPP* of about 17. Moreover, the analysis showed that these scientists have an *h*-index of about 23 (the *h*-index is about 12.5 for full professors in academic radiology in the USA (Rad et al., 2010) and 16 for full professors in neurosurgery (Lee et al., 2009)).

Values equal to or above these values for each parameter would indicate a very successful scientist. However, by comparing the different individuals (Table 1) it is somewhat surprising that the different parameters roughly lead to very similar results. Thus, individuals with a high productivity (high number of publications) also have a high personal impact factor (*C*) and often a high *CPP*. However, CPP can also penalize high productivity and reward low productivity (Table 1; case "M"). Based on that, it can be speculated that the *h*-index may indeed represent a better parameter for evaluation of scientific quality and sustainability of scientific output. The results obtained for the ranking based on the *h*-index differ somewhat from the ranking of the numbers of publications, but is close to the ranking based on the personal impact factor. Indeed, this strong relationship can be approximately described as $h_{cal} = \mathrm{SQRT}(C/3) \sim h$. The results obtained for h_{cal} are close to the values of the *h*-index.

However, a scientist with such a record (*P*: 113; *C*: 2200; *CPP*: 17; *h*-index: 23) would rarely send an application. The reason is that it would take more than 20 years to achieve such a record. However,

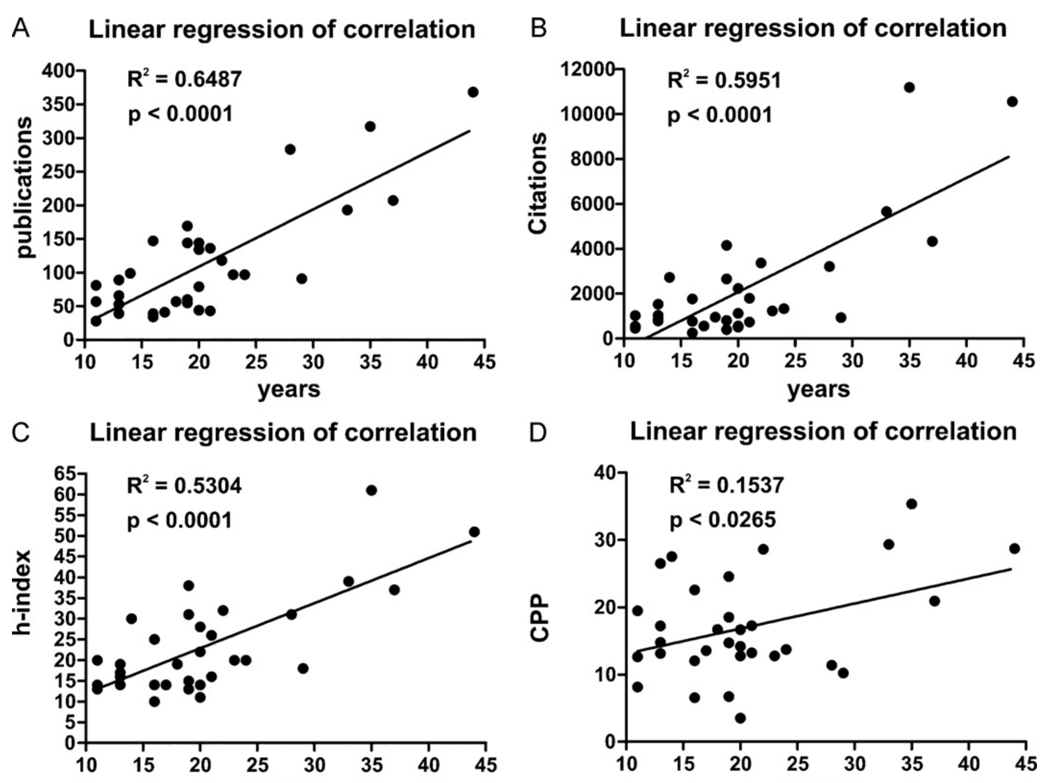

Fig. 2. Linear regression analysis for the correlation of the parameters: number of publications (A); number of citations (B), *h*-index (C) and CPP (D) with years showed that all of these parameters are strongly influenced by the time a scientist has worked in the scientific field (years).

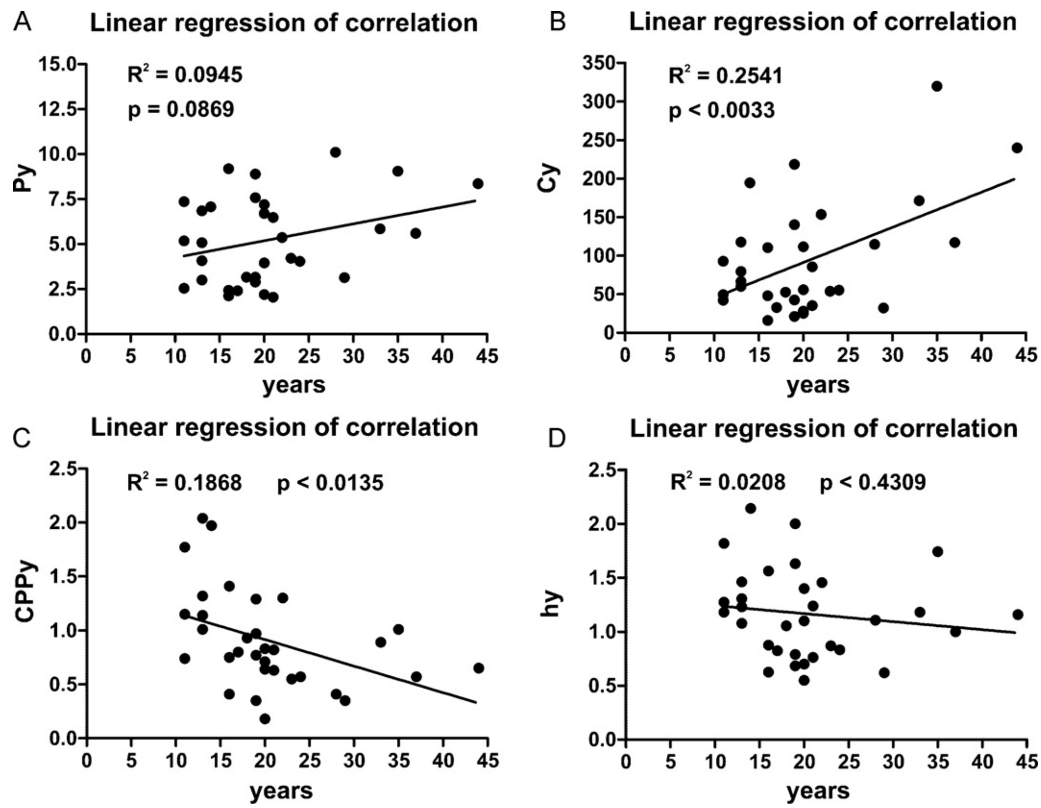

Fig. 3. Linear regression analysis for the parameters: publications, citations, *h*-index and CPP were corrected for time (*y*). *Py* was found not to be as strongly influenced by *y* (A) as *P* (see Fig. 2A for comparison). Time-corrections for *C* (*Cy*) as well as for CPP were still found to be influenced by *y* (B and C). Correction of the *h*-index for *y* (*hy*): linear regression analysis showed that *hy* was not strongly influenced by *y*.

in most cases the applicants are younger and thus have not worked for such a long time in that field. Therefore, these indices are not very suitable in this scenario.

Moreover, there is another drawback for these indices, since they do not take the time factor into account. The time-factor (y) is the number of years between the first published papers of a scientist until the present. Why is this time-factor important for such evaluation processes? The reason is very simple: scientists who have worked in their field for a longer period than younger scientists have a greater chance to publish more. Indeed, it was found that the time-factor has a great influence upon the *P*, *C*, and to a lesser degree also *CPP*. What is about the *h*-index? The *h*-index favours scholars who have consistently published papers over many years (Thompson et al., 2009) and is influenced by the duration of a scientific career. Thus, it looks as if it is not possible to compare scientists and their scientific records, if the different scientists have worked for different time periods in their scientific field. This would indicate that younger scientists always have less chances for appointments than older colleagues. This also would imply that, for those who have to make the decisions, they have no instrument to look for promising young scientists.

In his paper from 2005, Hirsch not only introduces the *h*-index, but also the *m*-index (Hirsch, 2005). The *m*-index is a correction of the *h*-index for time ($m = h/y$). According to Hirsch, the *m*-value is an “indicator of the successfulness of a scientist” and the parameter *m* should provide a useful yardstick to compare scientists of different seniority. Therefore, this value is not only suitable for identifying overall successful scientists, but also should help to get a measurement of the quality corrected for the time of scientific activity. Indeed, no correlation of y with m (=hy) was found ($R^2 = 0.0208$; $p = 0.4309$). Thus, for fair comparisons of individual scientist of one scientific field, the *m*-index will provide a measurement that is much more independent of the number of years between the first published paper of a scientist until the present than the *h*-index. Moreover, the *m*-index, unlike the *CPP* does not reward low productivity.

Since the *m*-index is corrected for age, it may be useful in predicting future achievement (which is relevant in this scenario). The *m*-index can be seen as an indicator for “scientific quality” with the advantage (as compared to the *h* index) that the *m*-index is corrected for age. In this scenario the mean value for this index was 1.16 ± 0.20. Thus, a successful scientist in the field of Anatomy & Cell Biology should have an *m* index of about 1. Interestingly, an *m* value of about 1 also characterizes, e.g. a successful physicist (Hirsch, 2005). Thus, this index may also be useful for comparisons of scientists of different fields.

The *Py*-index is also suitable to compare scientists of different seniority, since this factor is also corrected for the time. *Py*, however, is not a suitable indicator for “scientific quality”, but an indicator for “productivity”, since *Py* is the mean number of published items per year.

Based on the data obtained, it is suggested that in case of evaluation, the *m*-index and *Py* should be calculated and these indicators should also be used in case of a bibliometric based evaluation, since all other indicators depend on the time a scientists spends in the field. Thus, bibliometric indicators (with the exception of the *m*-index and *Py*) are not well-suited to compare scientists of different seniority and thus may have only limited informative value in case of an evaluation.

Until now, there is no reliable and unbiased “magic number” available that describes the “scientific quality and productivity”. The available indicators may be useful for a pre-selection, but it should kept in mind that all available indicators are based on data obtained from the former achievements of the scientists and may not predict their future achievements. If bibliometric indicators are used for pre-selection or evaluation, one should be aware of their specific limitations and one should focus on factors that may favour an unfair discrimination. However, despite the availability of these indicators, the best method to gain an impression of the quality is still the old-fashioned (and time-consuming) method of reading the papers.

References

Balandin, S., Stancliffe, R.J., 2009. Impact factors and the h-index: what researchers and readers need to know. Augment. Altern. Commun. 25, 1–3.

Ball, P., 2005. Index aims for fair ranking of scientists. Nature 436, 900.

Bornmann, L., Daniel, H.D., 2009. The state of h index research. Is the h index the ideal way to measure research performance? EMBO Rep. 10, 2–6.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. U.S.A. 102, 16569–16572.

Hirsch, J.E., 2007. Does the H index have predictive power? Proc. Natl. Acad. Sci. U.S.A. 104, 19193–19198.

Honekopp, J., Kleber, J., 2008. Sometimes the impact factor outshines the H index. Retrovirology 5, 88.

Hunt, G.E., Cleary, M., Walter, G., 2010. Psychiatry and the Hirsch h-index: the relationship between journal impact factors and accrued citations. Harv. Rev. Psychiatry 18, 207–219.

Lee, J., Kraus, K.L., Couldwell, W.T., 2009. Use of the h index in neurosurgery. Clinical article. J. Neurosurg. 111, 387–392.

Lehmann, S., Jackson, A.D., Lautrup, B.E., 2006. Measures for measures. Nature 444, 1003–1004.

Rad, A.E., Brinjikji, W., Cloft, H.J., Kallmes, D.F., 2010. The H-index in academic radiology. Acad. Radiol. 17, 817–821.

Rieder, S., Bruse, C.S., Michalski, C.W., Kleeff, J., Friess, H., 2010. The impact factor ranking – a challenge for scientists and publishers. Langenbecks Arch. Surg. 395 (Suppl. 1), 69–73.

Seglen, P.O., 1997. Why the impact factor of journals should not be used for evaluating research. BMJ 314, 498–502.

Thompson, D.F., Callen, E.C., Nahata, M.C., 2009. New indices in scholarship assessment. Am. J. Pharm. Educ. 73, 111.

van Raan, A.F.J., 2006. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. Scientometrics 67, 491–502.

Zhang, C.T., 2009. The e-index, complementing the h-index for excess citations. PLoS One 4, e5429.