Bibliometric indicators of young authors in astrophysics: Can later stars be predicted?

Frank Havemann · Birger Larsen

Received: 15 April 2014 / Published online: 27 November 2014 © Akadémiai Kiadó, Budapest, Hungary 2014

Abstract We test 16 bibliometric indicators with respect to their validity at the level of the individual researcher by estimating their power to predict later successful researchers. We compare the indicators of a sample of astrophysics researchers who later co-authored highly cited papers before their first landmark paper with the distributions of these indicators over a random control group of young authors in astronomy and astrophysics. We find that field and citation-window normalisation substantially improves the predicting power of citation indicators. The sum of citation numbers normalised with expected citation numbers is the only indicator which shows differences between later stars and random authors significant on a 1 % level. Indicators of paper output are not very useful to predict later stars. The famous $h$-index makes no difference at all between later stars and the random control group.

Keywords Bibliometric indicators · Research evaluation · Astrophysics · Citation analysis · Performance of young authors · Comparison of indicators

Introduction

Any indicator should actually indicate what it is made for. If an indicator is used for evaluation it should not provide an incentive for an unwanted behaviour. However, in scholarly publishing instances have been reported of such behaviours, including multiple publications, least publishable units, unjustified assignment of co-authorship, and different
tactical citation practices. Bibliometricians should strive to develop valid research indicators which have no unwanted adverse effects (Kreiman and Maunsell 2011).

Most bibliometric indicators are not developed for the evaluation of individual researchers (Costas et al. 2010, p. 1565), however individuals are increasingly being evaluated using such indicators. We test selected indicators with respect to their validity at the level of the individual researcher by estimating their power to predict later successful researchers. For this reason, we compare bibliometric indicators of a sample of astrophysics researchers who later co-authored highly cited papers (later stars, for short) before their first landmark paper with the distributions of these indicators over a random control group of young authors in astronomy and astrophysics.

Results obtained with some standard basic indicators have been presented as a poster at ISSI 2013. Here we extend the study to more sophisticated measures with the aim to find the best indicators for predicting later stars. We imagine a scenario where later stars apply for a job in an astrophysical research institute 5 years after their first paper in a journal indexed in Web of Science (WoS). Do they perform better bibliometrically than the average of applicants with the same period of publishing?

**Data and method**

**Sampling of authors**

To identify the sample of later stars and a random control sample, we inspected 64 astronomy and astrophysics journals to find researchers who started publishing after 1990 and had published for a period of at least 5 years in WoS journals. We excluded those who had more than 50 co-authors on average because evaluating those ‘big science’ authors at an individual level cannot be supported by bibliometrics. We draw a random sample of 331 authors mainly publishing in this field and affiliated longer in Europe than elsewhere. The latter criterion conflicts with the international character of astrophysics research but makes the sample more homogenous with respect to the educational and cultural background of the researchers. We avoid identifying homonymous authors by using the Author Search option of Web of Science and by selecting only authors who have not too frequent names.2

To find authors with highly cited papers, for each journal considered we ranked papers with more than four citations per year and less than ten authors according to their citations per year. We excluded papers with ten or more authors because we want to have later stars whose contributions to the successful papers are not too small. From the top 20 % of these paper rank-lists we extracted all European authors of highly cited papers. We obtained 362 candidates who published their first highly cited paper at least 5 years after their first paper in one of the 64 journals.

We ranked these later star candidates according to their number of highly cited papers. We went through this list and checked whether the authors had really 5 years or more to wait for the breakthrough if all their papers in WoS-journals are taken into account. We chose the first 40 authors to keep the effort manageable. For all WoS-papers of the 40 later stars and of the 331 random authors (downloaded at Humboldt-University, Berlin) all citing papers were identified by CWTS, Leiden. All bibliometric indicators presented

---

1 14th International Society of Scientometrics and Informetrics Conference in Vienna, Austria, 15th to 20th July 2013 (Havemann and Larsen 2013).

2 We exclude candidates with many paper sets offered by Author Search.
below are based on papers and their citations within the first 5 years of the author. To compare only authors with similar collaboration behaviour we restricted both samples to authors with less than four and more than one co-author on average ending up with 30 later stars and 179 random authors.

We further restricted both samples to authors starting before 1999 because there is only one star starting later (in 2002) but many random authors (more than 100). By this restriction to 29 stars and 74 authors in the control group we take into account that the citation behaviour of astrophysicists has changed remarkably during the last 25 years. The number of references has increased. The median of number of references of the 448 papers published in the 1986 volume of the *Monthly Notices of the Royal Astronomical Society* was 24. Until the year 2010 the median of reference numbers has doubled (calculated with 2,006 papers, data source: WoS). Longer lists of references induce higher citation numbers of papers. Thus, both samples still have a time variance of expected citation numbers. This time variance increases the overlap between the citation-indicator distributions of the samples when citation numbers are not normalised. In other aspects the union of our samples is more homogenous than many real groups of applicants (career duration, collaboration behaviour, geographical background).

We then tested whether the 29 authors in the sample are really later stars i.e. have co-authored a highly cited paper after their first 5 years but not within this period. To be specific, we define a later star as an author who has a paper with twice the yearly citation number of the most cited paper in the first 5 years. This is a relative definition but any absolute citation threshold could cause a subfield bias. Two authors in the sample do not match the definition. Their most cited papers have, respectively, 18.8 and 19.0 citations p.a. and their most cited papers within their first 5 years have 10.0 and 10.5 yearly citations. We omit these two authors from the sample of later stars. The citation data of the most cited papers co-authored by the remaining 27 later stars is given in section “Most cited papers of later stars“ of Appendix (p. 13).

An alternative data source for astrophysics publications and their citations is the *Astrophysics Data System* (ADS) delivered jointly by the US National Aeronautics and Space Administration (NASA) and the Smithsonian Astrophysical Observatory (Henneken et al. 2011). ADS includes also non-refereed publications. Any user can obtain a whole slew of bibliometric indicators for any set of selected publications.

**Statistics**

For each bibliometric indicator considered, we test whether both samples behave like random samples drawn from the same population by applying a one-sided Wilcoxon rank sum test with continuity correction. We test the null hypothesis that for both samples we have the same probability of drawing an author with a larger value in the other sample. The alternative hypothesis is that indicator values of later stars exceed the values of random authors (Sachs and Hedderich 2006, p. 391).

We have also tested the hypothesis that for both samples we have the same probability of drawing an author with a larger value of the *collaborative coefficient* (Ajiferuke et al.

---

3 cf. Henneken et al. (2011, p. 5).
4 Topics in physics as in astrophysics differ substantially in citation density (Radicchi and Castellano 2011; Pepe and Kurtz 2012).
5 http://adsabs.harvard.edu.
6 cf. the Wikipedia article http://en.wikipedia.org/wiki/Mann-Whitney-Wilcoxon_test.
In both samples we have a similar collaboration behaviour (cf. Fig. 1). If we refuse the null hypothesis we would fail in more than one half of possible cases (test probability $p = .581$). This result ensures that differences between both groups are not due to different typical team sizes.

All work was done using the free open-source statistics software R (which includes a graphics package).7

Selection of indicators

The indicators analysed here are listed together with their mathematical definitions in Table 1. In section “Descriptions of indicators” of Appendix we discuss the definition of each indicator.

We have calculated and tested two simple output indicators and nine indicators of influence. Beside pure numbers of papers and their citations within the first five publishing years of the authors we use fractionally counted papers and citations as the input for indicators of productivity and influence. Fractional counting is motivated by the assumption that the personal success of a co-author of a paper decays with the number of the paper’s authors. Its use in evaluation penalises unjustified assignment of co-authorship to others.

If we compare papers published in fields with different citation behaviour any citation indicator should be field normalised with expected citation numbers. Here we consider only one field but—as mentioned above—the citation behaviour of astrophysicists has changed dramatically within the last decades. We want to discriminate between the two samples but this is blurred by the broad variances of the distributions which are only

7 http://www.r-project.org (R-scripts for indicator calculation and sample data can be obtained from the first author of this paper).
caused by the changing citation behaviour. Hence, distributions of unnormalised citation indicators of the two samples overlap partly due to the changing citation behaviour.

Another desired effect of normalising with expected citation numbers is that we account for different citation windows of papers. Thus, citations to papers published in the beginning of a period obtain a lower weight than those to papers published in the last year. The estimation of expected citation numbers of papers is described in section “Expected citation numbers” of Appendix.

Another method to deal with varying citation behaviour is to determine each paper’s percentile in the citation distribution of a control sample of papers. Bornmann et al. (2013) compare five approaches to this promising method. Percentile ranking avoids the use of arithmetic means of heavily skewed citation distributions. We minimise the influence of skewness by calculating expected citation numbers by a linear regression over all years considered (see section “Expected citation numbers” of Appendix). We have to leave a test of the percentile method with our samples to further work due to a lack of citation data of control samples.

Recently, several authors tested a third approach to field normalisation of citation numbers. Here data on the citing side are normalised. Waltman and van Eck (2013, see also references of this paper) discuss three variants of this method. Also this approach cannot be tested with our data due to the same lack of citation numbers of control samples. We could

Table 1  List of author indicators: \(a_i\) is the number of authors of paper \(i\); \(c_i\) is the number of citations of paper \(i\); \(E(c_i)\) is the expected number of citations of paper \(i\) (cf. section “Expected citation numbers” of Appendix); we assume that papers of an author are ordered according to \(c_i\) and denote the paper’s rank with \(r\); the effective rank is defined as \(r_{eff}(r) = \sum_{i=1}^{r} 1/a_i\)

| Name               | Definition                                                                 |
|--------------------|---------------------------------------------------------------------------|
| **Productivity**   |                                                                           |
| Number of papers   | \(\sum_i 1 = n\)                                                         |
| Fractional score   | \(\sum_i 1/a_i = f\)                                                     |
| **Total influence**|                                                                           |
| Number of citations| \(\sum_i c_i = c\)                                                        |
| Norm. number cit.  | \(\sum_i c_i/E(c_i) = c_{norm}\)                                         |
| \(j\)-index        | \(\sum_i \sqrt{c_i}\)                                                   |
| Fract. citations   | \(\sum_i c_i/a_i\)                                                      |
| Fract. norm. cit.  | \(\sum_i c_i/(E(c_i)a_i)\)                                              |
| **Typical influence**|                                                                           |
| Mean cit. number   | \(\sum_i c_i/n = c/n\)                                                  |
| Mean fract. cit.   | \(\sum_i (c_i/a_i)/n\)                                                  |
| Median fract. cit. | median\((c_i/a_i)\)                                                     |
| Max. fract. cit.   | max\((c_i/a_i)\)                                                        |
| **\(h\)-type indices**|                                                                           |
| Hirsch index       | max\((r\lceil c_r \geq r \rceil) = h\)                                  |
| \(g\)-index        | max\((r\lceil \sum c_i/a_i \geq r \rceil)\)                           |
| **Fractional \(h\)-type**|                                                                           |
| \(h_m\)-index      | max\((r_{eff}\lceil c_{r(eff)} \geq r_{eff} \rceil)\)                  |
| \(g_{r}\)-index    | max\((r\lceil \sum c_i/a_i \geq r \rceil)\)                           |
| \(g_m\)-index      | max\((r_{eff}\lceil \sum r_{(r(eff))} c_i/a_i \geq r_{eff}^2 \rceil)\) |
| **Collaboration**  |                                                                           |
| Collab. coeff.     | \(1 - f/n\)                                                             |
test the simplest variant where each citation of a paper is divided by the number of all references of the citing paper (Zhou and Leydesdorff 2011; Pepe and Kurtz 2012). Waltman and van Eck (2013) and also Radicchi and Castellano (2012) found that this fractional counting of references does not properly normalise for field and subfield differences. A further drawback of this variant is that citation numbers are not corrected for the age of the cited paper. We therefore did not test it.

In addition to the eleven indicators of productivity and of influence we calculated the widely used Hirsch or $h$-index (Hirsch 2005), a number combining influence and output performance in an uncontrolled and arbitrary manner, and four variants of it which have been introduced to avoid disadvantages of the original Hirsch index.

We did not consider any indicator based on the number of highly cited papers because this contradicts our sampling procedure: we selected later stars who have no highly cited paper in their first 5 years of publishing.

Results

Medians of all 16 indicators of both samples are given in Table 2. We have also added a row for the controlling variable, the collaborative coefficient. In the next to last column of Table 2 we list the failure probability $p$ of rejecting the null hypothesis that both samples behave like random samples drawn from the same population. In the last column we give the rank $R$ according to $p$. For all but the collaboration indicator and the two indicators on least ranks (Hirsch index and median of fractional citation numbers) the stars’ sample has a higher median than the random sample.

The boxplots in Appendix allow a comparison of indicator distributions for both samples. The figures are ordered according to the ranking $R$. That means that $p$ values increase from the first to the last boxplot. The boxplots have a logarithmic scale because all indicator distributions are highly skewed. All citation indicators have zero values for some uncited authors in the control sample. Therefore we display the logarithm of indicator values $\log(\cdot)$ (Fig. 2).

The two indicators based on normalised citation numbers are the most useful among the 16 indicators considered (see Fig. 3). With respect to normalised numbers of citations both samples behave not like random samples from the same population. In this case, rejecting the null hypothesis has a failure probability below 1%.

The distributions of two further indicators differ at least on a 5% significance level (see Fig. 4). For the remaining 12 indicators there is no significant difference between distributions of later stars and of authors in the control group (see Figs. 5, 6, 7, 8, 9, 10). The Hirsch index is the worst indicator. It has very similar distributions for both samples ($p = 30.3\%$, see Fig. 10).

Discussion

Our results underline the necessity to correct citation indicators for the age of the cited papers and also for varying citation behaviour. The two indicators of total influence based on citation numbers normalised with expected citation numbers are the best ones among a

---

8 It would be interesting—from a theoretical point of view—to determine the influence of each correction separately.
total of 16 to help predict later stars. The normalised number of citations $c_{\text{norm}}$ is the only indicator which shows significant differences between later stars and random authors on a 1 % level. Thus, normalised citation indicators of total influence can indeed help to predict later successful authors. Despite this relative good performance of normalised citation indicators of total influence we cannot recommend to use them as the only basis for an evaluation of young authors in astrophysics and in similar fields of natural sciences. Normalisation at the field level cannot correct for a variability in citation numbers between different topics. Optof (2011) analysed the citation density in different topics of cardiovascular research papers and concluded that even normalised citation indicators “should not be used for quality assessment of individual scientists” (cf. his abstract and our footnote 4, p. 3). In each case, bibliometrics can only support evaluation and cannot replace individual peer review.

Table 2 Median indicators of samples, test probability $p$, and rank $R$ (according to $p$)

| Indicator          | Stars | Random | $p$  | $R$ |
|--------------------|-------|--------|------|-----|
| **Productivity**   |       |        |      |     |
| Number of papers   | 7     | 6      | .122 | 12  |
| Fractional score   | 2.67  | 1.86   | .135 | 13  |
| **Total influence**|       |        |      |     |
| Number of citations| 30    | 22.5   | .056 | 7   |
| Norm. number cit. | 5.89  | 3.83   | .008 | 1   |
| $j$-index          | 11.45 | 8.76   | .054 | 6   |
| Fract. citations   | 10.00 | 6.57   | .060 | 8   |
| Fract. norm. cit. | 1.71  | 1.10   | .015 | 2   |
| **Typical influence**|   |        |      |     |
| Mean cit. number   | 5.11  | 4.00   | .184 | 14  |
| Mean fract. cit.   | 1.20  | 0.99   | .101 | 11  |
| Median fract. cit. | 0.50  | 0.67   | .280 | 15  |
| Max. fract. cit.   | 4.50  | 3.00   | .064 | 9   |
| **h-Type indices** |       |        |      |     |
| Hirsch index       | 3     | 3      | .303 | 16  |
| $g$-index          | 5     | 4      | .072 | 10  |
| **Fractional h-type** |   |        |      |     |
| $h_m$-index        | 1.32  | 1.00   | .030 | 3   |
| $g_r$-index        | 3     | 2      | .046 | 4   |
| $g_m$-index        | 2.17  | 1.68   | .050 | 5   |
| **Collaboration**  |       |        |      |     |
| Collab. coeff.     | .683  | .683   | .581 | 17  |

None of the two output indicators have a significant difference below the 10 % level.\footnote{This is in accordance with the result obtained by Neufeld et al. (2013, cf. p. 9) when comparing successful with non-successful applicants of a funding programme for young researchers.} Thus, it is very unlikely to discover a later star in astrophysics by comparing her productivity with the productivity of a random author (Figs. 8, 9). The Hirsch index makes no difference at all ($p = 30.3 \%$, Fig. 10). This is in agreement with conclusions drawn by Lehmann et al. (2006) and also by Kosmulski (2012) who analysed small samples of
mature scientists and found that the number of publications “is rather useless” as a tool of assessment and that also the $h$-index is not really helpful. In contrast to these findings, Pudovkin et al. (2012) found that $h$-index and number of papers are indicators which differ most significantly between group leaders and other scientists at a medical research institution. This can surely be explained by real output differences of elder and younger researchers but maybe partly also by the assumption that more often group leaders have been working at the institute over the whole analysed 5-year period than other researchers.

We could have analysed the generalised $h$-index proposed by Radicchi et al. (2008) who use normalised citation and paper numbers. We did not because $h$ performs much worse than indicators of total influence (cf. Table 2).

The $g$-index proposed by Egghe (2006) to improve the $h$-index (by giving higher weight to highly cited publications) performs indeed better than the original ($p = 7.2\%$, Fig. 7). The same holds for the analysed three $h$-type indices which are based on fractional counting. They have been introduced by Egghe (2008) and by Schreiber (2008c; 2009) to account for varying collaboration behaviour.

There is no significant difference between the two samples when we compare citation indicators which are designed to reflect the mean influence of an author’s papers. We calculated three of them: the arithmetic mean of citation numbers ($p = 18.4\%$, Fig. 9), fractionally counted citations per paper ($p = 10.1\%$, Fig. 8), and the median of the fractionally counted citations ($p = 28\%$, Fig. 10). We wondered whether for a later star a large maximum of (fractional) citations is more typical than a large value of any measure of central tendency of citation numbers. The answer is yes. The maximum of fractional

---

**Fig. 2** Linear regressions and averages of citation numbers of papers of random authors in astrophysics after the first (the publication) year (red), the second year (orange), the third year (yellow), the fourth year (green), and the fifth year (blue). (Color figure online)
citations is a better indicator of typical influence \( (p = 6.4\% \quad \text{Fig. 7}) \). We could have analysed normalised indicators of typical influence, too. We did not because indicators of typical influence do not perform better than those of total influence.

We do not exclude self-citations when calculating citation indicators. There are arguments for their exclusion in evaluative bibliometrics but we assume that it would be difficult for young authors to massively cite their own papers within their first 5 years of publishing.

We expect that weighting (fractional) paper numbers with a measure of journal reputation would improve the predictive power of output indicators. We did not test this because the only journal-reputation indicator available to us was the journal impact factor which is not useful here—albeit often used for weighting paper numbers (Seglen 1997; Lozano et al. 2012, see also the references of these papers).

Analysing 85 researchers in oncology Hönekopp and Khan (2012) found that “a linear combination of past productivity and the average paper’s citation” is a better predictor of future publication success than any of the single indicators they had studied. We did not consider combinations of indicators of productivity and of mean influence because the simpler indicators of total influence also reflect productivity—as far as the produced papers have been cited. Neglecting uncited papers is a desired effect that is also quoted in favour of the \( h \)-index.

Hornbostel et al. (2009) found only small differences in numbers of publications and citations between approved and rejected applicants to a German funding programme for young researchers. In an earlier study, Nederhof and van Raan (1987) compared 19 Ph.D. graduates in physics with best degrees to 119 other graduates with lower grade. They considered the total number of papers before and after graduation and their total and average (short time) impact. The 19 best graduates performed significantly better but, interestingly, the impact of their papers declined and reached the level of the control-group papers a few years after graduation. The authors speculate about the reason of this phenomenon and suggest that better students could have been engaged for hot and therefore highly cited research projects. They conclude, that maybe “the quality of the research project, and not the quality of the particular graduate is the most important determinant of both productivity and impact figures” (Nederhof and van Raan 1987, p. 348). This hypothesis could also hold for the young astrophysicists we analysed. Its confirmation would further diminish the weight of bibliometric indicators in the evaluation of young researchers.

**Acknowledgments** We thank Jesper W. Schneider of Aarhus University for helpful discussions of an early draft and Paul Wouters at CWTS in Leiden for kindly providing citation data. We further thank two anonymous reviewers who helped us to sharpen our definition of later stars. The analysis was done for the purposes of the ACUMEN project, financed by the European Commission, cf. http://research-acumen.eu/.

**Appendix**

Most cited papers of later stars

In Table 3 the most cited papers of later stars are compared to their most cited papers published within the first 5 years they have papers in WoS-journals. The citation numbers \( p.a. \) (cit/year) are determined with a citation window of 5 years for the most cited papers in all years. For the most cited papers of the first 5 years the citation window is restricted to these 5 years because we suppose an evaluation directly after this period.
All authors in Table 3 have got more than twice the number of citations \( p.a. \) for their most cited paper than for their most cited paper in their first 5 years. The author on rank 8 has got more citations for the early most cited paper than the authors on ranks 2, 9, and 21 for their most cited paper in the whole period of analysis. Omitting these three later stars would result in a clearer picture but we cannot omit them without the risk of a bias with respect to subfields because differing citation numbers could be caused by different citation behaviours in different subfields (see footnote 4, p. 3). Our definition of later stars can only be relative: they have got substantially higher citation numbers for their most cited paper after their first 5 years than within these years.

In addition, some bibliometric indicators of the first 5 years of later stars are listed (cf. Table 1 for definitions). The table is ranked according to normalised number of citations \( c_{\text{norm}} \): \textit{Data source: Web of Science}.

### Table 3

Papers of later stars with highest annual citation numbers in all years and in the first 5 years of each later star (s.a. text)

| Rank | First year | \( n \) | \( c \) | \( h \) | \( c_{\text{norm}} \) | Most cited papers |
|------|------------|-------|-------|-------|-----------------|-----------------|
|      |            |       |       |       |     | All years | Year | cit/year | First 5 years | Year | cit/year |
| 1    | 1991       | 14    | 107   | 5     | 34.59 | 2002 | 54.2 | 1994 | 12.0 |
| 2    | 1992       | 15    | 157   | 8     | 27.55 | 2004 | 16.2 | 1992 | 6.6  |
| 3    | 1994       | 12    | 43    | 4     | 27.30 | 2007 | 24.6 | 1998 | 7.0  |
| 4    | 1992       | 12    | 72    | 5     | 17.42 | 1997 | 63.4 | 1993 | 4.5  |
| 5    | 1993       | 7     | 95    | 4     | 15.16 | 2004 | 28.8 | 1993 | 12.0 |
| 6    | 1992       | 27    | 88    | 6     | 13.43 | 2005 | 27.2 | 1993 | 4.3  |
| 7    | 1992       | 11    | 82    | 3     | 11.26 | 2004 | 27.6 | 1992 | 11.0 |
| 8    | 1996       | 5     | 30    | 2     | 10.47 | 2004 | 96.8 | 1999 | 16.5 |
| 9    | 1992       | 16    | 84    | 5     | 10.12 | 2001 | 14.4 | 1992 | 5.0  |
| 10   | 1994       | 9     | 46    | 3     | 10.03 | 2009 | 86.8 | 1994 | 4.4  |
| 11   | 1992       | 8     | 57    | 4     | 9.42  | 1999 | 57.2 | 1994 | 5.7  |
| 12   | 1991       | 5     | 28    | 3     | 9.05  | 1998 | 18.2 | 1992 | 2.8  |
| 13   | 1995       | 10    | 36    | 3     | 6.12  | 2006 | 24.2 | 1998 | 4.0  |
| 14   | 1992       | 9     | 27    | 3     | 5.89  | 2007 | 17.2 | 1995 | 3.5  |
| 15   | 1996       | 7     | 12    | 2     | 5.74  | 2007 | 24.6 | 1996 | 0.8  |
| 16   | 1995       | 11    | 27    | 3     | 5.24  | 2006 | 214.4 | 1996 | 3.5  |
| 17   | 1997       | 6     | 19    | 2     | 5.22  | 2006 | 18.2 | 1997 | 2.2  |
| 18   | 1998       | 3     | 47    | 3     | 4.93  | 2005 | 27.8 | 1999 | 8.0  |
| 19   | 1997       | 5     | 22    | 2     | 4.58  | 2006 | 21.0 | 2000 | 7.5  |
| 20   | 1991       | 3     | 30    | 2     | 4.11  | 1999 | 23.8 | 1991 | 4.2  |
| 21   | 1993       | 13    | 30    | 3     | 4.08  | 2003 | 10.6 | 1995 | 2.7  |
| 22   | 1994       | 4     | 40    | 3     | 3.90  | 2005 | 53.4 | 1994 | 5.0  |
| 23   | 1994       | 2     | 14    | 1     | 2.65  | 2007 | 24.6 | 1994 | 2.4  |
| 24   | 1996       | 6     | 13    | 2     | 2.13  | 2005 | 18.6 | 1999 | 2.0  |
| 25   | 1993       | 4     | 9     | 2     | 1.87  | 2004 | 80.6 | 1996 | 2.0  |
| 26   | 1991       | 3     | 1     | 1     | 1.85  | 2004 | 26.0 | 1995 | 1.0  |
| 27   | 1991       | 5     | 2     | 1     | 0.34  | 2002 | 125.0 | 1995 | 2.0  |
Nonetheless, we repeated all calculations for the smaller sample of 24 later stars. This gives us an impression how stable the results are under small changes of the sample. All \( p \) values are higher now——what is expected for a smaller sample if it would be generated by random reduction——but, again, the normalised number of citations \( c_{\text{norm}} \) is the best indicator \(( p = 1.8\% )\) and the \( h \)-index the worst measure \(( p = 50.8\% )\).

Descriptions of indicators

**Productivity indicators**

**Number of papers:** This elementary indicator of productivity belongs to a bygone era when co-authorship was the exception and not the rule. It has the unwanted adverse effects of multiple publishing of the same results and of honorary authorships.

**Fractional score:** Each paper \( i \) is divided into \( a_i \) fractions where \( a_i \) is the number of authors. These fractions are summed up for the papers of the evaluated author. We use the simplest variant where all fractions of a paper are equal: \( f = \sum i 1/a_i \). This indicator penalises honorary authorships and takes into account that larger teams can be more productive.

**Total influence**

All indicators of total influence tend to increase with the author’s number of papers. That means, they are also indicating productivity.

**Number of citations:** Each citation of a paper indicates that it has influenced the citing author(s). The sum \( c = \sum c_i \) of raw numbers \( c_i \) of citations of an author’s papers is highly field dependent. The paper’s number of citations \( c_i \) depends on the age of a paper at the time of evaluation. Highly cited papers have surely some quality but less cited ones can also be of high quality.

**Normalised numbers of citations:** We normalise each paper’s number of citations \( c_i \) by an expected number of citations \( E(c_i) \) which takes into account the paper’s age and the citation behaviour in astrophysics during the first 5 (calendar) years in the paper’s lifetime (cf. section “Expected citation numbers” of Appendix). After normalising each paper’s citation number we sum the ratios of observed and expected citation numbers:

\[
c_{\text{norm}} = \sum_{i=1}^{n} \frac{c_i}{E(c_i)}.
\]

Some bibliometrists do not calculate the sum of ratios but the ratio of sums (Schubert and Braun 1986): \( \frac{\sum c_i}{\sum E(c_i)} \). This procedure is thought to evaluate the whole oeuvre of an author but has been criticised recently for being not “consistent” (Opthof and Leydesdorff 2010; Waltman et al. 2011).\(^{10}\)

**The \( j \)-index:** The \( j \)-index is the sum of the square roots of citation numbers of the author’s papers

\[
j = \sum_{i=1}^{n} \sqrt{c_i}.
\]

It was proposed by Levene et al. (2012) to downgrade the influence of highly cited papers in the sum of citation numbers.

\(^{10}\) The \( h \)-index is also not consistent (Marchant 2009; Waltman and van Eck 2012).
**Fractional citations:** Analogously to the fractional score described above we distribute citations of each paper equally to its authors:

\[ \sum_{i=1}^{n} \frac{c_i}{a_i}. \]

**Fractional normalised citations:** The normalised numbers of citations can also be distributed among the authors involved (Radicchi and Castellano 2011):

![Box plots showing normalized citation distribution](image-url)
Fig. 4 The indicators on rank 3 and 4 according to $p$ values: $p<5\%$

$$\sum_{i=1}^{n} \frac{c_i}{E(c_i)a_i}.$$ 

Typical influence

Mean citation number: The arithmetic mean of citations of an author’s papers
is the simplest indicator of influence which does not tend to increase with the author’s productivity.

Mean fractional citations: The arithmetic mean of fractionally counted citations of an author’s papers:
Median of fractional citations: The median of fractionally counted citations of an author’s papers median\(\left(\frac{c_i}{a_i}\right)\) is considered because citation distributions are skewed.

Maximum of fractional citations: We wondered whether for a later star a large maximum of (fractional) citations \(\max\left(\frac{c_i}{a_i}\right)\) is more typical than a large value of any measure of central tendency of citation numbers (Lehmann et al. 2008, cf. p. 375).
Indices of h-type

Hirsch index: The $h$-index was introduced by Hirsch (2005) “to quantify an individual’s scientific research output.” It is defined as the maximum rank $r$ in a rank list of an author’s papers according to their citation numbers $c_i$ which is less than or equal to the citation number $c_r$ of the paper with rank $r$. $h = \max(r | c_r \geq r)$. The $h$-index has been criticised for its arbitrariness (van Eck and Waltman 2008). It is arbitrary because in the definition Hirsch “assumes an equality between incommensurable quantities” (Lehmann et al.
2008, p. 377), namely a rank and a citation number. Hirsch himself stated that his index depends on field-specific citation and collaboration behaviour (Hirsch 2005, p. 16571).

Egghe’s g-index: Egghe (2006) criticised the h-index for being insensitive to the citation frequency of an author’s highly cited papers. His g-index can be defined as the maximum rank \( r \) which is less than or equal to the mean citation number \( \frac{\sum_{i} c_i}{r} \) of papers till rank \( r \) (Schreiber 2008b). This condition is equivalent to \( \sum_{i} c_i \geq r^2 \). That means, \( g \) can also be defined as
Fractional indices of h-type

Schreiber’s $h_m$-index: Fractional counting of papers or of citations could be applied to define an $h$-index which takes multi-authorship into account (Egghe 2008; Schreiber 2007).
Schreiber (2008a) argued that fractionally counted citations could remove highly cited papers from the $h$-core if they have a lot of authors. This led him to define the $h_{\text{m}}$-index as the maximal effective rank $r_{\text{eff}} = \sum_i 1/a_i$ which is less than or equal to the number of citations $c_r$:

$$h_{\text{m}} = \max(r_{\text{eff}} | c_{r(r_{\text{eff}})} \geq r_{\text{eff}}).$$

**Egghe’s $g_f$-index:** Egghe (2008) proposed to define a fractional $g$-index $g_f$ as

![Box plots showing median of fractional citations](image)

**Fig. 10** The indicators on rank 15 and 16 according to $p$ values.
Here the citations are counted fractionally.

Schreiber’s $g_m$-index: Schreiber (2009) proposed a fractional $g$-index $g_m$ where both, papers and citations, are counted fractionally:

$$g_m = \max \left( r \sum_{i=1}^{r} \frac{c_i}{a_i} \geq r^2 \right).$$

Expected citation numbers

Usually, for field normalisation expected citation numbers of papers are calculated as arithmetic means of citation numbers of all papers (of the same document type) published in all journals of the field in the same year. There are two main technical problems with this method, the rough delineation of fields and the skewness of citation distributions.

We do not evaluate single authors but only want to show the influence of field normalisation on distributions of citation indicators of authors. Therefore we can use a random sample of papers (for which we have already the citation data) instead of all papers in the field. This sample contains papers published in the years 1991–2009 by all 331 random authors of our initial control sample. We only consider those 2,342 papers with at most 20 authors. Figure 2 shows the average cumulated citation numbers in the publication year, 1 year later, 2 years later etc. Due to the skewness of citation distributions these arithmetic means fluctuate. Therefore we made a linear regression for each of the five time series of citation numbers of papers (not of the averages) but restricted the analysis to the years 1995–2007 (coloured part of the regression lines) where we have more than 100 papers in each year. The interpolated citation numbers obtained by linear regression are used as expected citation numbers $E(c_i)$ of papers published in the corresponding years.

From these data we estimate a doubling of citation numbers in astrophysics in the two decades around the millennium. Calculating expected citation numbers as field averages is problematic because the arithmetic mean is not a good measure for the central tendency of skewed citation distributions. Lundberg (2007) therefore proposed to determine expected citation numbers as geometric means of citation numbers of papers in the field. Because papers can have zero citations he adds 1 to be able to calculate the geometric mean. This can be justified by saying that publishing a paper is the first citation of the published results.

References

Ajiferuke, I., Burrell, Q., & Tague, J. (1988). Collaborative coefficient: A single measure of the degree of collaboration in research. *Scientometrics, 14*, 421–433.

Bornmann, L., Leydesdorff, L., & Wang, J. (2013). Which percentile-based approach should be preferred for calculating normalized citation impact values? An empirical comparison of five approaches including a newly developed citation-rank approach (P100). *Journal of Informetrics, 7*(4), 933–944. s. a. http://arxiv.org/abs/1306.4454

Costas, R., van Leeuwen, T. N., & Bordons, M. (2010). A bibliometric classificatory approach for the study and assessment of research performance at the individual level: The effects of age on productivity and impact. *Journal of the American Society for Information Science and Technology, 61*(8), 1564–1581.

Egghe, L. (2006). An improvement of the $h$-index: The $g$-index. *ISSI Newsletter, 2*(2), 8–9.
Egghe, L. (2008). Mathematical theory of the h- and g-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology, 59*(10), 1608–1616.

Havemann, F., & Larsen, B. (2013). Bibliometric indicators of young authors in astrophysics: Can later stars be predicted? In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.), Proceedings of ISSI 2013 Vienna (Vol. 2, pp. 1881–1883).

Henneken, E. A., Kurtz, M. J., & Accamazzi, A. (2011). The ADS in the information age-impact on discovery, arXiv:1106.5644.

Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences, 102*(46), 16569–16572. http://arxiv.org/abs/physics/0508025.

Hönenkopp, J., & Khan, J. (2012). Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics, 90*(3), 843–853.

Hornbostel, S., Böhmer, S., Klingsporn, B., Neufeld, J., & von Ins, M. (2009). Funding of young scientist and scientific excellence. *Scientometrics, 79*(1), 171–190.

Kosmulski, M. (2012). Calibration against a reference set: A quantitative approach to assessment of the methods of assessment of scientific output. *Journal of Informetrics, 6*(3), 451–456.

Kreiman, G., & Mamssell, J. H. R. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience, 5*, article nr. 48 (6 pages).

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

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

Levene, M., Fenner, T., & Bar-Ilan, J. (2012). A bibliometric index based on the complete list of cited publications. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics (16), 1–6. s.a. arXiv:1304.6945.

Lozano, G. A., Lariivière, V., & Gingras, Y. (2012). The weakening relationship between the impact factor and papers’ citations in the digital age. *Journal of the American Society for Information Science and Technology, 63*(11), 2140–2145.

Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics, 1*(2), 145–154.

Marchant, T. (2009). Score-based bibliometric rankings of authors. *Journal of the American Society for Information Science and Technology, 60*(6), 1132–1137.

Nederhof, A. J., & van Raan, A. F. J. (1987). Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics. *Scientometrics, 11*(5–6), 333–350.

Neufeld, J., Huber, N., & Wegner, A. (2013). Peer review-based selection decisions in individual research funding, applicants’ publication strategies and performance: The case of the ERC starting grants. *Research Evaluation, 22*(4), 237–247.

Opthof, T. (2011). Differences in citation frequency of clinical and basic science papers in cardiovascular research. *Medical & Biological Engineering & Computing, 49*(6), 613–621.

Opthof, T., & Leydesdorff, L. (2010). Caveats for the journal and field normalizations in the CWTS (“Leiden”) evaluations of research performance. *Journal of Informetrics, 4*(3), 423–430.

Pepe, A., & Kurtz, M. J. (2012). A measure of total research impact independent of time and discipline. *PLoS One, 7*(11), e46428.

Pudovkin, A., Kretschmer, H., Stegmann, J., & Garfield, E. (2012). Research evaluation. Part I: Productivity and citedness of a German medical research institution. *Scientometrics, 93*(1), 3–16.

Radicchi, F., & Castellano, C. (2011). Rescaling citations of publications in physics. *Physical Review E, 83*(4), 046116.

Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics, 6*(1), 121–130.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences, 105*(45), 17268–17272.

Sachs, L., & Hedderich, J. (2006). *Angewandte Statistik. Methodensammlung mit R* (12th ed.). Berlin: Springer.

Schreiber, M. (2008a). A modification of the h-index: The h<sub>m</sub>-index accounts for multi-authored manuscripts. *Journal of Informetrics, 2*(3), 211–216.

Schreiber, M. (2008b). The influence of self-citation corrections on Egghe’s g index. *Scientometrics, 76*(1), 187–200. arXiv:0707.4577.

Schreiber, M. (2008c). To share the fame in a fair way, h<sub>m</sub> modifies h for multi-authored manuscripts. *New Journal of Physics, 10*(4), 040201.
Schreiber, M. (2009). Fractionalized counting of publications for the g-index. *Journal of the American Society for Information Science and Technology, 60*(10), 2145–2150.

Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics, 9*(5), 281–291.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ: British Medical Journal, 314*(7079), 498–513.

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

Waltman, 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.

Waltman, L., & van Eck, N. J. (2013). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *Journal of Informetrics, 7*(4), 833–849.

Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics, 87*(3), 467–481.

Zhou, P., & Leydesdorff, L. (2011). Fractional counting of citations in research evaluation: A cross- and interdisciplinary assessment of the Tsinghua University in Beijing. *Journal of Informetrics, 5*(3), 360–368.