

# Does the $h$ index have predictive power?

J. E. Hirsch\*

Department of Physics, University of California at San Diego, La Jolla, CA 92093-0319

Edited by Manuel Cardona, Max Planck Institute for Solid State Research, Stuttgart, Germany, and approved October 2, 2007 (received for review August 23, 2007)

**Bibliometric measures of individual scientific achievement are of particular interest if they can be used to predict future achievement. Here we report results of an empirical study of the predictive power of the  $h$  index compared with other indicators. Our findings indicate that the  $h$  index is better than other indicators considered (total citation count, citations per paper, and total paper count) in predicting future scientific achievement. We discuss reasons for the superiority of the  $h$  index.**

citations | prediction | achievement

The  $h$  index of a researcher is the number of papers coauthored by the researcher with at least  $h$  citations each (1). We have recently proposed it as a representative measure of individual scientific achievement. Other commonly used bibliometric measures of individual scientific achievement are total number of papers published ( $N_p$ ) and total number of citations garnered ( $N_c$ ). Recently, Lehmann *et al.* (2, 3) have argued that the mean number of citations per paper ( $n_c = N_c/N_p$ ) is a superior indicator. Here we wish to address the question: which of these four measures is best able to predict future scientific achievement?

For the purposes of this article, we do not wish to dwell on the controversial question of what is the optimal definition of scientific achievement.<sup>†</sup> We are not interested in measuring the *past* achievement of an individual, e.g., for the purpose of awarding a prize or for election to a prestigious academy, but rather in predicting *future* achievement. So we could simply bypass this question by defining “scientific achievement” by the bibliometric measure under consideration and ask: which measure is better able to predict its future values? For example, how likely is a researcher who today has a large number of citations to gain a large number of citations in future years? To the extent that a bibliometric measure reflects particular traits of the researcher rather than random events, it should have higher predictive power than another measure that is more dependent on random events. For example, we argued in ref. 1 that the total number of citations,  $N_c$ , “may be inflated by a small number of ‘big hits,’ which may not be representative of the individual if he/she is coauthor with many others on those papers.” For that individual, the present  $N_c$  value is not likely to be a good predictor of his/her future  $N_c$  values.

Alternatively, among the indicators listed in the first paragraph, it may be argued that the total number of citations,  $N_c$ , is the one that best reflects scientific achievement because it gives a measure of the integrated impact of a scientist’s work on the work of others. Then, we would like to know: which indicator is best able to predict  $N_c$  at a future time? It is certainly not obvious that the answer is  $N_c$  itself.

There are two slightly different questions of interest. (i) Given the value of an indicator at time  $t_1$ ,  $I(t_1)$ , how well does it predict the value of itself or of another indicator at a future time  $t_2$ ,  $I'(t_2)$ ? This question is of interest, for example, in trying to decide between different candidates for a faculty position. A possible consideration might be: how likely is each candidate to become a member of the National Academy of Sciences 20 years down the line? For that purpose, one would like to rank the candidates by their expected *cumulative* achievement after 20 years. This

means, in particular, that citations obtained *after* time  $t_1$  to papers written *before* time  $t_1$  are relevant. (ii) How well do the different indicators predict *future* scientific output? To award grant money or other resources for future research at time  $t_1$ , one would like to rank the candidates by their expected scientific output *from* time  $t_1$  *on* to some future time. In deciding who should get a grant, it should be irrelevant how many more citations the earlier papers of that individual are expected to collect in future years.

## Procedure

We use the ISI Web of Science database in the “General Search” mode (<http://isiwebofknowledge.com>).<sup>‡</sup> ISI has recently incorporated tools under “Author Finder” that help to discriminate between different researchers with the same name. Once the publications of a researcher are identified, ISI provides in the “Citation Report” the total number of citations  $N_c$ , total number of papers  $N_p$ , citations per paper  $n_c = N_c/N_p$ , and the  $h$  index, all *at the present time*.

ISI also allows one to restrict the time frame of the papers’ publication date, so it is easy to find  $N_p(t_1)$ , the number of papers published up to year  $t_1$ , or  $N_p(t_1, t_2)$ , the number of papers published between times  $t_1$  and  $t_2$ . However, ISI does not provide a similarly simple way to obtain these values for the other indicators under consideration. To obtain this information, one needs to export (save) the information in the Citation Report into an Excel file and add up the citations in the time interval desired. It is a straightforward but tedious procedure.

For illustration, we show in Fig. 1 the  $h$  index and the number of citations as a function of time for two prominent physicists: a theorist and an experimentalist. As we conjectured in ref. 1, the  $h$  index follows approximately a linear behavior with time, and the total number of citations is approximately quadratic with time. Similar behavior is found in many other cases.

## Sample PRB80

Because citation patterns vary between fields and also between subfields, and there are also trends with time, we chose to look at authors in a single subfield and of comparable scientific age, to compare the predictive power of the various indicators. Ideally we would like to pick a random subset of all physicists who earned their Ph.D. in a given subfield in a given year and published in that subfield throughout their career. However, we have no practical way to make such a selection. As an

Author contributions: J.E.H. designed research, performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

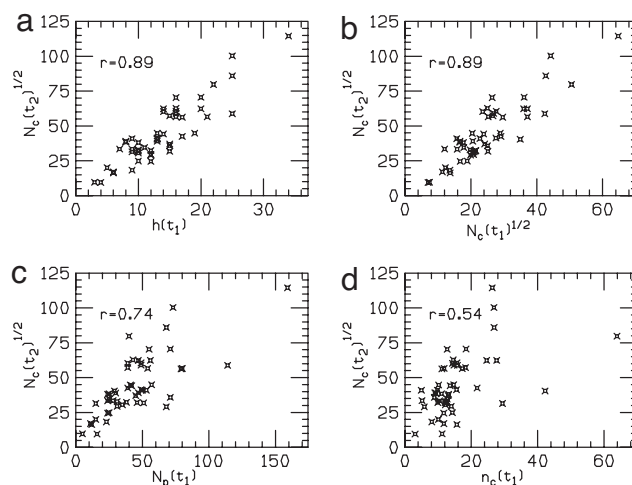
This article is a PNAS Direct Submission.

\*E-mail: [jhirsch@ucsd.edu](mailto:jhirsch@ucsd.edu).

<sup>†</sup>Ball P, Meeting of the Deutsche Physikalische Gesellschaft, March 26–30, 2007, Regensburg, Germany.

<sup>‡</sup>In using the very valuable ISI resource for individual evaluations, one should keep in mind that it has limitations, e.g., (i) it will, of course, miss citations where the author’s name is misspelled; (ii) books, book chapters, and most conference proceedings are not included; (iii) citations to “Rapid Communications” papers in *Phys Rev B* that include (R) in the citation are currently not counted by ISI.

© 2007 by The National Academy of Sciences of the USA



**Fig. 2.** Scatter plots of total number of citations,  $N_c$  after  $t_2 = 24$  yr vs. the value of the various indicators at  $t_1 = 12$  yr ( $t$  measured from the date of the first publication).  $h$ ,  $h$  index;  $N_p$ , number of papers;  $n_c$ , mean number of citations per paper;  $r$ , correlation coefficient, with  $r = 1$  corresponding to perfect linear correlation.

As discussed earlier, we would like to know:

- How well does the performance during the first 12 years predict the cumulative achievement over the entire 24-yr period?
- How well does the performance during the first 12 years predict performance in the subsequent 12 years?

Because the number of citations is expected to grow quadratically with time, we used  $\sqrt{N_c}$  as a measure of total citations.

First, we consider the predictive power of the various indicators after the first 12 years ( $t_1$ ) for the cumulative achievement in the 24-yr period ( $t_2$ ). In Fig. 2 we show the total number of citations after 24 years vs. each indicator after 12 years for each member of the sample, and their correlation coefficient,  $r$  (= covariance/product of standard deviations). It can be seen that the  $h$  index and the number of citations  $N_c$  at time  $t_1$  are the best predictors of cumulative citations at the future time  $t_2$ , with correlation coefficient  $r = 0.89$ . The number of papers correlates somewhat less ( $r = 0.74$ ), and the number of citations per paper,  $n_c$ , has lowest correlation with cumulative citations, with  $r = 0.54$ .

According to these results, if one wishes to select from among various candidates at time  $t_1$  the one(s) who will have the largest number of citations at the later time  $t_2$ , the  $h$  index or the number of citations at time  $t_1$  are good selection criteria. A candidate with low  $h$  or low  $N_c$  at time  $t_1$  will not have a high  $N_c$  at time  $t_2$ . Instead, a candidate with low  $N_p$  or low  $n_c$  at time  $t_1$  has a much higher chance of ending up with high  $N_c$  at time  $t_2$ .

Fig. 3 shows the ability of each indicator to predict its own cumulative value. Here, the differences between indicators is smaller and the correlation coefficient is high in all cases. Still, the  $h$  index shows the largest predictive power, with  $r = 0.91$ . That is, a researcher with a high  $h$  index after 12 years is highly likely to have a high  $h$  index after 24 years.

It is more difficult for the indicators at time  $t_1$  to predict scientific achievement occurring only in the subsequent period, i.e., without taking into account the citations after time  $t_1$  to work performed prior to  $t_1$ . As discussed, one would like to make such predictions to decide on allocation of research resources. In Fig. 4, the ability of the indicators at time  $t_1$  to predict citations to papers written in the  $t_1 - t_2$  time interval is considered. The highest correlation coefficient occurs for the  $h$  index ( $r = 0.60$ ) and the lowest for mean number of citations per paper ( $r = 0.21$ ).

| Time frame, yr | $h$ , $\sigma_h$ | $N_c$ , $\sigma_{N_c}$ | $N_p$ , $\sigma_{N_p}$ | $n_c$ , $\sigma_{n_c}$ | $a$ , $\sigma_a^*$ |
|----------------|------------------|------------------------|------------------------|------------------------|--------------------|
| 1–12           | 14, 6            | 740, 720               | 45, 26                 | 16, 10                 | 3.5, 1.2           |
| 1–24           | 24, 12           | 2,452, 2,471           | 90, 51                 | 25, 14                 | 4.1, 1.7           |
| 13–24          | 12, 8            | 759, 831               | 45, 33                 | 13, 10                 | 3.8, 1.7           |

19194 | www.pnas.org/cgi/doi/10.1073/pnas.0707962104



**Table 2. Averages and standard deviations in different time frames for the four indicators considered for sample APS95**

| Time frame | $h, \sigma_h$ | $N_c, \sigma_{N_c}$ | $N_p, \sigma_{N_p}$ | $n_c, \sigma_{n_c}$ | $a, \sigma_a^*$ |
|------------|---------------|---------------------|---------------------|---------------------|-----------------|
| Up to 1994 | 22, 7         | 1,887, 1,348        | 114, 53             | 17, 7               | 3.6, 0.9        |
| Up to 2006 | 36, 13        | 5,741, 4,468        | 205, 122            | 32, 41              | 4.2, 1.6        |
| 1995–2006  | 19, 11        | 1,718, 1,760        | 91, 87              | 19, 12              | 4.1, 2.1        |

\*This column gives  $a = N_c/h^2$ .

The data for cumulative achievement up to 2006 are shown in Figs. 9 and 10. It can be seen that the pattern is similar to Figs. 2 and 3, the corresponding graphs for sample PRB80.

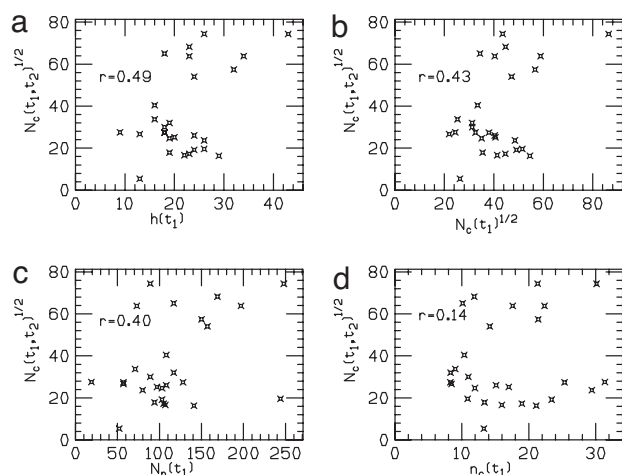
It is easy to understand why the correlations here are weaker than in the first example. Scientists are elected to APS fellowship at very different stages in their careers, so the horizontal axis variables in these figures are not time-normalized. For example, a member of this group might have had a large  $N_c$  in 1994 because he/she had been publishing for many years at a slow rate, and his/her productivity in the subsequent 12 years would not be expected to be larger than that of another scientist of this group who started his/her career much later and had a higher publication rate.

Note also that the 12-yr productivity and impact of the APS fellows sample (Table 2) is, on average, substantially higher than that of the random sample PRB80 (Table 1) and that there are no points on the  $x$  axes in the figures for the period  $t_1 - t_2$  for the APS sample (Figs. 7 and 8), in contrast to those of the PRB80 sample (Figs. 4–6). These differences are to be expected because election to APS fellowship is not a random process.

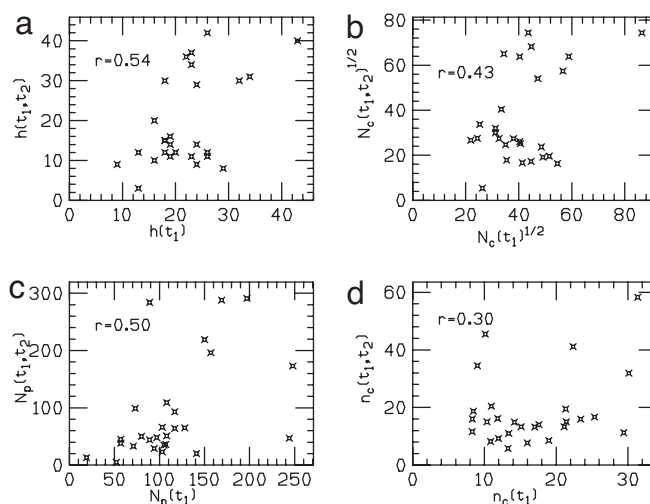
### Combining $h$ and $N_c$

Our results indicate that the  $h$  index and the total number of citations are better than the number of papers and the mean citations per paper to predict future achievement, with achievement defined by either the indicator itself or the total citation count,  $N_c$ . Furthermore, we found a small consistent advantage of the  $h$  index compared with  $N_c$ .

It has been argued in the literature that one drawback of the  $h$  index is that it does not give enough “credit” to very highly cited papers, and various modifications have been proposed to correct this, in particular, Egghe’s  $g$  index (4), Jin *et al.*’s AR index (5), and Komulski’s  $H^{(2)}$  index (6). These modified indices reward authors with higher citation numbers in the papers that contribute to the  $h$  count.



**Fig. 7.** Predictive power of each indicator at year 1994 for the number of citations to papers published in the 1995–2006 time interval, for sample APS95.



**Fig. 8.** Predictive power of each indicator at year 1994 for the value of the same indicator for the papers published in the 1995–2006 time interval, for sample APS95

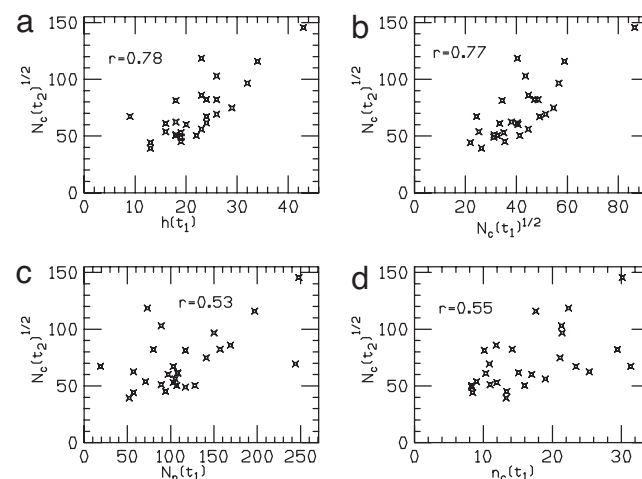
To test the possibility that giving a higher weight to highly cited papers may enhance the predictive power of the  $h$  index, we considered the following expression:

$$h_g \equiv \sqrt{h^2 + \alpha N_c} \quad [1]$$

and asked the question: which value of  $\alpha$  will result in  $h_\alpha(t_1)$  best predicting the citation count of future work,  $N_c(t_1, t_2)$ ? That is, we considered the cases of Figs. 4a and 7a with  $h_\alpha(t_1)$  in the abscissa instead of  $h(t_1)$ .

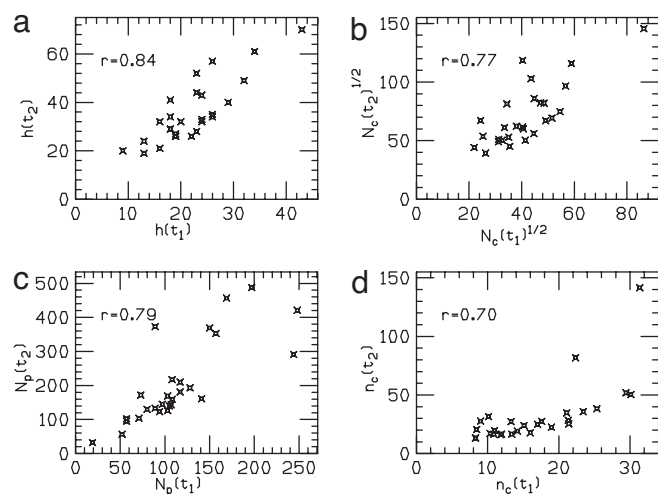
The resulting correlation coefficients as a function of  $\alpha$  are shown in Fig. 11. Surprisingly, a small *negative*  $\alpha$  ( $\alpha \approx -0.1$ ) yields the largest correlation coefficient in both samples considered. For positive  $\alpha$ , the correlation coefficients decrease monotonically and approach the values corresponding to the predictor  $N_c$ ,  $r = 0.53$  and  $r = 0.43$ , respectively, corresponding to Figs. 4b and 7b.

Consequently, the best predictor of future achievement (with achievement defined as number of citations) inferred from our data (e.g., sample PRB80) would be a linear regression fit to  $\sqrt{N_c(t_1, t_2)}$  vs.  $h_\alpha(t_1)$  with  $\alpha = -0.1$  ( $r = 0.62$ ), leading to the paradoxical result that, given two researchers with the same  $h$



**Fig. 9.** Predictive power of each indicator at year 1994 for the number of citations to all papers published up to 2006, for sample APS95.





**Fig. 10.** Predictive power of each indicator at year 1994 for the value of the same indicator at year 2006, for sample APS95.

index, the one with *lower*  $N_c(t_1)$  should be expected to earn a *higher* number of citations in the subsequent time period.

By using the relationship  $N_c = ah^2$ , we can rewrite Eq. 1 as

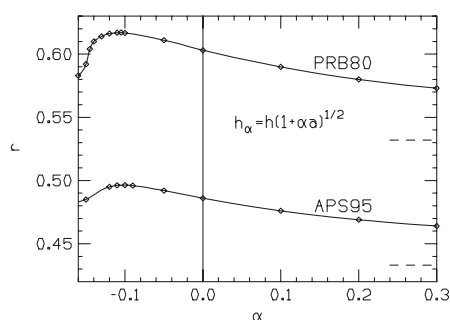
$$h_{\alpha} = h \sqrt{1 + \alpha a} . \quad [2]$$

The fact that a negative  $\alpha$  yields larger predictive power indicates that authors with large values of  $a = N_c/h^2$  are, on average, less likely to earn a larger number of citations in future work than authors with smaller  $a$ . We believe that this effect is principally due to the effect of coauthorship, as discussed in the next section.

## Discussion

In summary, we found that the  $h$  index appears to be better able to predict future achievement than the other three indicators—number of citations, number of papers, and mean citations per paper—with achievement defined by either the indicator itself or the total citation count,  $N_c$ . In addition, the  $h$  index was found to be a better predictor of productivity ( $N_p$ ) than  $N_p$  itself. Furthermore, in attempting to combine  $h$  with  $N_c$  to enhance the predictive power of  $h$ , we found that  $N_c$  should enter with a *negative* weight.

It is interesting, and not obvious, that the  $h$  index is able to predict both itself and the productivity  $N_p$  better than  $N_p$  can predict itself. Perhaps it indicates that some of the prolific authors with small citation counts feel less incentive to continue being prolific because they perceive that their work is not having an impact.



**Fig. 11.** Correlation coefficient,  $r$ , between  $N_c(t_1, t_2)$  and  $h_\alpha$  defined in Eq. 2, for samples PRB80 and APS95. As  $\alpha$  increases, the curves approach the asymptotic values given by the dashed lines,  $r = 0.53$  and  $r = 0.43$ , respectively.

We believe the superiority of  $h$  compared with  $N_c$  as a predictor is due to the issue of coauthorship, touched on in ref. 1 and in the Introduction. Let us elaborate on this further.

Consider a paper  $j$  with  $N_j^c$  citations coauthored by several scientists with different levels of seniority and ability, each of whom made different contributions to this paper. If we are counting citations, each coauthor gets the same “credit,” i.e., adds  $N_j^c$  citations to his/her total citation count, independent of his/her individual contribution to this paper.

Instead, if we are considering  $h$ , this paper will or will not contribute to the  $i$ th author's  $h$  index,  $h_i$ , depending on whether  $N_c^j > h_i$  or  $N_c^j < h_i$ . If it contributes, that author only "needs"  $h_i$  of the  $N_c^j$  citations to increase his/her  $h$  by 1. So one may say that each author  $i$  gets "allocated" only  $h_i$  of the  $N_c^j$  citations. Both junior and less able authors are likely to have a lower  $h$  than senior and more able authors, *and* they are likely to have made a lesser contribution to the paper.<sup>8</sup> Hence, it is appropriate that they benefit from a smaller portion of the total  $N_c^j$ .

In other words, to “first order,” using  $h$  rather than  $N_c$  as a measure of scientific achievement automatically reduces an important source of distortion when multiply coauthored papers are involved, by allocating a smaller portion of the credit to those authors who are likely to have contributed less. The argument is not foolproof, and exceptions undoubtedly will occur, but the “injustice” done to powerful junior coauthors in the early stages of their careers will automatically be remedied in due time as their  $h$  index rapidly increases.

Furthermore, it is interesting and revealing that the advantage of  $h$  over  $N_c$  in predicting future  $N_c$  values is lost when *cumulative* citations, rather than only citations to new papers, are considered (Figs. 2 vs. 4 and Figs. 7 vs. 9). We suggest that this also reflects the effect of coauthorship. Highly cited papers in the initial period will usually continue to garner a high number of citations in the subsequent period also for those among the coauthors who made only minor contributions to the paper. Although the cumulative citations of those individuals will be high, they should be less likely to make major *new* contributions in the subsequent period. Thus we argue that, even for a decision focused on optimizing expected *cumulative* achievement,  $h$  should be favored as an indicator because it appears better able to predict *individual* cumulative achievement.

Other recently proposed bibliometric measures that give more weight to very highly cited papers, such as Egghe's  $g$  index (4), Jin *et al.*'s AR index (5), and Komulski's  $H^2$  index (6), are likely to suffer from the same drawback as  $N_c$  because they will assign the citations of highly cited papers equally to all coauthors without discrimination. Thus, we conjecture that the predictive power of these modified indices is likely to be worse than that of the  $h$  index, as our analysis of the  $h_o$  index (Eq. 2) also suggests.

With respect to the indicator  $n_c$ , mean number of citations per paper, our results indicate that it has very little predictive value. The low correlation found between  $n_s$  in the different timeframes (initial 12 years and subsequent 12 years) is due to a variety of reasons. In some cases, the individual's productivity,  $N_p$ , remained similar but the total impact,  $N_c$ , changed substantially; sometimes productivity declined and the total impact declined even more; and sometimes productivity increased and the mean impact per paper also increased.

These results are at odds with the conclusions of the recent studies by Lehmann *et al.* (2, 3). They start from the reasonable assumption that “the quality of a scientist is a function of his or her full citation record” and address the question of which

<sup>5</sup>Of course, it will often be the case that a junior coauthor will have performed most of the actual work for the paper. Nevertheless, if the paper has senior coauthors and ended up with a large number of citations, it will often be the case that the senior coauthor(s) will have played the crucial role.

single-number indicator is best to discriminate between scientists, aiming “to assign some measure of quality.” They argue that an indicator is of no practical use unless “the uncertainty in assigning it to individual scientists is small.” They perform a Bayesian analysis of citation data from a large sample extracted from the SPIRES database ([www.slac.stanford.edu/spires/hep/](http://www.slac.stanford.edu/spires/hep/)) and find that, among the three indicators (i) mean number of citations per paper,  $n_c$ , (ii) number of papers published per year,  $N_p/n$ , and (iii)  $h$  index,  $n_c$  is far superior in discriminating between scientists. They conclude that “*compared with the  $h$  index, the mean number of citations per paper is a superior indicator of scientific quality, in terms of both accuracy and precision*,” and hence that “the mean or median citation counts (per paper) can be a useful factor in the academic appointment process.”

Bornmann and Daniel (7) echo their conclusions and state that the Lehmann *et al.* (2, 3) study “raises some doubt as to the accuracy of the ( $h$ ) index for measuring scientific performance,” that “the mean, median, and maximum numbers of citations are reliable and permit accurate measures of scientific performance,” and, instead, that “the  $h$  index is shown to lack the necessary accuracy and precision to be useful.”

We argue that these conclusions are deeply flawed. Our results here have shown that the  $h$  index is a far better predictor of future scientific achievement than the mean number of citations per paper, and surely the same would hold for the median. For example, the correlation coefficient between the number of citations in the subsequent 12 years and the  $h$  index in the previous 12 years in sample PRB80 was found to be  $r = 0.60$ , much larger than  $r = 0.21$ , the correlation found with the mean number of citations per paper in the previous 12 years. The  $h$  index was also far superior at predicting itself,  $r = 0.61$  vs.  $r = 0.23$  for  $n_c$ . A similar pattern was found in our other sample, and

it is likely that similar results would be obtained with the sample used by Lehmann *et al.* (2, 3).

This example illustrates the danger in using sophisticated mathematical analysis to jump to practical conclusions (of sometimes life-changing consequences) in the delicate issues under consideration. Although the Lehmann *et al.* (2, 3) study may be correct in concluding that the mean number of citations is better to “discriminate” between scientists for a given fixed time period according to their definition, the fallacy in their argument appears to be that this does not imply that the indicator is associated with an identifiable individual trait that would be expected to persist with time and certainly not with “scientific quality.” Else, one is forced to conclude, in light of the results of the present article, that scientific quality (as defined by Lehmann *et al.*) in the past is nearly uncorrelated with scientific quality in the future for individual scientists, a conclusion that defies common sense.

Instead, a variety of studies [refs. 1, 8 (including an extensive list of references), and 9] have shown that the  $h$  index by and large agrees with other objective and subjective measures of scientific quality in a variety of different disciplines (10–15), and the present study shows that the  $h$  index is also effective in discriminating among scientists who will perform well and less well in the future. We conclude tentatively (assuming that future empirical studies will corroborate the results of this article) that the  $h$  index is a useful indicator of scientific quality that can be profitably used (together with other criteria) to assist in academic appointment processes and to allocate research resources.

I thank Marie McVeigh for helpful advice on extracting information from the ISI Web of Science database, P. Ball for calling ref. 2 to my attention, and M.C. for stimulating discussions.

1. Hirsch JE (2005) *Proc Natl Acad Sci USA* 102:16569–16572.
2. Lehmann S, Jackson AD, Lautrup BE (2007) <http://arxiv.org/abs/physics/0701311>.
3. Lehmann S, Jackson AD, Lautrup BE (2006) *Nature* 444:1003–1004.
4. Egghe L (2006) *Scientometrics* 69:131–152.
5. Jin B-H, Liang L-M, Rousseau R, Egghe L (2007) *Chin Sci Bull* 52:855–863.
6. Kosmulski M (2006) *ISSI Newslett* 2(3):4.
7. Bornmann L, Daniel HD (2007) *J Am Soc Inf Sci Technol* 58:1381–1385.
8. Bornmann L, Daniel HD (2005) *Scientometrics* 65:391–392.

9. van Raan AFJ (2006) *Scientometrics* 67:491–502.
10. Kelly CD, Jennions MD (2006) *Trends Ecol Evol* 21:167–170.
11. Jeang KT (2007) *Retrovirology* 4:42.
12. Cronin B, Meho LI (2006) *J Am Soc Inf Sci Technol* 57:1275–1278.
13. Iglesias JE, Pecharroman C (2006) [arXiv:physics/0607224](http://arxiv.org/abs/physics/0607224).
14. Oppenheim C (2007) *J Am Soc Inf Sci Technol* 58:297–301.
15. Van Noorden R (2007) *Chemistry World* 4(5). Available at [www.rsc.org/chemistryworld/Issues/2007/May/index.asp](http://www.rsc.org/chemistryworld/Issues/2007/May/index.asp).