

Field dependence of citation data validation of premier peer-review decisions in India

Gangan Prathap*

Research evaluation based on citation data identifies excellence using quantity and quality proxies as the main orthogonal dimensions. Citation data are readily available from aggregators such as the Web of Science or Scopus in the form of number of papers P , number of citations C and h -index. However, most of the prestigious award-giving bodies, e.g. for the Nobel Prizes or for the Shanti Swarup Bhatnagar (SSB) Prize in India, rely on rigorous peer evaluation rather than on citation-based performance. While peer-review judgements are holistic but also subjective, citation-based evaluations tend to be reductive and quantitative. In this study, we look at three of the most prestigious premier peer-review decision-making exercises in India, the SSB Prize, and election Fellowship of the Indian National Science Academy, New Delhi and the Indian Academy of Sciences, Bengaluru. By comparing the awardees and successfully elected Fellows against a list of outstanding citation-based top performers in the world, we report a marked field dependence with citation data predicting 0% to less than 40% of the peer-review selections depending on the field. This study highlights two mixed dangers: one of compressing performance to numbers in an essentially reductive process, and the other of the possible bias and prejudices involved in what is often a subjective peer-review process. Thus, type I and type II errors cannot be avoided in any peer-review or citation evaluation taken in isolation, and vary from as much as 100% to 63%, depending on the field.

Keywords: Bibliometrics, citations data, peer-review evaluation, field dependence, publications.

THE Shanti Swarup Bhatnagar Prize for Science and Technology (SSB Prize) is arguably the most prestigious multidisciplinary science award in India. It recognizes excellence in scientific research in India in seven major categories, and the selection is regarded as the most rigorous peer-review exercise in the country. Anecdotal evidence suggests that the committees use citation data, but this does not play a significant role in arriving at the final selection of the prize winners. Since the institution of the SSB Prize in 1958, about 535 awards have been given and a full list can be found at: <http://ssbprize.gov.in/Content/AwardeeList.aspx>, with additional information available at https://en.wikipedia.org/wiki/List_of_Shanti_Swarup_Bhatnagar_Prize_recipients.

The SSB Prize can only be awarded to a ‘citizen of India engaged in research in any field of science and technology up to the age of 45 years’, but this restriction does not apply to equally prestigious peer-review decision making exercises in India: election to the Fellowship of the Indian National Science Academy (INSA), New Delhi and the Indian Academy of Sciences (IASc), Ben-

galuru. The SSB Prize, along with election to INSA and IASc, are among the highest recognitions a scholar can get in the country. We therefore also study the strike rate of the peer-review committees of various sections of INSA and IASc during the 1990–1996 window for elections to these learned societies.

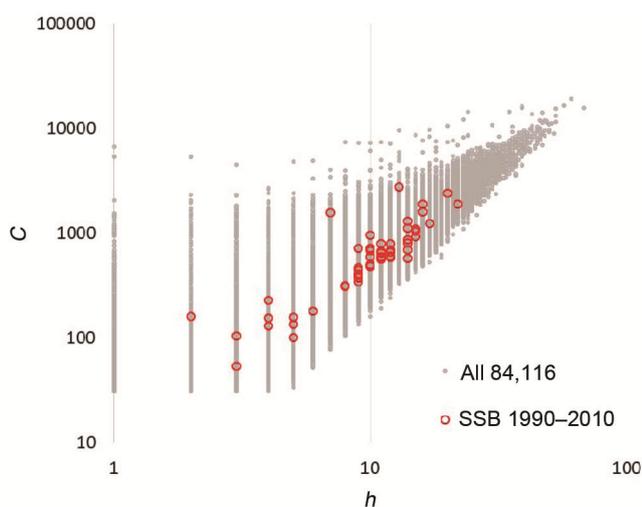
Peer-review exercises are holistic but subjective, while citation analysis is quantitative and reductive. Each has its advantages and disadvantages. The latter tries to reconcile quantity of output (usually the number of publications P in a prescribed window called the publications window) and the quality or impact of publication of an individual scientist or aggregates (e.g. institutions, journals, countries, etc.). Impact or quality is usually measured by the ratio of citations C earned by the publications during a prescribed citation window to the number of publications¹, i.e. $i = C/P$. A recent comprehensive and rigorous citation data-based effort was the appraisal of published work of scientists using citations and h -indices as the measure of total impact². These two indicators embed size and quality in a composite way combining size-independent and size-dependent parameters¹. The impact term i is interpreted as the expression of a thermodynamic relationship through the constitutive equation $C = iP$. The thermodynamic paradigm then allows

The author is in the A. P. J. Abdul Kalam Technological University, Thiruvananthapuram 695 016, India.

*e-mail: gangan_prathap@hotmail.com

Table 1. Category-wise error rate of Shanti Swarup Bhatnagar (SSB) Prize peer-review decisions from 1990 to 2010 if citation validation is used

Category	Listed in Ioannidis <i>et al.</i> ²	SSB awardees	Error
Engineering Sciences	11	30	0.63
Chemical Sciences	14	41	0.66
Physical Sciences	11	37	0.70
Medical Sciences	7	27	0.74
Total 1990–2010	52	224	0.77
Biological Sciences	5	39	0.87
Earth, Atmosphere, Ocean and Planetary Sciences	2	21	0.90
Mathematical Sciences	2	29	0.93

**Figure 1.** The 52 SSB 1990–2010 awardees on the h - C map against the background of the 84,116 scientists in the list of Ioannidis *et al.*².

one to view the two-dimensional i - P maps as phase diagrams³.

Instead of i - P maps, it is also possible to view all scientists on a h - C map by combining the h -index (a composite indicator) and citation (C) data (also a composite indicator). The data from Ioannidis *et al.*² which is available in the public domain, allow us to generate two-dimensional maps of citation performance for 84,116 leading scientists in the world. These scatter plots of h versus C allow us to use composite indicators combining quality and quantity on a two-dimensional map so that citation performance can be compared for different levels of aggregation, from scholars to journals, institutions and countries⁴.

Convergent validity

We thus have two constructs to evaluate scientific excellence. One is based on peer review and the other on ranking using citation data. Convergent validity can be established if these two similar constructs correspond with one another^{5,6}.

Many attempts have been made in the past to determine if peer-review decisions of scientific excellence can be validated using citation data^{5,6}. Bornmann and Daniel⁵ studied the convergent validity of decisions for awarding long-term fellowships (Boehringer Ingelheim Fonds (BIF)) to postdoctoral researchers using the h -index. Of the 414 BIF applicants, 64 were approved and 350 rejected. Applicants had h -indices that correlated well with standard bibliometric indicators, and on average the h -indices of approved BIF applicants were higher than those of rejected applicants, thus confirming the validity of the funding decisions. However, there were noticeable type-I (falsely drawn approval) and type-II (falsely drawn rejection) errors, with one-third of the decisions to award a fellowship to an applicant showing the former error, and another one-third of the decisions not to award a fellowship to an applicant showing the latter error. Bornmann and Daniel⁵ also showed that the field of study of an applicant can increase or decrease the risks for type-I and type-II errors.

In this study, we shall use citation data from Ioannidis *et al.*² to determine the field dependence of the validity of decision of the SSB peer-review committees, and the selection committees of INSA and IASc. Also, the danger of relying on citation data alone or on peer review alone becomes clear. These are at the same time type-I and type-II errors depending on the point of view. It is seen that the field or discipline increases the risks of type-I and type-II errors, as noticed earlier⁵.

Two-dimensional h - C maps using examples from Ioannidis *et al.*²

Ioannidis *et al.*² provides excellently curated bibliometric data for 84,116 scientists (those among the top 30,000 for impact in a single year (2013) in at least one of the six indicators) and separately across 12 scientific fields. We first take cognizance of P , the number of papers indexed by Scopus by each scientist published up to and including 2013. C is the cumulative number of citations per author in 2013. Since Scopus is reliable only for documents published since 1996, the impact calculations, i.e. based on

Table 2. Category-wise error rate of Indian National Science Academy peer-review decisions from 1990 to 1996 if citation validation is used

Section	Listed in Ioannidis <i>et al.</i> ²	Elected Fellows	Error
V-Engineering and Technology	5	18	0.72
II-Physics	5	20	0.75
III-Chemical Sciences	4	19	0.79
I-Mathematical Sciences	2	11	0.82
Total	20	220	0.86
VII-Plant Sciences	1	8	0.88
VIII-Animal Sciences	1	13	0.92
X-Cell and Biomolecular Sciences	1	16	0.94
XII-Agricultural Sciences	1	91	0.99
IV-Earth and Planetary Sciences	0	11	1.00
XI-Health Sciences	0	13	1.00

Table 3. Category-wise error rate of Indian Academy of Sciences peer-review decisions from 1990 to 1996 if citation validation is used

Section	Listed in Ioannidis <i>et al.</i> ²	Elected Fellows	Error
Chemistry	9	37	0.76
Physics	9	45	0.80
Engineering and Technology	7	37	0.81
Animal Sciences	2	11	0.82
Total	34	212	0.84
Plant Sciences	1	8	0.88
Earth and Planetary Sciences	2	18	0.89
General Biology	3	28	0.89
Mathematical Sciences	1	12	0.92
Medicine	0	16	1.00

$i = C/P$, favour those who have published or continue to publish after 1996. In the present exercise to measure citation data-based performance, we rely only on the h and C data from Ioannidis *et al.*².

We first consider the h - C maps, superimposing the cohort of SSB awardees against a background of all 84,116 scientists, where h and C data are available. We use logarithmic axes, and plot h on the x -axis and C on the y -axis, as with the same h a higher C implies a higher impact i . As the original data in the Ioannidis *et al.*² are based on the Scopus database, citation data are reliable only for the period 1996–2013, although publication data up to 2103 are used to construct the h -indices. For this reason, we shall restrict our study to only those awardees that have been recognized during the period 1990–2010 and Fellows of the Academies who were elected during the phase 1990–96. There were a total of 224 SSB awardees in seven categories (Table 1). The list of Ioannidis *et al.*² offers data in 12 major disciplines and by cross-checking each name and with disambiguation needed in a few cases, it was possible to find only 52 awardees in list. Figure 1 shows the 52 SSB awardees during 1990–2010 against the entire 84,116 scientists. Note that both h and C are composite indicators combining quality and quantity, and at the tip of the funnel, the top global performers are seen.

Thus, reliance only on citation cut-offs of the rigour used in Ioannidis *et al.*² would mean that only 23% of the awardees would have been validated by a citation data-based exercise many years after the fact. At the same time, from the citation data point of view, 77% of awardees have been falsely included (a type-I error). The ratio also changes considerably with category. The Mathematical Sciences category is most poorly served (only 7% make the citation cut; that is, there is a possible 93% error rate), and the Biological Sciences and the Earth, Atmosphere, Ocean and Planetary Sciences categories are also poorly served by reliance on citation data. However, relatively, the Engineering Sciences, Chemical Sciences, Physical Sciences and Medical Sciences respond much better to citation validation, but it is never better than a 63% error rate (for Engineering Sciences).

We repeated the exercise for the election of Fellows to two of the premier science academies in the country, namely INSA and IASc. We choose the window 1990–1996 so that there was sufficient time (effectively the window 1996–2013) for the elected incumbents to register their academic prowess in terms of citations in the list of Ioannidis *et al.*². Tables 2 and 3 summarize the results. Out of 220 Fellows elected through the peer-review committees of INSA during 1990–1996, only 20 were listed in the list of Ioannidis *et al.*². Again, out of 212

Fellows short-listed by the selection committees of IASc during 1990–96, only 34 were present in the list. Marginally, the strike rate in the SSB exercise is better compared to the efforts of the science academies to choose and recognize top scientists of the country. When it comes to field dependence, both INSA and IASc committees in Physics, Engineering Science and Technology and Chemistry were better than the average while the Biological, Life Sciences and Medical Sciences performed poorly compared to the average.

Concluding remarks

Most award-granting bodies, e.g. for the Nobel Prizes, or for the SSB Prize in India, rely more on the expert judgement of peer reviewers rather than blind reliance on quantitative data which are usually citation-based. There are inherent dangers in both, as type-I and type-II errors are inherited in such exercises⁵.

In this study, we use citation data from a well-curated exercise² to determine how the errors operate if the review decisions are validated using citation data-based rankings. Depending on the point of view (type-I from the citation data point of view and type-II from the peer-review point of view), errors vary across disciplines – from 100% for the Earth and Planetary Sciences and Health Sciences sections of INSA to 63% for the award in Engineering Sciences and Technology of the SSB Prize.

Some caveats are in order. This exercise is not undertaken to compare the *h*-indices and total citation counts of SSB awardees and elected Fellows of the science academies to those in a much larger sample of scientists chosen

globally. Rather, the goal is to determine how many of those who have successfully navigated the peer-review process in India also go on to register high citation and *h*-index performance in order to appear in the list of Ioannidis *et al.*² much after the fact. While the peer-review process is a holistic and subjective enterprise, citation-based metrics are quantitative and reductive but not necessarily more objective. Depending on whether we take the peer-review choices as the underlying ground truth, or conversely, citation-based choices as the ground truth, what is inevitable is that there are large type-I and type-II errors⁵.

1. Prathap, G., Eugene Garfield: from the metrics of science to the science of metrics. *Scientometrics*, 2018, **114**(2), 637–650.
2. Ioannidis, J. P. A., Klavans, R. and Boyack, K. W., Multiple citation indicators and their composite across scientific disciplines. *PLoS Biol.*, 2016, **14**(7), e1002501; doi:10.1371/journal.pbio.1002501.
3. Kaur, J., Ferrara, E., Menczer, F., Flammini, A. and Radicchi, F., Quality versus quantity in scientific impact. *J. Informetr.*, 2015, **9**, 800–808.
4. Crespo, J. A., Ortuno-Ortí, I. and Ruiz-Castillo, J., The citation merit of scientific publications. *PLoS ONE*, 2012, **7**(11), e49156.
5. Bornmann, L. and Daniel, H. D., Convergent validation of peer-review decisions using the *h*-index: extent of and reasons for type I and type II errors. *J. Informetr.*, 2007, **1**, 204–213.
6. Bornmann, L., Tekles, A. and Leydesdorff, L., How well does I3 perform for impact measurement compared to other bibliometric indicators? The convergent validity of several (field-normalized) indicators. *Scientometrics*, 2019; <https://doi.org/10.1007/s11192-019-03071-6>.

Received 25 May 2019; accepted 14 December 2019

doi: 10.18520/cs/v118/i8/1169-1172