Journal of Informetrics 7 (2013) 318–324

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

A bibliometric analysis of academic publication and NIH funding

Jiansheng Yang[a,b], Michael W. Vannier[c], Fang Wang[d], Yan Deng[d], Fengrong Ou[d], James Bennett[a], Yang Liu[d,**], Ge Wang[a,*]

[a] VT-WFU School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, USA
[b] School of Mathematical Sciences, Peking University, Beijing, China
[c] Department of Radiology, University of Chicago, Chicago, USA
[d] School of Public Health, China Medical University, Shenyang, China

ARTICLE INFO

Article history:
Received 19 September 2012
Received in revised form 24 October 2012
Accepted 2 November 2012
Available online 29 January 2013

Keywords:
Bibliometric analysis
Citation analysis
h-Index
a-Index
Research funding
Race/ethnicity

ABSTRACT

Academic productivity and research funding have been hot topics in biomedical research. While publications and their citations are popular indicators of academic productivity, there has been no rigorous way to quantify co-authors' relative contributions. This has seriously compromised quantitative studies on the relationship between academic productivity and research funding. Here we apply an axiomatic approach and associated bibliometric measures to revisit a recent study by Ginther et al. (Ginther et al., 2011a,b) in which the probability of receiving a U.S. National Institutes of Health (NIH) R01 award was analyzed with respect to the applicant's race/ethnicity. Our results provide new insight and suggest that there is no significant racial bias in the NIH review process, in contrast to the conclusion from the study by D. K. Ginther et al. Our axiomatic approach has a potential to be widely used for scientific assessment and management.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

In a recent study (D. K. Ginther et al.: "Race, ethnicity, and NIH research awards," Science, 19 August 2011, p. 1015) (Ginther et al., 2011a,b), the probability of receiving a U.S. National Institutes of Health (NIH) R01 award was related to the applicant's race/ethnicity. The paper indicated that black/African-American applicants were 10% less likely than white peers to receive an award after control for background and qualification, and suggested "*leverage points for policy intervention*" (Ginther et al., 2011a,b). These findings have generated a widespread debate regarding the unfairness of the NIH grant review process and its correction. The moral imperative is clear that any racial bias is not to be tolerated, particularly in the NIH funding process. However, the question of whether such a racial bias truly exists requires rigorous and systematic evaluation.

NIH director Francis Collins and Deputy Director Lawrence Tabak reiterated that the Ginther study revealed "*from 2000 to 2006, black grant applicants were significantly less likely to receive NIH research funding than were white applicants. The gap in success rates amounted to 10 percentage points, even after controlling for education, country of origin, training, employer characteristics, previous research awards, and publication record. Their analysis also showed a gap of 4.2 percentage points for Asians; however, the differences between Asian and white award probabilities were explained by exclusion of noncitizens from the analysis*" (Tabak & Collins, 2011). NIH officials admitted "*the gap could also result from 'insidious' bias favoring whites in a peer-review system that supposedly ranks applications only on scientific merit*" (Kaiser, 2011).

* Corresponding author. Tel.: +1 540 998 9198.
** Corresponding author. Tel.: +86 133 8688 5612.
E-mail addresses: jsyang.rena@gmail.com (J. Yang), mwvannier@gmail.com (M.W. Vannier), wangfang-libo@163.com (F. Wang), dengyan6028@163.com (Y. Deng), ofrcmu@163.com (F. Ou), jrbennet@vt.edu (J. Bennett), cmuliuyang@yahoo.com (Y. Liu), ge-wang@ieee.org (G. Wang).

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2012.11.006

In a Letter to Editor of Science, Voss expressed uneasiness about proposals which address implications of the Ginther study (Voss, 2011). He warned that "*disparity-reduction policies represent social experiments with tremendously important consequences, the effects of which could take decades to identify. . .much of the racial disparity reported could be attributed to black R01 applicants having half the citation count and one-fifth as many last-authored publications as white applicants from similarly ranked institutions. Coupled with the finding that R01s were awarded to highly ranked applications irrespective of race, this suggests that R01 disparity is due to lower research success among black applicants rather than to any problems with NIH review*" (Voss, 2011). In another Letter to Editor, Erickson pointed out that the citation analysis defined in the Ginther study was not relevant to competitive scientists, the number of citations under consideration should be about 1000, instead of being about 84, and the number of citations should be normalized to the career length. The opinion was expressed that similarly qualified scientists "*would be equally successful in grant funding, with no disparity for race and ethnicity*" (Erickson, 2011).

D.K. Ginther et al. wrote a defensive response to these letters. They disagree with Voss about his explanation, because "*there is substantial evidence that affirmative action does not explain the results*" (Ginther et al., 2011a,b). They found that "*blacks and whites were equally likely to receive tenure at higher education institutions that are research intensive*", and "*a bad match for research careers will have most likely been weeded out earlier*". "*There is a case to be made for positive selection of black scientists – that they are the best of the best – as opposed to being bad matches resulting from affirmative action*." Also, they disagree with Erickson about the citation issue, because their data included about 300 early-career individuals who had ∼1000 citations, being in the top 1% of the pool. Furthermore, a recent evaluation of the NIH K program (http://grants.nih.gov/training/K_Awards_Evaluation_FinalReport_20110901.pdf) showed that awardees published about 10 papers in the 5 years after the award and attracted about 150 citations per person. Furthermore, they did not think that age-normalizing citations would change their results for early-career investigators.

Sherley commented on the view from Tabak and Collins (Tabak & Collins, 2011), "*the limited public discussion on the possible underlying factors has focused on the NIH review process. Although this is an obvious place to continue the investigation, the explanation may lie elsewhere*" (Sherley, 2011). "*Barriers at the home institution*" were mentioned for "*minority investigators pursuing primarily cancer health disparities research*". For example, "*although NIH requires the writing of minority recruitment plans by its grantee institutions, it currently neither evaluates how nor even whether such plans are implemented*." Collins and Tabak did not agree with Sherley, "*the plans on all NRSA training grants are rigorously reviewed, and if they are deficient, the grants are not funded until corrective action is taken on the part of the grantee. Awarded training grants that are subsequently submitted for renewal are reviewed for the recruitment plan's results. If the plans are judged ineffective, this assessment affects its likelihood of being funded again*" (Collins & Tabak, 2011).

Good and bad biases in grant review processes have been extensively studied over past decades. An excellent review and a long list of references can be found in (Bornmann, 2011). Up to now, at least 25 sources of bias have been identified that may compromise peer-review outcomes. Although many studies have shown an association between various biases and peer-review results, general interpretations remain indefinite in many cases. An interesting phenominon is that existing studies have often yielded heterogeneous conclusions, due to disparate definitions of what are under investigation, different research designs and statistical procedures (Bornmann, 2011). With the meta-analysis approach, Bornmann and colleagues are trying to arrive at generalized statements from heterogeneous data on the effect of potential biases (Bornmann, 2011), and much more efforts are clearly yet to be made. In this context, our study is in contrast to the recent study by D. K. Ginther et al. (Ginther et al., 2011a,b), and contributes to make the issue open on whether or not the probability of receiving a U.S. National Institutes of Health (NIH) R01 award is related to the applicant's race/ethnicity.

In this paper, we are motivated to revisit the Ginther study using an axiomatic approach and paired statistical analysis. Given the complexity of this problem, we report our initial study that has already generated very interesting information. In the following sections, we describe our method, human subjects, data processing, and key results.

2. Axiomatic method

The number of publications and the number of co-authors have rapidly increased over the past decades (Greene, 2007), and the competition for academic resources has intensified with the current budgetary limitation. To manage scientific activities in general, and optimize the resource allocation in particular, individualized assessment of research results is being actively studied (Anonymous, 2007; Ball, 2008; Campbell, 1999; Foulkes & Neylon, 1996; Hirsch, 2005; Hirsch, 2007; Zhang, 2009). However, the current indices, such as the number of papers, the number of citations, the h-factor and its variants (Hirsch, 2005; Hirsch, 2007) have limitations, especially their inability to quantify co-authors' credit shares (Bornmann & Daniel, 2009; Williamson, 2009).

In the *h*-index system, a scholar having an *h* value means that he/she has published *h* papers each of which has been cited at least *h* times and his/her other papers are cited less than *h* times. Hence, the *h*-index measures both productivity and impact of his/her work (Hirsch, 2005; Hirsch, 2007). While the idea is insightful and increasingly used (Ball, 2005; Ball, 2007; Kinney, 2007; Pilc, 2008; Sebire, 2008), the *h*-index is quite volatile (Dodson, 2009) and subject to various biases (Baldock et al., 2009; Bornmann & Daniel, 2009; Engqvist & Frommen, 2008; Jeang, 2007; Kelly & Jennions, 2007; Lehmann et al., 2006; Mishra, 2008; Radicchi et al., 2008; Todd & Ladle, 2008; Wendl, 2007). A major obstacle to significant improvement of the *h*-index and other popular indices of this type has been the lack of assessment of co-authors' individual contributions. It is well recognized that the quantification of individual co-authors' credits in a publication is extremely important (Anonymous, 2007; Ball, 2008; Campbell, 1999; Foulkes & Neylon, 1996; Zhang, 2009). Current perception of a researcher's qualification relies to a great degree on either inflated or fractional counting methods (Hagen, 2008). While the former method gives the full credit to any co-author (for example, it is only stated in a biography how many papers one has published), the latter method distributes an equally divided credit to each co-author (as in some bibliometric analyses). Neither of these methods is ideal, because the order or rank of co-authors and the corresponding authorship are almost exclusively used to indicate the relative contributions of co-authors. Except for special arrangements such as specification of corresponding authors, the further down the list of co-authors for a publication, the less credit he or she deserves. Quite commonly in the biomedical research settings, the first author and the corresponding author are considered the most prominent.

To quantify co-authors' relative contributions, there are various methods available (Abbasi et al., 2011; Bellis, 2009; Egghe et al., 2012; Liu & Fang, 2012; Rehn et al., 2007; Vinkler, 2010). For example, the harmonic counting method was proposed (Hagen, 2008) to avoid the equal-share bias of the fractional counting method. A more heuristic variant was also suggested (Zhang, 2009). While the harmonic counting method has a flexibility permitting equal rankings for subsets of co-authors, it has a major limitation directly related to the fairness of credit-sharing. Without loss of generality, let us assume that the order of co-authors' names is consistent with their credit ranking, and that there are totally n co-authors on a publication whose shares are presented as a vector $\vec{x} = (x_1, x_2, \cdots, x_n)$ $(1 \le i \le n)$. Then, the k-th author's harmonic credit x_k is defined as

$$x_k = \alpha \frac{1}{k}, \ \ where\ \alpha = \frac{1}{\sum_{j=1}^{n} \frac{1}{j}}, 1 \le k \le n \tag{1}$$

Evidently, there is no rationale behind the proportionality that the k-th author contributes 1/k as much as the first author's contribution. Realistically, there are many possible ratios between the k-th and the first authors' credits, which may be equal or may be rather small such as in the cases of data sharing or technical assistance. Despite its superiority to the fractional method, the harmonic method has not been practically used, because of its subjective nature.

In order to present a rigorous solution for academic credit sharing, we present the following axiomatic system (Wang & Yang, 2010). Let us assume that each publication has n co-authors in m subsets ($n \ge m$) where co-authors in the i-th subset have the same credit x_i in $x=(x_1, x_2 \ldots, x_m)$ $(1 \le i \le m)$. The axiomatic system consists of the following three postulations:

Axiom 1 (Ranking Preference): $x_1 \ge x_2 \ge \ldots \ge x_m \ge 0$;

Axiom 2 (Credit Normalization): $c_1 x_1 + c_2 x_2 + \ldots + c_m x_m = 1$;

Axiom 3 (Maximum Entropy): x is uniformly distributed in the domain defined by Axioms 1 and 2.

The first two axioms are self-evident. The last axiom asserts that all the cases permitted by Axioms 1 and 2 are equally likely by the maximum entropy principle, since there is no basis for assuming otherwise. Therefore, the fairest estimation of co-authors' credits must be the expectation of all possible credit vectors. In other words, the k-th co-author's credit must be the elemental mean, which is referred to as the a-index for its axiomatic foundation. Please see the closed form formulas and their derivations in (Wang & Yang, 2010).

Naturally, three individualized scientific productivity measures can be defined based on our a-index. First, the productivity measure in terms of journal reputation, or the *Pr*-index, is the sum of the journal impact factors (IF) of one's papers weighted by his/her a-indices respectively. Second, the productivity measure in terms of peers' citations, or the *Pc*-index, is the sum of the numbers of citations to one's papers weighted by his/her a-indices respectively. While the *Pr*-index is useful for immediate productivity measurement, the *Pc*-index is retrospective and generally more relevant. Finally, the *Pc*IF* index is the sum of the numbers of citations after being individually weighted by both the a-index and the journal impact factor. When papers are cited, the Pc*IF index credits high-impact journal papers more than low-impact counterparts, as higher-impact papers generally carry higher relevance or offer stronger support to a citing paper.

3. Human subjects

This study targeted the top 92 American medical schools ranked in the 2011 US News and World Report, from which 31 odd-number-ranked schools were selected for paired analysis (schools were excluded if they did not provide online faculty photos or did not allow 1:2 pairing of black versus white faculty members). Data were gathered from September 1 to 5, 2011 on black and white faculty members in departments of internal medicine, surgery, and basic sciences in the 31 selected schools. White and black/African American faculty members were confirmed by their photos, names, and resumes as needed, and department heads/chairs were excluded. These schools were categorized into three tiers according to their ranking: 1st-31st as the first tier, 33rd-61st as the second tier, and 63rd-91st as the third tier. After 130 black faculty members were found from these schools, 40 black faculty members were randomly selected. The selected 40 black faculty members were 1:2 paired with white peers, yielding 120 samples as our first pool. The pairing criteria include the same gender, degree, title, specialty, and university. The ratio of 1:2 was chosen to represent white faculty members better, since the number of white faculty members is much more than that of black faculty members. Any additional major constraint, such as the number of papers, would prevent us from having a sufficient number of pairs.

Among the 130 black samples in the initial list, 14 faculty members were funded by NIH during the period from 2008 to 2011. Two of 14 black samples were excluded because of failure in matching with any white faculty member. Furthermore, an additional black faculty member was excluded because he only published at conference without any Science Citation Index (SCI) record in this period (http://sub3.webofknowledge.com). This zero productivity cannot be used as the denominator for our bibliometric analyses (see the tables below). Note that this exclusion is actually in favor of drawing a conclusion more favorable to support the conclusion from the study by D. K. Ginther et al. (Ginther et al., 2011a,b), and yet as shown below our conclusion is different from that by D. K. Ginther et al. Consequently, 11 funded black faculty members were kept. Among them, 10 were from the first tier, and 1 from the second tier. These 11 funded black faculty members were 1:1 paired with white samples that both met the pairing criteria and were funded by NIH in the same period. Consequently, there were 11 pairs of black and white investigators, which is our second pool.

4. Data processing

Using the Web of Knowledge (http://sub3.webofknowledge.com), datasets were systematically collected for the two pools of faculty members. In our study, the funding and publication records were produced to cover the period from January 2008 to August 2011. Each dataset corresponded to a single black-white combination, and included bibliographic information, such as co-authors, assignment of the corresponding author(s), journal impact factors, and citations received from 2008 to 2011. The journal impact factors were obtained from Journal Citation Reports (http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports).

The a-index values were computed using our axiomatically based formula (Wang & Yang, 2010). In computing a-index values, the first author(s) and the corresponding author(s) were treated with equal weights in this context. For the NIH-funded samples, individual numbers of funded proposals and individual funding totals were found via the NIH Reporter system (http://projectreporter.nih.gov/reporter.cfm).

Features of interest included the number of journal papers, number of citations, Pr-, Pc-, and Pc*IF-indices. For the second pool samples, additional features were numbers of NIH funded proposals and NIH funding totals per person and per racial group, respectively.

The paired t-tests were performed using SPSS 13.0 on the datasets from the first and second pools. In the first pool, the average data of two white professors were paired to individual data of the corresponding black professor. The tests were specifically performed by professional rank and school reputation, gender and integrated for racial groups.

5. Results and discussions

The scientific productivity was evaluated using the Pr-, Pc-, and Pc*IF-indices. Statistical significance levels are indicated by "*" for p<0.05 and "**" for $p < 0.01$. Table 1 suggests that higher scientific productivity was positively correlated with more senior professional titles or more prestigious institutional tiers. Furthermore, the analysis shows the male investigators were statistically more productive than the female colleagues, and the black faculty members statistically less productive than the white colleagues. The distribution of professional titles (Full, Associate, and Assistant Professor) for the black faculty members was 3:12:25, indicating an imbalance in the higher ranks. Despite that more than a half of the black samples were from first tier institutions, 14 were assistant professors. Thus, the numbers of black associate and full professors were insufficient for us to devise title-specific conclusions with statistical significance.

Table 2 focuses on the scientific productivities of the NIH funded black and white investigators, and indicates similar racial differences in scientific productivity. Although statistical significance cannot be established per professional title due

Table 1
Scientific publication measures for the black and white faculty members in the first pool.

	Race	Number of samples	Mean				
			Mean of papers	Number of citations	Pr-index	Pc-index	Pc×IF-index
Full Professor	Black	3	16.33 ± 17.24	120.67 ± 144.36	17.62 ± 23.21	33.24 ± 50.06	130.51 ± 202.80
	White	6	17.67 ± 22.87	197.83 ± 279.04	17.49 ± 19.77	20.96 ± 26.88	260.35 ± 326.53
Associate Professor	Black	12	5.83 ± 5.75	30.00 ± 37.10	4.73 ± 5.25	4.69 ± 5.35	31.32 ± 42.73
	White	24	9.08 ± 8.63	52.25 ± 55.76	5.38 ± 4.55	7.78 ± 6.04	41.23 ± 58.22
Assistant Professor	Black	25	2.44 ± 3.11**	8.88 ± 20.35*	1.71 ± 2.17**	0.86 ± 1.29*	2.87 ± 5.49*
	White	50	5.18 ± 4.86	31.94 ± 52.94	6.05 ± 6.42	7.05 ± 11.23	48.42 ± 107.01
First Tier (Groups 1–21)	Black	21	5.19 ± 8.18**	27.62 ± 63.63*	5.29 ± 9.92*	6.09 ± 19.63	29.13 ± 82.78
	White	42	10.02 ± 10.66	70.31 ± 118.28	9.22 ± 9.38	11.07 ± 14.88	87.12 ± 168.07
Second Tier (Groups 22–29)	Black	8	6.00 ± 6.28	36.50 ± 45.26	3.41 ± 3.36	4.91 ± 6.08	24.14 ± 29.35
	White	16	5.69 ± 5.32	26.44 ± 26.85	6.20 ± 5.51	6.71 ± 5.77	37.82 ± 51.48
Third Tier (Groups 30–40)	Black	11	2.09 ± 1.81	6.55 ± 8.66	1.26 ± 1.42	0.94 ± 1.38	3.12 ± 6.82
	White	22	3.23 ± 2.79	30.09 ± 53.54	2.28 ± 2.33	4.21 ± 6.10	32.22 ± 64.83
Male	Black	22	6.14 ± 7.91*	36.55 ± 65.60	4.72 ± 9.17**	6.60 ± 19.27	32.58 ± 81.54*
	White	44	9.68 ± 10.42	66.25 ± 111.14	8.79 ± 8.82	9.93 ± 11.21	75.90 ± 135.35
Female	Black	18	2.50 ± 4.16	7.78 ± 11.79	2.69 ± 4.71	1.79 ± 2.93	6.81 ± 11.68
	White	36	4.36 ± 4.50	31.19 ± 59.12	4.16 ± 5.60	6.33 ± 12.44	45.37 ± 123.49
Total	Black	40	4.50 ± 6.68**	23.60 ± 50.87*	3.81 ± 7.49**	4.44 ± 14.48	20.98 ± 61.71*
	White	80	7.29 ± 8.63	50.48 ± 92.12	6.71 ± 7.81	8.31 ± 11.77	62.16 ± 129.42
	Ratio	0.5	0.62	0.47	0.57	0.53	0.34

Table 2
Scientific publication measures for the black and white faculty members in the second pool.

Race	Number of samples	Mean				
		Number of papers	Number of citations	Pr-index	Pc-index	Pc×IF-index
Black	11	10.45 ± 9.02	88.64 ± 98.30*	11.13 ± 12.47	14.96 ± 24.11*	90.43 ± 124.94
White	11	18.64 ± 14.18	203.73 ± 189.02	18.03 ± 13.24	34.39 ± 43.82	318.42 ± 474.53
Ratio	1	0.56	0.44	0.62	0.44	0.28

Table 3
Ratios between the total funding amount and the accumulated scientific publication measurement for racial groups (not individuals) in the second pool.

Race	Number of samples	Funding total	Funding total normalized by Pr-index	Funding total normalized by Pc-index	Funding total normalized by Pc×IF-index
Black	11	20,140,082	164,565.69	122,423.76	20,247.54
White	11	43,796,537	220,860.92	115,781.91	12,503.74
Ratio	1	0.46	0.75	1.06	1.62

Table 4
Ratios between the total number of NIH-funded projects and the accumulated scientific publication measurement for racial groups (not individuals) in the second pool.

Race	Number of samples	Number of projects	Number of projects normalized by Pr-index	Number of projects normalized by Pc-index	Number of projects normalized by Pc×IF-index
Black	11	22	0.180	0.134	0.022
White	11	37	0.187	0.098	0.011
Ratio	1	0.59	0.96	1.37	2.0

to the limited numbers of samples, the differences between the racial groups are significant in terms of the number of citations and the Pc-index. In the following analysis, these scientific productivity measures will serve as the base to evaluate the fairness of the NIH funding process. Note that the racial/ethnic differences in Pr and Pc (Tables 1 and 2) are consistent with the citation analysis performed in (Ginther et al., 2011a,b).

In Tables 3 and 4, the funding support and the number of funded projects for each racial group were normalized by Pr, Pc and Pc*IF respectively. In addition to the racial difference in the R01 success rates (Ginther et al., 2011a,b), it can be seen in Tables 3 and 4 that the funding total and the number of funded projects for black NIH investigators were only 46% and 62% of that for whites, respectively. However, when these funding totals and numbers of funded projects were normalized by Pr, the ratios between black and white faculty members were narrowed. Furthermore, the normalization by the citation-oriented indices Pc and Pc*IF indicates that black faculty members had more favorable ratios from 1.06 to 2.00.

There are apparent differences in research performance by major racial groups based on individual scientific publication measures. The application of the new scientific productivity indices to the racial groups (Tables 1 and 2) clarifies the source of discrepant funding successes. When the total grant amounts and the number of funded projects were racial-group-wise normalized by these indices, the NIH review process does not appear biased against black faculty members (Tables 3 and 4).

The key results achieved statistical significance in the paired analysis that was capable of sensing differences with adequate specificity and sensitivity. There is a potential for the axiomatic approach to produce more comprehensive results with expansion of the sample sets. The construction of the databases used in this study took our 10 students' efforts over about three months, and yet is still much smaller than that used in the Ginther study (The Ginther study "*included 83,188 observations with non-missing data for the explanatory variables*" (Ginther et al., 2011a,b)). On the other hand, if detailed information were used on educational background, training, prior awards, and related variables, pairing of black and white investigators would become impossible in many cases. It is underlined that our critical abstraction has been axiomatically-formulated scientific productivity and accordingly-defined funding normalization. This perspective has allowed us to evaluate the fairness of the NIH review process in a more straightforward way, yielding statistical significance with smaller sample sizes.

The limitations of the current study are multiple, and could have made our results sub-optimal to different degrees. The research disciplines, specific institutes, other grant mechanisms (e.g., P and K awards) were not separately considered. The prior training (T and K awards), longitudinal trends, and review process changes were not analyzed. When the samples were selected, the unavailability of some faculty photos was a difficulty. Since the number of white faculty members is large, it was hoped to use more white samples for a better representation. However, the pairing criteria prevented us from including white faculty members beyond the 1:2 and 1:1 ratios for the first and second pools, respectively. The existing online searching systems do not support the computation of the axiomatic indices. The tedious data entry and analysis tasks are error-prone. Cross validation steps were performed to produce data up to a high standard. Ideally, an automated exclusive study using the axiomatic approach and paired analysis should be performed to generate the highest possible statistical confidence.

In principle, the normalization of bibliometric indicators is desirable with respect to the research field, article type, publication year, and so on (Rehn, Kronman et al., 2007). Our axiomatically based approach for credit-sharing among coauthors is also a normalization tool (Wang & Yang, 2010), and more reasonable than the harmonic counting method. The purpose of our work is to examine the recent study by D. K. Ginther et al. (Ginther et al., 2011a,b), and we did not perform any other normalization, such as with respect to the research field and so on, since D. K. Ginther et al. did not do that either. In fact, our human subjects (black and white faculty candidates) were selected from major departments (internal medicine, surgery, and basic sciences), and paired analyses were performed. Hence, the effect of differences among research fields and other factors is not considered sufficiently significant in our study. Nevertheless, we agree that this and other improvements should be made in future studies.

The axiomatic approach can be useful in multiple ways. For example, it may help streamline and monitor peer-review and research execution. Optimization of the NIH funding process has been a public concern. The NIH Grant Productivity Metrics and Peer Review Scores Online Resource stimulated hypotheses that can be tested using the axiomatic indices. For example, will new investigators be more influential than senior researchers? Will large grant mechanisms such as U01 (Research Project Cooperative Agreement) and P01 (Research Program Project Grant) be more productive than R01 (Research Project Grant) and R21 (Exploratory/Developmental Research Grant)? Will renewed projects be more cost-effective than initially funded projects? Although any bibliometric measure is subject to model mismatch and performance fluctuation, several important problems of common interest can be studied with the same individual or team as its own control. These questions and many others can be addressed using the approach proposed in this paper.

6. Conclusions

We have proposed an axiomatic approach for individualized academic evaluation in the teamwork context. Our methodology has potential to be widely used for scientific assessment and management. As an initial application, our axiomatic approach has been applied to revisit a recent high-profile study by Ginther *et al.* (Ginther et al., 2011a,b). In contrast to their Science paper, our results suggest that there is no significant racial bias in the NIH review.

Competing interests

None.

Acknowledgments

The authors thank Dr. Z. J. Lin for discussion and help with implementing a Monte-Carlo algorithm for evaluation of co-authors' credits in this axiomatic framework. The authors are grateful for valuable comments by Drs. Lizhi Sun, Scott Verbridge, Renee Prater, and others, as well as diligent data entry and analysis work performed by our research assistants especially W. Y. Wang, S. H. Tong, Y. Y. Gao, and G. C. Liu.

References

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, *5*(4), 594–607.
Anonymous. (2007). Who is accountable? *Nature*, *450*(7166), 1.
Baldock, C., Ma, R., & Orton, C. G. (2009). Point/counterpoint. The h index is the best measure of a scientist's research productivity. *Medical Physics*, *36*(4), 1043–1045.
Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, *436*(7053), 900.
Ball, P. (2007). Achievement index climbs the ranks. *Nature*, *448*(7155), 737.
Ball, P. (2008). A longer paper gathers more citations. *Nature*, *455*(7211), 274–275.
Bellis, N. D. (2009). *Bibliometrics and citation analysis – from the science citation index to cybermetrics*. Plymouth, UK: Scarecrow Press.
Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, *45*, 199–245.
Bornmann, L., & Daniel, H. D. (2009). The state of h index research. Is the h index the ideal way to measure research performance? *EMBO Reports*, *10*(1), 2–6.
Campbell, P. (1999). Policy on papers' contributors. *Nature*, *399*(6735), 393.
Collins, F. S., & Tabak, L. A. (2011). Race disparity in grants: Oversight at home response. *Science*, *334*(6058).
Dodson, M. V. (2009). Citation analysis: Maintenance of h-index and use of e-index. *Biochemical and Biophysical Research Communications*, *387*(4), 625–626.
Egghe, L., Guns, R., & Rousseau, R. (2012). Measuring co-authors' contribution to an article's visibility. *Scientometrics*.
Engqvist, L., & Frommen, J. G. (2008). The h-index and self-citations. *Trends in Ecology & Evolution*, *23*(5), 250–252.
Erickson, H. P. (2011). Race disparity in grants: Check the citations. *Science*, *334*(6058), 899.
Foulkes, W., & Neylon, N. (1996). Redefining authorship. Relative contribution should be given after each author's name. *BMJ*, *312*(7043), 1423.
Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F. Y., Haak, L. L., et al. (2011a). Race disparity in grants: Empirical solutions vital response. *Science*, *334*(6058), U154–U899.
Ginther, D. K., Schaffer, W. T., Haak, L. L., & Kington, R. (2011b). Race, ethnicity, and NIH research awards. *Science*, *333*(6045), 1015–1019.
Greene, M. (2007). The demise of the lone author. *Nature*, *450*(7173), 1165.
Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLoS ONE*, *3*(12), e4021.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.
Hirsch, J. E. (2007). Does the H index have predictive power? *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19193–19198.
http://grants.nih.gov/training/K_Awards_Evaluation_FinalReport_20110901.pdf
http://projectreporter.nih.gov/reporter.cfm
http://sub3.webofknowledge.com
http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports

Jeang, K. T. (2007). Impact factor, H index, peer comparisons, and retrovirology: Is it time to individualize citation metrics? *Retrovirology*, *4*, 42.
Kaiser, J. (2011). NIH uncovers racial disparity in grant awards. *Science*, *333*(6045), 925–926.
Kelly, C. D., & Jennions, M. D. (2007). H-index: Age and sex make it unreliable. *Nature*, *449*(7161), 403.
Kinney, A. L. (2007). National scientific facilities and their science impact on nonbiomedical research. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 17943–17947.
Lehmann, S., Jackson, A. D., & Lautrup, B. E. (2006). Measures for measures. *Nature*, *444*(7122), 1003–1004.
Liu, X. Z., & Fang, H. (2012). Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution. *Journal of Informetrics*, *6*(4), 557–565.
Mishra, D. C. (2008). Citations: Rankings weigh against developing nations. *Nature*, *451*(7176), 244.
Pilc, A. (2008). The use of citation indicators to identify and support high-quality research in Poland. *Archivum Immunologiae et Therapiae Experimentalis*, *56*(6), 381–384.
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 of the United States of America*, *105*(45), 17268–17272.
Rehn, C., Kronman, U., & Wadskog, D. (2007). Bibliometric indicators - definitions and usage at Karolinska Institutet Stickholm, Sweden, Karolinska Institutet University Library.
Sebire, N. J. (2008). H-index and impact factors: Assessing the clinical impact of researchers and specialist journals. *Ultrasound in Obstetrics and Gynecology*, *32*(7), 843–845.
Sherley, J. L. (2011). Race disparity in grants: Oversight at home. *Science*, *334*(6058)
Tabak, L. A., & Collins, F. S. (2011). Weaving a richer tapestry in biomedical science. *Science*, *333*(6045), 940–941.
Todd, P. A., & Ladle, R. J. (2008). Citations: Poor practices by authors reduce their value. *Nature*, *451*(7176), 244.
Vinkler, P. (2010). *The evaluation of research by scientometric indicators*. Oxford, UK: Chandos Publishing.
Voss, J. L. (2011). Race disparity in grants: Empirical solutions vital. *Science*, *334*(6058), 899.
Wang, G., & Yang, J. S. (2010). *Axiomatic quantification of co-authors' relative contributions.* , arXiv:1003.3362v1001 [stat.AP].
Wendl, M. C. (2007). H-index: However ranked, citations need context. *Nature*, *449*(7161), 403.
Williamson, J. R. (2009). My h-index turns 40: My midlife crisis of impact. *ACS Chemical Biology*, *4*(5), 311–313.
Zhang, C. T. (2009). A proposal for calculating weighted citations based on author rank. *EMBO Reports*, *10*(5), 416–417.