

5-2016

Performance benchmarks for scholarly metrics associated with fisheries and wildlife faculty

Robert K. Swihart
Purdue University, rswihart@purdue.edu

Mekala Sundaram
Purdue University

Tomas O. Höök
Purdue University

James Andrew DeWoody
Purdue University

Kenneth F. Kellner
Purdue University

Follow this and additional works at: <http://docs.lib.purdue.edu/fnrpubs>

 Part of the [Genetics Commons](#)

Recommended Citation

Swihart R.K., Sundaram M., Höök T.O., DeWoody J.A., Kellner K.F. Performance benchmarks for scholarly metrics associated with fisheries and wildlife faculty. PLOS ONE Volume 11, Issue 5, 1 May 2016, Article number e0155097. <http://dx.doi.org/10.1371/journal.pone.0155097>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

RESEARCH ARTICLE

Performance Benchmarks for Scholarly Metrics Associated with Fisheries and Wildlife Faculty

Robert K. Swihart^{1*}, Mekala Sundaram¹, Tomas O. Höök^{1,2}, J. Andrew DeWoody¹, Kenneth F. Kellner¹

1 Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, United States of America, **2** Illinois-Indiana Sea Grant Program, Purdue University, West Lafayette, Indiana, United States of America

* rswihart@purdue.edu



Abstract

Research productivity and impact are often considered in professional evaluations of academics, and performance metrics based on publications and citations increasingly are used in such evaluations. To promote evidence-based and informed use of these metrics, we collected publication and citation data for 437 tenure-track faculty members at 33 research-intensive universities in the United States belonging to the National Association of University Fisheries and Wildlife Programs. For each faculty member, we computed 8 commonly used performance metrics based on numbers of publications and citations, and recorded covariates including academic age (time since Ph.D.), sex, percentage of appointment devoted to research, and the sub-disciplinary research focus. Standardized deviance residuals from regression models were used to compare faculty after accounting for variation in performance due to these covariates. We also aggregated residuals to enable comparison across universities. Finally, we tested for temporal trends in citation practices to assess whether the “law of constant ratios”, used to enable comparison of performance metrics between disciplines that differ in citation and publication practices, applied to fisheries and wildlife sub-disciplines when mapped to Web of Science Journal Citation Report categories. Our regression models reduced deviance by $\frac{1}{4}$ to $\frac{1}{2}$. Standardized residuals for each faculty member, when combined across metrics as a simple average or weighted via factor analysis, produced similar results in terms of performance based on percentile rankings. Significant variation was observed in scholarly performance across universities, after accounting for the influence of covariates. In contrast to findings for other disciplines, normalized citation ratios for fisheries and wildlife sub-disciplines increased across years. Increases were comparable for all sub-disciplines except ecology. We discuss the advantages and limitations of our methods, illustrate their use when applied to new data, and suggest future improvements. Our benchmarking approach may provide a useful tool to augment detailed, qualitative assessment of performance.

OPEN ACCESS

Citation: Swihart RK, Sundaram M, Höök TO, DeWoody JA, Kellner KF (2016) Performance Benchmarks for Scholarly Metrics Associated with Fisheries and Wildlife Faculty. PLoS ONE 11(5): e0155097. doi:10.1371/journal.pone.0155097

Editor: George Tserpes, Hellenic Centre for Marine Research, GREECE

Received: January 5, 2016

Accepted: April 25, 2016

Published: May 6, 2016

Copyright: © 2016 Swihart et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Author-based data are all contained within the Supporting Information file [S1 Data](#). Institutional codes are omitted to protect the anonymity of faculty at institutions where few fisheries and wildlife faculty occur.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

University administrators seek evidence of scholarly productivity and impact when making decisions on, among other things, faculty hiring, salary increases, tenure, and promotion. Ideally, evidence is multi-faceted and includes data collected objectively from multiple sources [1] as well as insights provided by peers or client groups. Evidence of productivity and impact in teaching, for example, may include student credit hours generated, number of courses taught, student and peer evaluations, accounts of mentoring, curriculum development and innovation, funds obtained in support of teaching/learning, and publication of teaching practice or scholarship.

In research, evidence of scholarship is largely (though not exclusively) dependent on a faculty member's publication and citation record. To the extent that science operates on an "economy of reputation" [2], indexes that quantify performance in relation to publications and citations provide a convenient currency that increasingly is used to approximate the overall impact of a researcher [3]. Perhaps not surprisingly given the demand and ease with which information on publication and citation records is obtained, the number and variety of these "bibliometric" indexes has exploded in the last decade (reviewed by [3–6]). A necessary condition for effective use of such metrics is an ability to place values into an appropriate context [7]. The fact that a faculty member averages 35 citations per year is meaningful only to the extent that her performance can be compared to a reference group. Numerous studies have established performance benchmarks for particular disciplines to enable comparison to peers [7–13] or to highly regarded scientists [14–16]. Here, we provide a model-based process to benchmark performance of faculty in fisheries and wildlife disciplines.

Although the notion of benchmarking is intuitive, it can be difficult to implement. Within a discipline, two faculty members may differ in a bibliometric score for reasons that have nothing to do with innate differences in their ability or the quality of their work. For instance, the number of years over which a scientist has been active in a chosen profession is related positively or in an asymptotic fashion to indexes that rely on accumulated counts of publications or citations [17,18] and negatively to indexes expressed as rates that are influenced by temporal trends in publication and citation practices [19]. In addition, few faculty members devote all of their time to research; instead, many have responsibilities associated with teaching, extension and outreach, or service. Not surprisingly, those with more non-research related responsibilities tend to produce lower bibliometric scores than faculty with greater research appointments [20]. Sex differences in bibliometric scores also are common within a discipline; in several disciplines males tend to generate higher performance scores than females [10,15,21]. Even within a discipline, sub-disciplines can differ dramatically in publication and citation practices [22]. Despite the importance of these covariates on bibliometric scores, most benchmarking studies have failed to incorporate their effects or focused on a single covariate, typically professional age [9], but see [15].

One approach to account for such covariates is to control for them statistically: "The residuals of a regression of productivity on age and other control variables might provide a promising avenue for a tailor-made individual adjustment for every individual in the sample" [18]. We apply a standardized residual approach to regression models of bibliometric performance by fisheries and wildlife faculty that incorporate potentially influential covariates. We also consider the feasibility of comparing performance across various scientific sub-disciplines within fisheries and wildlife.

Materials and Methods

Benchmarking against peer performance

In a separate study [20], we identified 'best' models to predict publication and citation metrics for 437 faculty at 33 research-extensive universities in the United States that were members of

the National Association of University Fisheries and Wildlife Programs (NAUFWP). For each of 437 faculty members we collected bibliometric information from Web of Sciences searches and obtained information on faculty appointments by contacting institutional administrators. We determined Ph.D. date for faculty from department webpages and by searching for electronic dissertations in ProQuest database. For faculty with common last names, we refined searches as described in [20] and compared results to personal web pages and publication lists. Sub-disciplines of research were assigned after examination of faculty webpages and selected publications. Finally, we developed regression models for Hirsch's h-index and m quotient [14], the h_b -index [23], number of publications, and annual citation rate [20]. An individual's h-index is defined as the number of publications with at least h citations each [14]. Hirsch argued that two individuals with the same h-index were similar in terms of overall scientific impact, even if they differed in excess citations (or excess publications) [14]. Brown disagreed and defined the h_b index as $h_b = h + \sqrt{e}$, where e is the sum of all citations in excess of h^2 for those articles that contribute to h [23,24]. Hirsch [14] further defined the m quotient as the annual rate of increase in h over the years of scientific activity for an individual. For m quotient and annual citation rate, we equated years of scientific activity to years since conferral of the Ph.D. [20]. For the population of 437 faculty members, significant variation in some or all of these performance metrics was explained by time allocated to research (i.e., research appointment percentage), elapsed years since Ph.D. conferral (i.e., academic age), a quadratic term associated with academic age, sex, the sub-disciplines in which research was conducted (8 sub-disciplinary categories were incorporated into models as indicator variables to which a faculty member could be assigned—disease, genetics, social sciences, management, ecology, quantitative, conservation and aquatic sciences), and how early the first publication appeared in one's academic career [20].

The generalized linear models developed previously [20] form the basis for our benchmarking approach. Herein, we develop similar models for three additional performance metrics. Total citations were included as a cumulative measure of the impact of an individual's scholarship [3]. The median number of citations for publications in the h-core, i.e. the m index, was included because it has been recommended as a measure of impact that complements h [3,5]. Finally, the r index, defined as the square root of the sum of citations for papers with h or more citations [25], was computed to measure citation intensity in the h core.

Details on model development are provided elsewhere [20]. Briefly, we constructed a set of nested candidate models, with an intercept-only model followed by sequential incorporation of academic age (and a centered quadratic term if deemed appropriate after inspection of residuals), percent of appointment allocated to research, sex (0 = female, 1 = male), and sub-discipline. Unlike some previous analyses [20], we did not include the difference between the years of Ph.D. conferral and first publication as a covariate, because doing so to benchmark using standardized residuals would unjustly penalize individuals for publishing at an early career stage. Negative binomial models were fitted, after rounding to the nearest integer, for h, h_b , m, and r indexes, number of publications, number of citations, and annual citation rate, computed as the quotient of citation count and elapsed years since attainment of Ph.D. Models were fitted with function `glm.nb` from the MASS library [26] in R 3.0.2 [27], as considerable over-dispersion was evident in models fitted to a Poisson distribution. Gaussian regression models were fitted for the m quotient, which we defined as the quotient of h-index and years since Ph.D. conferral [5], after transformation of academic age to $\ln(\text{age} + 0.5)$ to reduce heteroscedasticity. We set academic age to zero for one individual who obtained the Ph.D. in 2015. Nested models were compared using likelihood-ratio tests to determine which covariates to retain, and a final model was fitted to each of the response variables for benchmarking purposes.

To benchmark against peer performance, the final models were used to compute standardized deviance residuals for each of the 437 faculty members [28], where deviance is the difference in log likelihoods between a proposed model and a saturated model with one parameter fitted for every observation in the dataset. Standardized deviance residuals weight each observation's deviance by the influence of the observation on the model [29]. The standardized residuals form the basis for comparisons among peers, as they represent deviations in performance after accounting for variation due to academic age, sex, and other factors. The sign of the residual indicates direction of deviation from expectation. Positive and negative residuals suggest better and worse performance relative to model prediction, respectively. More specifically, the standardized residual for an individual faculty member can be compared to a histogram or distribution of standardized residuals for all faculty members to assess the percentile of the distribution with which she is associated. For each model we ranked faculty in descending order of standardized residuals to generate percentile rankings. For a large enough sample size in regressions, internally standardized residuals approximate a standard normal distribution [28,29]. Thus, we also computed percentile rankings for faculty based on a standard normal curve.

When considered as an evaluation tool, use of multiple indicators is recommended to provide a more accurate depiction of performance [3,30]. Thus, we also used the standardized residuals to compute benchmarks for individual faculty by combining results across models. First, we used the percentiles of standardized residuals for the 8 metrics to compute the mean, range, and 95% confidence interval of percentile ranks for each faculty member. Equal weighting of individual metrics to compute a mean percentile is most appropriate when metrics are uncorrelated, but some of our measures exhibited strong correlations (S1 Table). Therefore, in a second approach, we reduced the dimensionality of the standardized residuals via a factor analysis with varimax rotation. The resulting factor scores were displayed in two dimensions, with convex hulls for the innermost 100, 95, 90, 75, and 50 percent of factor scores superimposed on bivariate plots. We computed a joint percentile based on a mean factor score for each faculty member that was weighted by the relative proportion of variation explained by each factor. To facilitate the use of our benchmarking approach by other individuals, including faculty not in the population for which models were developed, we provide R code to perform the modeling and computations described above, given a set of values for response variables and covariates (S1 Code).

To illustrate the use of these models as a benchmark for peers, we consider a hypothetical male associate professor of fisheries and wildlife 15 years post-Ph.D. that conducts ecologically based research on fish diseases (Table 1). We also assume that the individual develops quantitative tools associated with analyses used in disease ecology. The faculty member would thus be categorized as active in the sub-disciplines of ecology, disease, and quantitative sciences. Further, suppose that the faculty member has 25% effort allocated to research. In terms of publication and citation metrics, suppose the citation record for the faculty member, for 30 publications ranked in descending order of number of citations, is (50, 40, 30, 20, 18, 16, 15, 14, 12, 10, 9, 8, 6, 5, 5, 4, 3, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0) (Table 1). Ten publications have 10 or more citations, thus the h-index is 10. The minimum number of citations needed for an h-index of 10 is $10^2 = 100$, so the number of excess citations in the h core is $e = 225 - 100 = 125$, and $h_b = 10 + 125^{1/2} = 21$. The m quotient is the ratio of h-index and academic age, so m quotient = $10/15 = 0.67$. The annual citation rate is $270/15 = 18$. The median number of citations for papers appearing in the h-core, i.e., the m index, is $(18+16)/2 = 17$. The r index is $225^{1/2} = 15$.

Benchmarking institutions

Aggregation of individual-level performance metrics can be used to make comparisons across institutions (e.g., [18,31]). For each performance metric we ranked universities with at least

Table 1. Standardized Deviance Residuals and Benchmarking Percentiles for Hypothetical Faculty Member in Fisheries and Wildlife.

Metric	Value	Residual	Benchmarking Percentiles ^a
h-index [14]	10	-1.05	16.0
h_p-index [23]	21	-1.40	9.6
m quotient [14]	0.67	-0.96	9.1
Publications	30	-1.02	17.8
Citations	270	-1.36	12.8
Citations/year	18	-1.48	11.0
m index [3,5]	17	-1.46	8.7
r index [26]	15	-1.34	11.2
Mean/Percentile		-1.26	12.0/8.4

Residuals and percentiles computed from regression models for a hypothetical faculty member with the bibliometric values listed above and the following characteristics: male, 25% research appointment, and a sub-disciplinary research focus in ecology, disease, and quantitative methods.

^aBenchmarking percentiles represent the percentage of all faculty in the database (n = 438) with residuals as small or smaller than the residual of the hypothetical faculty. The value of 12.0% averaged across the 8 bibliometrics is as large as 8.4% of averages for other faculty.

doi:10.1371/journal.pone.0155097.t001

five faculty in fisheries and wildlife according to their mean standardized residuals. For each university we then depicted with a boxplot the median, interquartile width, and range of the mean ranks of residuals across the 8 performance metrics. We assessed the importance of university affiliation as a covariate in regression models using likelihood ratio tests.

Benchmarking across disciplines

Faculty in fisheries and wildlife are heterogeneous with respect to training and research interests, with disciplinary ties that range from genetics to ecology to social sciences. Publication and citation practices differ substantially across disciplines [30]. Hence, comparison of faculty from different disciplines requires normalization that accounts for variation in citation and publication practices. We believe that our regression approach provides an improvement over prior attempts at normalization when considering peers within a discipline (fisheries and wildlife in this case). As one comparison, we extended the ‘law of constant ratios’ [32] to disciplines with which faculty and fisheries and wildlife are aligned.

Podlubny examined the number of citations produced in each of 6 years spanning the period 1992–2001 for 9 different disciplines [32]. He found that even though the number of citations in the disciplines increased over time, the ratio of citations for any two disciplines remained stable over time. Consequently, he suggested a simple normalization, i.e., the ratio of total citations in a discipline to total citations in mathematics (chosen as a reference), to enable comparisons of scientists from different disciplines. Subsequent study has shown that normalization to a mean number of citations per article published per year works as well or better than more complicated scaling methods [33]. Thus, we collected number of citations per article per year for the fisheries and wildlife sub-disciplines examined previously [20] and as classified in the Web of Science Journal Citation Reports (JCR). Specifically, we mapped 18 JCR categories to 7 of the 8 sub-disciplines used in our benchmarking models (Table 2); no JCR categories characterized our management sub-discipline. Following convention [16,32], we used the field of mathematics as a reference and computed the ratios of number of citations per article per year for each mapped set of JCR categories relative to mathematics. We computed these ratios for every year from 2005 to 2014 and used linear regression of year (centered) to determine whether the ratio remained constant across years, with interaction terms to test whether sub-disciplines differed in temporal trends of citations/article (relative to mathematics). As per the

Table 2. Classification Used to Map Journal Citation Report® Categories onto Sub-disciplines in Fisheries and Wildlife.

Sub-disciplines	Journal Citation Report Categories
Disease	Infectious diseases, Parasitology, Environmental sciences, Nutrition & dietetics, Anatomy & morphology
Genetics	Genetics & heredity, Evolutionary biology, Biochemistry & molecular biology
Social sciences	Sociology, Social sciences interdisciplinary, Agricultural economics & policy
Ecology	Ecology
Quantitative	Statistics & probability, Remote sensing
Conservation	Biodiversity conservation
Aquatic sciences	Limnology, Fisheries, Marine & freshwater biology

doi:10.1371/journal.pone.0155097.t002

Purdue University Senior Protocol Analyst, Human Research Protection Program, this study does not qualify as human subjects research as defined by the U.S. Department for Health and Human Services and hence was not subjected to IRB review.

Results

Benchmarking against peer performance

All final models included academic age, percent of appointment allocated to research, and sub-discipline (Table 3). Number of publications was the only response variable for which sex was significant (Table 3). Models reduced deviance by 25.2–52.4% relative to intercept-only models (Table 3).

After appending vectors of covariate values and bibliometric values for the hypothetical faculty member to the corresponding vectors for the 437 faculty members in our population (S1 Data), we refit the final models to obtain coefficients that incorporate data for the hypothetical faculty member and enable computation of standardized residuals (S1 Code, S2 Data). Of the metrics, the *m* quotient was least correlated with other variables (S1 Table) and hence may merit special attention. In our example, the hypothetical individual exhibited a standardized residual of -0.96 for *m* quotient and -1.05 for Hirsch’s *h*-index (Table 1), corresponding to percentile rankings of 9.1% and 16.0%, respectively, for these two complementary variables [14]. When all 8 bibliometrics were considered, the hypothetical faculty member exhibited an average standardized residual of -1.26. His average percentile rank of 12.0 across the 8 metrics is greater than only 8.4% of peers (Table 1). A similar value, 10.4%, is obtained for the percentile of a standard normal variable with a value of -1.26. Across all faculty, percentiles obtained empirically and via the normal approximation were highly correlated ($r = 0.989$, $p < 0.0001$).

For individual faculty, standardized deviance residuals from the models of the 8 bibliometrics exhibited considerable correlation (S1 Table). A factor analysis produced two factors that explained 53.8% and 34.2% of the variation, respectively. Factor 1 was correlated most highly with the performance metrics that emphasize citation counts, whereas Factor 2 was most correlated with publication count (Table 4). The *h*-index and *m* quotient exhibited correlations that were similar between the factors (Table 4). Our hypothetical faculty member exhibited scores that indicated below-average performance for both factors relative to peers; i.e., he fell in quadrant III of the bivariate scatter plot (Fig 1). Indeed, only 12.1% of peers scored at least as low on factor 1 (score ≤ -1.07), whereas 35.8% scored at least as low on factor 2 (score ≤ -0.36). To consider the factors jointly, we computed a mean score weighted by the relative variation explained by each factor. The first factor accounted for $.538 / (.538 + .342) = .611$ of

Table 3. Fitted Models for Eight Bibliometrics of Scholarly Performance by Fisheries and Wildlife Faculty.

Covariate	Bibliometric Response Variable Coefficients ^a (Standard Error)							
	h-index ^b	h _b -index	m quotient	Publications	Citations	Citations/ year	m index	r index
Intercept	1.48 (.09)	2.35 (.09)	2.08 (.18)	2.40 (.13)	4.23 (.17)	2.60 (.16)	2.27 (.10)	1.92 (.09)
Academic age	0.04 (.002)	0.04 (.002)	-0.52 (.04)	0.05 (.003)	0.09 (.004)	0.02 (.004)	0.04 (.002)	0.04 (.002)
Age ²	-0.001 (.0002)	-0.001 (.0002)		<i>-0.001</i> (.0002)	-0.002 (.0004)		-0.001 (.0002)	-0.001 (.0001)
Research	0.01 (.001)	0.01 (.001)	0.005 (.001)	0.008 (.001)	0.01 (.002)	0.01 (.002)	0.004 (.001)	0.01 (.001)
Sex				<i>0.21</i> (.08)				
Disease	0.21 (.06)	0.22 (.06)	<i>0.18</i> (.07)	0.29 (.08)	0.43 (.12)	0.42 (.12)	<i>0.18</i> (.07)	0.21 (.06)
Genetics	0.27 (.07)	0.30 (.08)	0.30 (.09)	<i>0.26</i> (.10)	0.63 (.14)	0.61 (.14)	0.34 (.08)	0.32 (.07)
Social	-0.25 (.08)	-0.26 (.08)	<i>-0.23</i> (.09)	<i>-0.21</i> (.10)	-0.48 (.15)	-0.45 (.14)	-0.29 (.08)	<i>-0.25</i> (.08)
Management	-0.22 (.04)	-0.20 (.05)	-0.20 (.06)	<i>-0.13</i> (.06)	-0.39 (.09)	-0.38 (.09)	-0.21 (.05)	-0.23 (.05)
Ecology	0.15 (.05)	0.18 (.05)	<i>0.12</i> (.06)	<i>0.11</i> (.07)	0.33 (.10)	0.34 (.09)	0.23 (.05)	0.21 (.05)
Quantitative	<i>0.18</i> (.06)	0.24 (.06)	<i>0.16</i> (.07)	<i>0.11</i> (.08)	0.50 (.12)	0.50 (.11)	0.30 (.06)	0.28 (.06)
Conservation	<i>0.10</i> (.05)	<i>0.13</i> (.05)	<i>0.07</i> (.06)	<i>0.07</i> (.07)	<i>0.27</i> (.10)	<i>0.27</i> (.09)	0.18 (.05)	<i>0.16</i> (.05)
Aquatics	<i>0.16</i> (.09)	<i>0.18</i> (.10)	<i>0.12</i> (.12)	<i>0.08</i> (.14)	<i>0.43</i> (.20)	<i>0.39</i> (.19)	<i>0.29</i> (.11)	<i>0.22</i> (.10)
θ	9.17 (1.04)	5.40 (0.44)	na	2.84 (0.20)	1.28 (0.08)	1.47 (0.10)	5.05 (0.41)	6.21 (0.54)
Improvement in Deviance ^c	52.4%	48.4%	43.4%	44.5%	47.0%	25.2%	42.4%	49.2%

Academic age = years since conferral of Ph.D.; ln(academic age + 0.5) was used to model m quotient. Age² = square of age, after centering; Research = percentage of appointment allocated to research; Sex = 0 if female, 1 if male; θ is the over-dispersion parameter for negative binomial models (all except m quotient, which was fitted via linear regression). Disease, . . . , Aquatics are binary variables for the 8 sub-disciplines.

^aBoldface signifies p < 0.001; italics signifies p < 0.05.

^bThe predicted h-index for a faculty member of average academic age (= 18.6 years) specializing in genetics and with a 52% research appointment is computed as exp(1.48 + .04*18.6 + .01*52 + .27*1) = 20. In this example, age² is 0 since the variable age² is centered on the mean and the faculty member is assumed to be of average academic age. No exponentiation is applied to the linear regression for m quotient. Additional examples are given in [20]. However, note that the coefficients and models in [20] include publication precocity (or year of researcher’s first publication relative to PhD attainment) as a variable.

^cRelative to intercept-only model.

doi:10.1371/journal.pone.0155097.t003

Table 4. Factor Analysis on Standardized Deviance Residuals from Regression Models of Bibliometric Performance by Fisheries and Wildlife Faculty.

Bibliometric	Factor Loadings	
	Factor 1	Factor 2
h-index	0.703	0.648
h _b -index	0.779	0.598
m quotient	0.486	0.430
Publications	0.339	0.895
Citations	0.801	0.593
Citations per year	0.794	0.577
m index	0.893	0.284
r index	0.884	0.457
Proportion of variation explained	0.538	0.342

doi:10.1371/journal.pone.0155097.t004

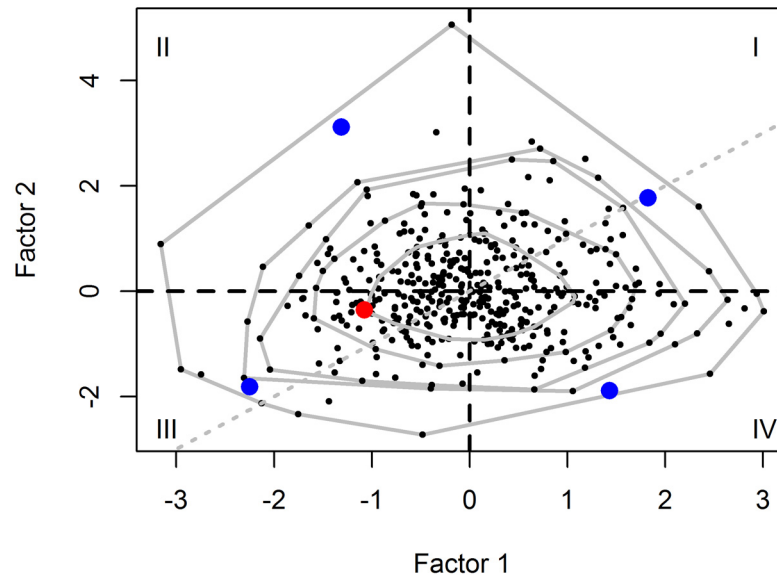


Fig 1. Benchmarking of faculty performance with factor analysis on residuals from models for 8 bibliometric variables. Factor 1 is correlated most highly with performance metrics that emphasize citation counts, whereas Factor 2 is most correlated with publication count. Convex hulls are superimposed for the innermost 100, 95, 90, 75, and 50 percent of factor scores to facilitate comparison. The dashed 45-degree line represents performance that is equally good (Quadrant I) or poor (Quadrant III) for both factors. The hypothetical faculty member (Table 1) is depicted with a red circle in Quadrant III. Representative faculty between the 95 and 100 percent hulls are highlighted in each quadrant with blue circles.

doi:10.1371/journal.pone.0155097.g001

explained variation, whereas the second factor accounted for the remaining .389. Thus, the weighted mean for the hypothetical faculty member is $-1.07(.611) + -0.36(.389) = -0.80$, yielding a ranking of 48, i.e., 11.0% of peers produced a joint weighted score at least as low. For the faculty highlighted in blue in each quadrant (Fig 1), the percentiles computed from the weighted linear combination of factor scores are 99.5 (quadrant I), 73.0 (II), 0.9 (III), and 61.6 (IV). Rankings from the two aggregative approaches, i.e., factor analysis and the unweighted means of percentiles, were highly correlated ($r = 0.985$, $p << 0.0001$).

Residuals for each bibliometric variable were computed from models in Table 3.

Benchmarking institutions

Likelihood-ratio tests indicated that university affiliation explained significant amounts of variation among institutions for all performance metrics except m quotient. When average standardized residuals were considered as ranks for the 8 performance metrics, University of California-Davis exhibited the best median ranking across the 8 metrics, followed by North Carolina State University and Colorado State University (Fig 2). Some programs (e.g., Cornell University, Colorado State University, University of Minnesota) with strong median ranks exhibited substantial variation in rankings across the 8 metrics, whereas other highly ranked programs (e.g., University of California-Davis, North Carolina State, Purdue University) were notable for consistency in rankings and hence low variation across the metrics (Fig 2).

Benchmarking across disciplines

When the mapped set of JCR categories was considered (Table 2), annual values for citations per article from 2005–2014 were 1.19 to 4.18 times greater than the citations per article in mathematics. The normalization ratios differed significantly between genetics and all other

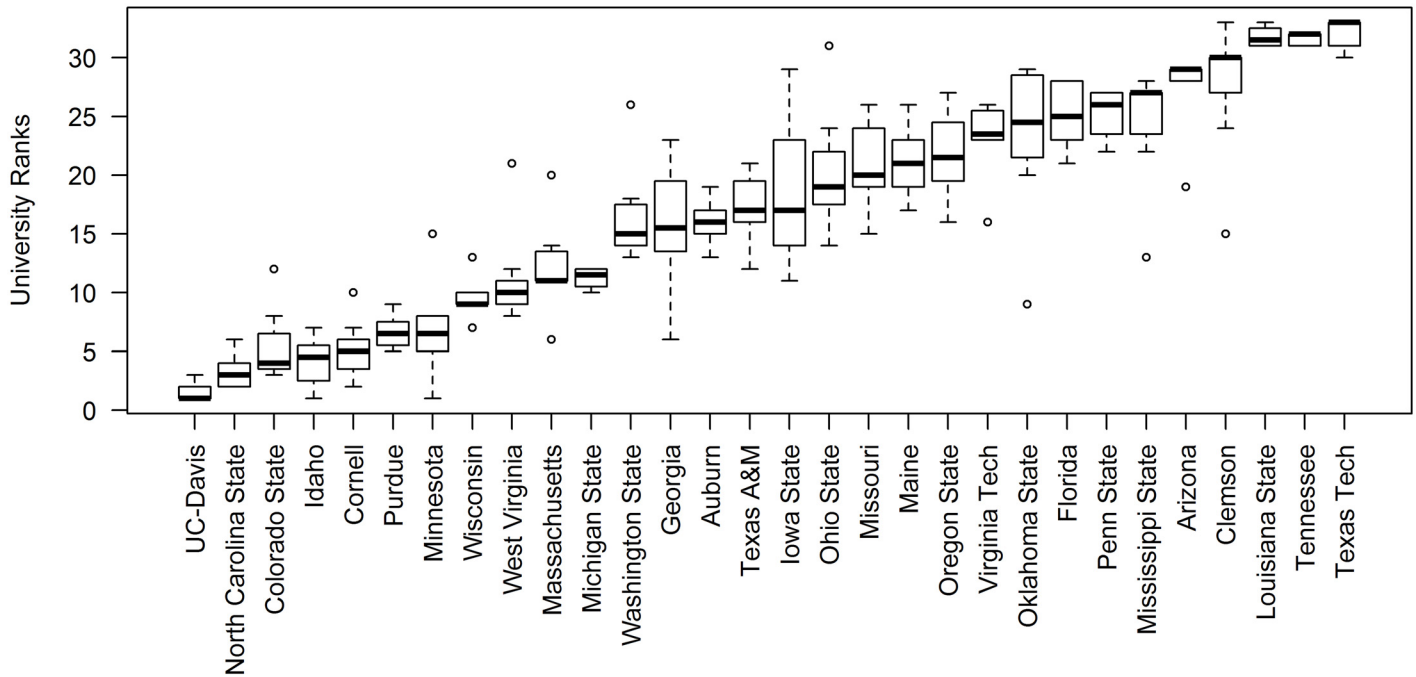


Fig 2. Performance metrics of fisheries and wildlife faculty vary with university affiliation. The box plot depicts median, interquartile width, and range of the mean ranks (1 = best) of standardized deviance residuals across 8 performance metrics for 30 universities as computed from models fitted to data for 437 faculty. Three universities (Connecticut, Kentucky, Nebraska) are not shown because they had fewer than five faculty members in fisheries and wildlife.

doi:10.1371/journal.pone.0155097.g002

sub-disciplines (all $P < 0.001$), with highest values for genetics and lowest values for social sciences (Fig 3). Ten-year mean ± 1 SD values for normalization ratios were 3.73 ± 0.29 for genetics, 3.01 ± 0.45 for ecology, 2.17 ± 0.35 for conservation, 2.15 ± 0.24 for aquatic sciences, 2.03 ± 0.11 for quantitative, 1.96 ± 0.19 for health sciences, and 1.45 ± 0.13 for social sciences.

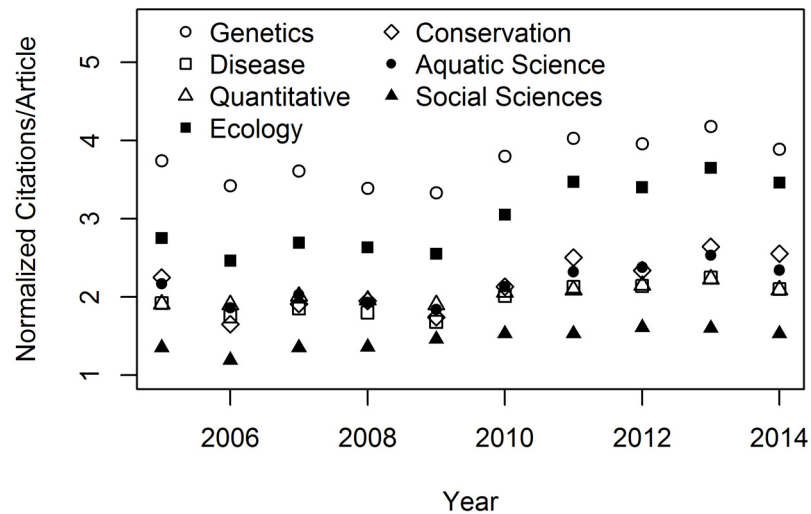


Fig 3. Citation practices in fisheries and wildlife sub-disciplines do not obey the law of constant ratios. Eighteen categories from Web of Science Journal Citation Reports® were clustered into the 7 sub-disciplinary categories (Table 2) to derive annual averages for citations per article, normalized to mathematics. Ratios increased over time, with comparable rates except for ecology.

doi:10.1371/journal.pone.0155097.g003

A significant increase in normalization ratios occurred over time ($t = 3.32$, $df = 56$, $p = 0.002$), indicating that citations per article for the disciplines in question increased relative to mathematics over the 10-year period (Fig 3). Decadal sets of annual normalization ratios were highly correlated for all pairs of disciplines (mean $r = 0.90$, range: 0.75–0.98). Moreover, disciplines did not differ in the rate at which their normalization ratios changed relative to genetics (all $p > 0.20$), except for ecology ($t = 2.28$, $df = 56$, $p = 0.03$). Ecology experienced a more rapid increase in its normalization ratio than genetics, driven primarily by increases from 2010–2014 (Fig 3).

Discussion

Benchmarking basics

Our approach to benchmarking differs from previous studies because it explicitly attempts to account for variation in publication and citation metrics due to academic age, sex, research appointment, and sub-disciplinary focus. Collectively, these covariates reduced deviance by roughly $\frac{1}{4}$ to $\frac{1}{2}$, although sex was only important when predicting number of publications. Such levels of explanatory power may seem modest, but we contend that incorporation of these covariates into benchmarking efforts is a marked improvement over prior methods because doing so enables comparison of individuals with differing academic age or research appointment and focus. Career duration has long been recognized as a factor that influences performance metrics, especially those that increase monotonically over time [14,19,34,35]. Prior benchmarking studies occasionally have accounted for age effects by conducting cohort analyses, but this approach is inferior to one that uses covariates to allow for “tailor-made” adjustments for every individual [18]. For soil scientists, a simple linear regression of h-index against academic age was provided, but other covariates were not considered nor were residuals used for comparisons of individuals [9]. A regression of h-index against academic age and sex has also been published [15], but for a sample limited to members of editorial boards of seven evolution and ecology journals. Likewise, sub-disciplinary effects have been noted in other fields [22,36,37], but attempts to account for variation due to sub-disciplinary focus thus far have relied on reference groups such as Nobel laureates [22] or named professors [20]. We are unaware of attempts to incorporate a faculty member’s research appointment, perhaps because this information is not provided on bibliometric search engines.

Standardized residuals offer a convenient statistic with which to compare discrepancies between observed and expected performance. At the level of individual faculty, the simplest approach is to compute standardized residuals and associated percentiles for each performance metric. We presented models for 8 metrics because each has its own strengths and supporters. Our hypothetical faculty member ranked ahead of only 8.7% of the 437 “peers” with whom he was compared using the m index, but he ranked ahead of 17.8% of them when number of publications was considered (Table 1). If general trends are more important than consideration of individual metrics, a mean percentile can be computed across all 8 metrics for each faculty member, or for some subset of metrics of interest. Such an approach provides an overall benchmark of performance as well as an indication of the degree to which the performance varied among the metrics under consideration. For the hypothetical faculty member, the overall mean performance ranked ahead of 8.4% of peers and was fairly consistent across the 8 metrics (95% CI: 6.2–11.6%).

Factor analysis is a more sophisticated approach with the same goal as above, namely, to express performance meaningfully in a reduced number of dimensions. Factor scores have been used to demonstrate the utility of performance metrics in predicting scores derived independently via peer assessment for applicants to a long-term fellowship program in molecular biology [38,39]. We interpret the factors extracted from our data as differentiating between

faculty who publish more articles and hence are more productive (Factor-2) and faculty who have higher citations or impact independent of productivity (Factor-1). Our 2 factors explained 88% of the variation in the standardized residuals for the 8 metrics (Table 4). The hypothetical faculty member fell within the convex hull containing 75% of all faculty members, but his performance was below average for both factors (Fig 1). Indeed, joint consideration of the two factors reveals that only 11.0% of faculty exhibited a weighted mean score as low or lower than the hypothetical faculty member. In general, the quadrant in which a faculty member resides conveys information on the joint distribution of the factor scores. Quadrant I is occupied by faculty who exceed the average in productivity and citations, whereas Quadrant III is populated by faculty who fall below the average in both of these dimensions. Quadrant II contains faculty members who produce more publications than average but fewer citations, i.e., they exhibit performance that emphasizes quantity over impact on peers. Although they publish as much as their otherwise comparable peers in Quadrant I, each of their articles is cited less. Finally, Quadrant IV contains faculty who produce fewer-than-expected publications but greater-than-expected citation metrics. Faculty members in Quadrant IV seem to emphasize impact of publications, as measured by citation metrics, over quantity. Although they produce fewer publications than their otherwise comparable counterparts in Quadrant I, each of their publications receives more attention from peers.

Faculty members exhibit considerable variation in performance after accounting for variation explained by age, research sub-discipline(s), percent research appointment, etc. Far less variation was noted among institutions, although some differences were evident (Fig 2). Students and prospective faculty should exercise caution when examining these metric-based rankings of institutions. They represent a snapshot of performance that will change as faculty retire and are hired. Moreover, they are silent on many of the other factors that are important to quality educational and career goals including social climate, administrative support, curricular offerings, and depth of expertise, to name a few [40]. Nonetheless, they do suggest that expectations for performance can be institutionalized to some degree, perhaps due to differences in infrastructure and research support, reward systems, or the standards used for hiring, promotion and tenure processes.

Normalized citations per year were not constant over time for the sub-disciplines affiliated with fisheries and wildlife (Fig 3), in contrast to previous work [32]. However, the rate of change in citations per article relative to mathematics did not differ across sub-disciplines, except for ecology. Thus, a “quick-and-dirty” comparison between disciplines other than ecology is possible for those who eschew model-based benchmarking. To illustrate, suppose that faculty member X in genetics has 1000 citations and 100 publications (10 cites/article), whereas faculty member Y in social sciences has 500 citations and 100 publications (5 cites/article). The raw metric for citations per article indicates that faculty X produces cites/article at twice the level of faculty Y. But the normalization ratios, 3.73 for genetics and 1.45 for social science, suggest that faculty Y has had greater impact than faculty X, after taking into account differences in citation and publication practices between disciplines. Specifically, when considered in reference to mathematics, faculty X produces $10/3.73 = 2.68$ cites/article, whereas faculty Y produces $5/1.45 = 3.45$ cites/article. The application of this approach should be limited to broad comparisons that inform discussion among academic units, because, as we have shown, other factors (sex, research appointment, academic age) can exert effects independent of discipline.

Benchmarking caveats

We believe that results from our benchmarking study are useful as indicators of past performance based on citations or publications after accounting for variation due to academic age,

sex, research appointment, and sub-disciplinary focus. However, they represent indices of performance at a particular point and should not be used to forecast future success in academia. Cumulative measures of achievement, like the h-index and its variants, tend to be poor predictors of future success due to temporal fluctuations in productivity [41]. Importantly, assessment of predictive power of regressions for such indexes is invariably inflated by their inherent autocorrelation [42]. Higher levels of variation in m-quotient values for faculty early in their academic careers also limits its utility for prediction with this group [20]. If the intent is to predict future success, a focus on rates of past achievement may be more appropriate [41,43]. Moreover, the definition of success may influence the predictive utility of various metrics. For >25,000 scientists in PubMed, the probability of becoming a principal investigator was modeled with relatively high accuracy (area under the logistic regression curve = 0.83) as a function of performance metrics such as number of first-authored publications, number of publications, and h-index [44]. When success was defined as the number of publications produced over the decade following conferral of a Ph.D., the number of prior publications and an earlier date of first publication relative to the Ph.D. were positive predictors of success for a sample of academics in biological and environmental sciences, but the best models only explained about 15% of the deviance [21].

Estimates from our regression models are of course dependent on the accuracy of bibliometric and covariate data used in analyses. We relied on data from the Web of Science to compile the data from which response metrics were computed. We chose Web of Science because it is used routinely. However, it is important to note that rankings produced by data from Scopus or Google Scholar may differ from Web of Science due to differences in the database coverage [45,46]. For any given database, blind reliance on bibliometrics computed automatically by search engines can lead to inaccuracies caused by errors of omission or commission [47]. To minimize this problem, we followed the recommendation [47] to conduct partially manual searches [20]. Still, we suspect that accuracy declined for more senior faculty in our sample, presumably corresponding to poorer coverage of old publications by Web of Science. We attempted to compensate for this perceived bias by conducting more individualized searches for these faculty, but errors may still have occurred.

Inaccuracies also could affect data collected for covariates. For instance, we relied on administrators from the target universities to provide information on research appointments of faculty. At least three factors could increase noise in these data. First, accounting practices associated with research appointments may differ among institutions. Second, research appointments may change over time; in our study we used the most recent appointment unless information provided in correspondence with administrators indicated that doing so was inappropriate due to a faculty member's much longer period of service in a different research appointment. Finally, faculty appointments may not accurately reflect how a faculty member allocates time to research versus other activities. Given these caveats, it is noteworthy that we found strong, consistent effects of research appointment for all metrics considered.

Our assignment of faculty according to sub-disciplines was subjective and based on assessment of websites and publications. Hence, it certainly is possible that faculty were misclassified in some instances. Moreover, our sub-disciplinary categories may themselves represent heterogeneous mixtures of research focus. For example, the quantitative sub-discipline included research on quantitative methods and on geospatial science, even though it could be argued that these warrant separate sub-disciplinary designations. Future attempts to incorporate sub-disciplinary effects could benefit from more objective classification, perhaps using visualization of key words [37]. Still, the similarity in citation rates for JCR categories that were mapped onto individual sub-disciplines (S2 Table) provides some support for our classification. Individual faculty members are likely to work in multiple sub-disciplines [20], with associated

effects on benchmarking metrics; indeed, a unimodal relationship has been shown between research breadth and h-index [48]. In situations where benchmarking is desired for faculty who are difficult to categorize using our sub-disciplinary groupings, regression models can easily be re-fitted without sub-disciplinary covariates.

We believe that our method improves on earlier efforts at benchmarking because it addresses several of the issues raised in the Leiden Manifesto [30]. For instance, those who rely on quantitative indicators of productivity and impact should choose methods that account for variation in citation and publication practices across fields of inquiry, adopt multiple indicators to offer a more well-rounded description of performance, and ensure transparency in data and analyses [30]. Nonetheless, we have devoted considerable discussion to caveats of our benchmarking method because, to be blunt, quantitatively based rankings of performance are prone to abuse. We strongly agree with other researchers [30] that quantitative benchmarks should supplement, not supplant, expert assessment of productivity and impact by an individual's peers. Similar to the caution provided by Engqvist and Frommen [49], it would be inappropriate to base important decisions (e.g., hiring, promotion, allocation of grant funds) on small differences in rankings from our benchmarking models. Even if our models could perfectly account for variation in publication and citation metrics, they should not replace critical evaluation by peers. A faculty member's reputation often is built on more than publications and citations; for instance, broader impacts that affect individual behavior or social policy may not necessarily be related to bibliometrics [3]. And of course our models, as with all models, are imperfect simplifications of reality. While models can provide important insights into performance, they are silent on numerous sources of variation in performance that may be recognized only by knowledge gained from careful inspection of an individual's record. Such sources of variation for an individual researcher may include details on the specific research questions asked and approaches used, philosophies on student mentoring and co-authorship, levels of commitment to professional service, or institutional inequities in research support. Our benchmarking methods can provide crucial information, but they should be viewed as one set of tools in a comprehensive evaluation system [30].

Supporting Information

S1 Code. Manual and R (3.0.2) code to perform comparisons and analyses with additional faculty members.

(PDF)

S1 Data. Database of 437 faculty used for model-based benchmarking of fisheries and wildlife.

(CSV)

S2 Data. Sample database of “new” individuals for inclusion in analysis after re-fitting of models with [S1 Code](#).

(CSV)

S1 Table. Correlations of standardized deviance residuals from 8 models for faculty in fisheries and wildlife.

(DOCX)

S2 Table. Average annual normalized citations per article for JCR categories and associated fisheries and wildlife sub-disciplines as grouped in [Table 2](#). Normalization was accomplished by dividing values by the corresponding citations per article for mathematics.

(DOCX)

Acknowledgments

We thank the dozens of administrators at participating universities who responded to our request for information on functional appointments of faculty. S. A. Hamilton helped with collection of covariate data and organization of the database. JD Antonides, JL Berl, JM Doyle, ZS Feiner, NB Fernandez, J Hönekopp, Y Ji, NI Lichti, S Mathur, DL Nelson, M Nizamlioglu, TJ Smyser, JR Willoughby, and an anonymous reviewer provided helpful comments on an earlier version of the manuscript.

Author Contributions

Conceived and designed the experiments: RKS MS. Performed the experiments: RKS MS TOH JAD. Analyzed the data: RKS MS KFK. Contributed reagents/materials/analysis tools: RKS MS KFK. Wrote the paper: RKS MS TOH JAD KFK.

References

1. Acuna DE, Allesina S, Kording KP. Future impact: Predicting scientific success. *Nature*. 2012; 489: 201–202. doi: [10.1038/489201a](https://doi.org/10.1038/489201a) PMID: [22972278](https://pubmed.ncbi.nlm.nih.gov/22972278/)
2. Whitley R. *The intellectual and social organization of the sciences*. Oxford: Oxford University Press; 2000.
3. Wildgaard L, Schneider JW, Larsen B. A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics*. 2014; 101: 125–158.
4. Alonso S, Cabrerizo FJ, Herrera-Viedma E, Herrera F. h-index: A review focused in its variants, computation and standardization for different scientific fields. *J. Informetr*. 2009; 3: 273–289.
5. Bornmann L, Mutz R, Hug SE, Daniel H-D. A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *J Informetr*. 2011; 5: 346–359.
6. Liu Y, Zuo W, Gao Y, Qiao Y. Comprehensive geometrical interpretation of h-type indices. *Scientometrics*. 2013; 96: 605–615.
7. Abramo G, D'Angelo CA, Viel F. A robust benchmark for the h- and g-indexes. *J Am Soc Inf Sci Technol*. 2010; 61: 1275–1280.
8. Benway BM, Kalidas P, Cabello JM, Bhayani SB. Does citation analysis reveal association between h-index and academic rank in urology? *Urology*. 2009; 74: 30–33. doi: [10.1016/j.urology.2008.10.045](https://doi.org/10.1016/j.urology.2008.10.045) PMID: [19567283](https://pubmed.ncbi.nlm.nih.gov/19567283/)
9. Minasny B, Hartemink AE, McBratney A. Soil science and the h index. *Scientometrics*. 2007; 73: 257–264.
10. Pagel PS, Hudetz JA. H-index is a sensitive indicator of academic activity in highly productive anaesthesiologists: results of a bibliometric analysis. *Acta Anaesthesiol Scand*. 2011; 55: 1085–1089. doi: [10.1111/j.1399-6576.2011.02508.x](https://doi.org/10.1111/j.1399-6576.2011.02508.x) PMID: [22092205](https://pubmed.ncbi.nlm.nih.gov/22092205/)
11. Pagel PS, Hudetz JA. An analysis of scholarly productivity in United States academic anaesthesiologists by citation bibliometrics. *Anaesthesia*. 2011; 66: 873–878. doi: [10.1111/j.1365-2044.2011.06860.x](https://doi.org/10.1111/j.1365-2044.2011.06860.x) PMID: [21864299](https://pubmed.ncbi.nlm.nih.gov/21864299/)
12. Selek S, Saleh A. Use of h index and g index for American academic psychiatry. *Scientometrics*. 2014; 99: 541–548.
13. Sharma B, Boet S, Grantcharov T, Shin E, Barrowman NJ, Bould MD. The h-index outperforms other bibliometrics in the assessment of research performance in general surgery: A province-wide study. *Surgery*. 2013; 153: 493–501. doi: [10.1016/j.surg.2012.09.006](https://doi.org/10.1016/j.surg.2012.09.006) PMID: [23465942](https://pubmed.ncbi.nlm.nih.gov/23465942/)
14. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*. 2005; 102: 16569–16572. PMID: [16275915](https://pubmed.ncbi.nlm.nih.gov/16275915/)
15. Kelly CD, Jennions MD. The h index and career assessment by numbers. *Trends Ecol Evol*. 2006; 21: 167–170. PMID: [16701079](https://pubmed.ncbi.nlm.nih.gov/16701079/)
16. Podlubny I, Kassayova K. Towards a better list of citation superstars: compiling a multidisciplinary list of highly cited researchers. *Res Eval*. 2006; 15: 154–162.
17. Liang L. h-index sequence and h-index matrix: Constructions and applications. *Scientometrics* 2006; 69: 153–159.
18. Perianes-Rodriguez A, Ruiz-Castillo J. Within- and between-department variability in individual productivity: the case of economics. *Scientometrics*. 2015; 102: 1497–1520.

19. Pepe A, Kurtz MJ. A measure of total research impact independent of time and discipline. *PLoS One*. 2012; 7: e46428. doi: [10.1371/journal.pone.0046428](https://doi.org/10.1371/journal.pone.0046428) PMID: [23144782](https://pubmed.ncbi.nlm.nih.gov/23144782/)
20. Swihart RK, Sundaram M, Hook TO, DeWoody JA. Predictors of scholarly productivity and impact by faculty in fisheries and wildlife: a bibliometric analysis. *J Wildl Manage*. 2016; 80: 563–572.
21. Laurance WF, Useche DC, Laurance SG, Bradshaw CJA. Predicting publication success for biologists. *Bioscience*. 2013; 63: 817–823.
22. Sorzano COS, Vargas J, Caffarena-Fernández G, Iriarte A. Comparing scientific performance among equals. *Scientometrics*. 2014; 101: 1731–1745.
23. Brown OR. The h_b -index, a modified h-index designed to more fairly assess author achievement. *Redox Rep*. 2012; 17: 176–178.
24. Zhang C-T. The e-index, complementing the h-index for excess citations. *PLoS One*. 2009; 4: e5429. doi: [10.1371/journal.pone.0005429](https://doi.org/10.1371/journal.pone.0005429) PMID: [19415119](https://pubmed.ncbi.nlm.nih.gov/19415119/)
25. Jin BH, Liang LL, Rousseau R, Egghe L. The R and AR indicators: Complementing the h-index. *Chin Sci Bull*. 2007; 52: 855–863.
26. Venables WN, Ripley BD. *Modern applied statistics with S*. 4th ed. New York: Springer Science and Business Media; 2002.
27. R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Available: <http://www.R-project.org/>. 2015.
28. Zuur AF, Ieno EN, Smith GM. *Analysing ecological data*. New York: Springer Science and Business Media; 2007.
29. Hilbe JM, Robinson AP. Residual Analysis. In: *Methods of statistical model estimation*. Boca Raton: CRC Press; 2013. pp. 141–142.
30. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. The Leiden Manifesto for research metrics. *Nature*. 2015; 520: 429–431. doi: [10.1038/520429a](https://doi.org/10.1038/520429a) PMID: [25903611](https://pubmed.ncbi.nlm.nih.gov/25903611/)
31. Laband DN, Zhang D. Citations, publications, and perceptions-based rankings of the research impact of North American forestry programs. *J For*. 2006; 104: 254–261.
32. Podlubny I. Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*. 2005; 64: 95–99.
33. Zhang Z, Cheng Y, Liu NC. Comparison of the effect of mean-based method and z-score for field normalization of citations at the level of Web of Science subject categories. *Scientometrics*. 2014; 101: 1679–1693.
34. Petersen AM, Penner O, Stanley HE. Methods for detrending success metrics to account for inflationary and deflationary factors. *Eur Phys J B*. 2011; 79: 67–78.
35. Aoun SG, Bendok BR, Rahme RJ, Dacey RG Jr, Batjer HH. Standardizing the evaluation of scientific and academic performance in neurosurgery—critical review of the “h” index and its variants. *World Neurosurg*. 2013; 80: E85–E90. doi: [10.1016/j.wneu.2012.01.052](https://doi.org/10.1016/j.wneu.2012.01.052) PMID: [22381859](https://pubmed.ncbi.nlm.nih.gov/22381859/)
36. Thompson DF, Nahata MC. Pharmaceutical science faculty publication records at research-intensive pharmacy colleges and schools. *Am J Pharm Educ*. 2012; 76: 173. doi: [10.5688/ajpe769173](https://doi.org/10.5688/ajpe769173) PMID: [23193337](https://pubmed.ncbi.nlm.nih.gov/23193337/)
37. van Eck NJ, Waltman L, van Raan AFJ, Klautz RJM, Peul WC. Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLoS One*. 2013; 8: e62395. doi: [10.1371/journal.pone.0062395](https://doi.org/10.1371/journal.pone.0062395) PMID: [23638064](https://pubmed.ncbi.nlm.nih.gov/23638064/)
38. Bornmann L, Wallon G, Ledin A. Is the h index related to (standard) bibliometric measures and to the assessments by peers? An investigation of the h index by using molecular life sciences data. *Res Eval*. 2008; 17: 149–156.
39. Bornmann L, Mutz R, Daniel H-D, Wallon G, Ledin A. Are there really two types of h index variants? A validation study by using molecular life sciences data. *Res Eval*. 2009; 18: 185–190.
40. Hattie, Marsh HW. The relationship between research and teaching: a meta-analysis. *Review Educ Res*. 1996; 66: 507–542.
41. Hönekopp J, Khan J. Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics*. 2012; 90: 843–83.
42. Penner O, Pan RK, Petersen AM, Kaski K, Fortunato S. On the predictability of future impact in science. *Sci Rep*. 2014; 3: 3052.
43. Mazloumian A. Predicting scholars' scientific impact. *PLoS One*. 2012; 7: e49246. doi: [10.1371/journal.pone.0049246](https://doi.org/10.1371/journal.pone.0049246) PMID: [23185311](https://pubmed.ncbi.nlm.nih.gov/23185311/)
44. van Dijk D, Manor O, Carey LB. Publication metrics and success on the academic job market. *Curr Biol*. 2014; 24: R516–R517. doi: [10.1016/j.cub.2014.04.039](https://doi.org/10.1016/j.cub.2014.04.039) PMID: [24892909](https://pubmed.ncbi.nlm.nih.gov/24892909/)

45. Henzinger M, Suñol J, Weber I. The stability of the h-index. *Scientometrics*. 2010; 84: 465–479.
46. Sharma B, Boet S, Grantcharov T, Shin E, Barrowman NJ, Bould MD. The h-index outperforms other bibliometrics in the assessment of research performance in general surgery: A province-wide study. *Surgery*. 2013; 153: 493–501. doi: [10.1016/j.surg.2012.09.006](https://doi.org/10.1016/j.surg.2012.09.006) PMID: [23465942](https://pubmed.ncbi.nlm.nih.gov/23465942/)
47. Jasco P. Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for F.W. Lancaster. *Libr Trends*. 2008; 56: 784–815.
48. Belmaker J, Cooper N, Ming Lee T, Wilman H. Specialization and the road to academic success. *Front Ecol Environ*. 2010; 8: 514–515.
49. Engqvist L, Frommen JG. The h-index and self-citations. *Trends Ecol Evol*. 2008; 23: 250–252. doi: [10.1016/j.tree.2008.01.009](https://doi.org/10.1016/j.tree.2008.01.009) PMID: [18367289](https://pubmed.ncbi.nlm.nih.gov/18367289/)