Inequality in measuring scholarly success: Variation in the h-index within and between disciplines

Ryan Light[1*], Aaron Gullickson[2], Jill Ann Harrison[3]

1 University of Oregon, Sociology, 632 PLC, Eugene, 97405, OR, USA,
2 University of Oregon, Sociology, 727 PLC, Eugene, 97405, OR, USA,
3 University of Oregon, Sociology, 604 PLC, Eugene, 97405, OR, USA,

☯These authors contributed equally to this work.

¤Current Address: Dept/Program/Center, Institution Name, City, State, Country

†Deceased

¶Membership list can be found in the Acknowledgments section.

* light@uoregon.edu

Abstract

Scholars and university administrators have a vested interest in building equitable valuation systems of academic work for both practical (e.g., resource distribution) and more lofty purposes (e.g., what constitutes "good" research). Well-established inequalities in science pose a difficult challenge to those interested in constructing a parsimonious and fair method for valuation as stratification occurs within academic disciplines, but also between them. The h-index, a popular research metric, has been formally used as one such method of valuation. In this article, we use the case of the h-index to examine how the distribution of research metrics reveal within and between discipline inequalities. Using ~~data from~~ bibliometric data from 1960-2019 on over 50,000 high performing scientists - the top 2% most frequently cited authors - across 174 disciplines, we construct random effects within-between models predicting the h-index. Results suggest significant within-discipline variation in several forms, specifically sole-authorship and female penalties~~, as well as significant between discipline variation in terms of sole authorship~~. Results also show that a sole authorship penalty plays a signficant role in well-known between-discipline variation. Field-specific models emphasize the "apples-to-oranges~~"~~," or incommensurable, property of cross-discipline comparison with significant heterogeneity in sole-authorship and female penalities within fields. Conclusions include continued caution when using the h-index or similar metrics for valuation purposes.

Introduction

From teaching to research, systematic evaluation of academic work presents a unique set of challenges as the academic disciplines that broadly organize scholarly labor differ ~~upon~~ by several relevant factors. ~~Differences in demographic characteristics, like race and gender, capture processes of racism and sexism~~ Inequalities experienced by minoritized and women scholars point to processes of social closure or exclusionary practices that limit the opportunities for some scholars and not others across disciplines

[1, 2, 3]. Cultural and economic differences between disciplines, like the prevalence of team science or even the definition of success, can affect the quality and quantity of publication [4, 5, 6]. Scholars have invented dozens of metrics that are formally or informally used in the evaluation of scholarly research, despite warnings about apples-to-oranges comparisons given within discipline and between discipline differences [7, 8]. The most widely used metric of this type is the h-index, a simple score where a scholar with 10 published articles with at least 10 citations receives a h-index of 10. The limitations of the h-index are well understood, but less is known about how processes of inequality are embedded both within and between disciplines. ~~Here, we ask three related questions:~~

1. ~~What is the variation within and between disciplines in the h-index?~~
2. ~~What within and between discipline factors contribute to this variation?~~
3. ~~How do within and between discipline factors vary across fields (e.g., social sciences, medical sciences, natural sciences)?~~

~~These questions arise from at least two important considerations. First, like other forms of work, academic labor is beset by inequalities. A large body of research describes the factors that contribute to inequality in academia [see 10, 13, for reviews] .~~ From well-known pipeline effects to publication differences to varying effects of parenthood, ~~scholars have~~ research has shown the multifaceted ways some scholars are obstructed from paths to success at work ~~[e.g. 9, 12, 14, among others]~~ [9, 10, 12, 13, 14]. At the same time, differences exist in the distribution of resources to disciplines~~as can be seen in variation in~~ , such as the prioritization of the natural sciences and engineering in federal funding and the ~~way resourcesare distributed~~ unequal distribution of resources, like salaries, to departments, the proxies for disciplines on university campuses [15]. Second, numerous scholars have drawn attention to the increasing push to quantify aspects of social life [16]. Research on the risks of quantification points to the crude ways that quantification reduces complex social phenomena, often with implications for inequality and nearly always with implications for valuation, or what is considered good or bad. In the case of scholarly metrics, critics have raised concerns about how these metrics transform scholarship into a capitalist-like market (e.g., neoliberalization), build unproductive individual constraints into academic labor - such as the anxiety connected to hyper-competitiveness - and ultimately result in less productive and innovative scholarship [17, 18] and efforts to "game metrics" at either the journal or individual-level [19]. ~~Turning to~~ Examining within and between discipline variation in the h-index offers insight into how metrics relate to inequalities and contributes to research on the limitations of simplifying summaries of academic labor through quantification. Here, we ask three related questions:

1. What is the variation within and between disciplines in the h-index?
2. How do social factors, specifically gender and sole authorship, contribute to this variation?
3. How do within and between discipline factors vary across fields (e.g., social sciences, medical sciences, natural sciences)?

While prior work examines the h-index by field, discipline, or gender, less is known about how ~~within and~~ gender and sole authorship contribute to *both* within *and* between discipline differences ~~affect~~ in h-index scores ~~in relationship to one another [20]. To understand factors that contribute to differences in scholar's~~ [20]. This study extends prior work on metrics in science generally and the h-index ~~, or Hirsch index, we use data on over 50,000 high-performing scientists across 174 disciplines in the United States. This analysis draws from Ioannidis et al.'s dataset identifying the top~~

~~scientists in the Scopus database [8]. To understand within and between discipline differences, we use random effects within-between (REWB) models predicting~~ specifically by conceptualizing the inequality process as occurring via intra-disciplinary and inter-disciplinary dynamics. Variation in the h-index ~~with a range of factors including disciplinary age, the number of sole publications, a female name index, and a measure of university productivity. The REWB model accounts for distinct within and between disciplinary effects on h-index scores, consistent with calls to develop multilevel analyses of gender disparities in science [21]. Results indicate that substantial within discipline differences exist, including both female and sole authorship penalties. Results further show substantial between discipline differences related to sole authorship. Moreover, roughly a third of the variation in h-index is between rather than within disciplines, suggesting an important~~ may result from "apples-to-oranges~~problem when comparing the h-index across disciplines . This apples-to-oranges problem results in significant differences across fields as well, especially in terms of sole authorship. Conclusions highlight how the continued use of the h-index or similar metrics for valuation perpetuates inequality in science both within and between disciplines~~" comparisons across disciplines and fields due to specific factors, like gender and coauthorship, but also occur within disciplines, as scholars experience differing publishing trajectories dependent on gender and coauthorship.

Inequality in Science

Inequality in science, and in academia generally, persists within and between disciplines and along well-known axes. Gender is one of the most studied axes of inequality in science. Research on the scientific pipeline, for example, shows how obstacles, like cultural stereotypes about skills differences in math and science and related social psychological factors such as a sense of belonging, limit pathways to particular scientific fields for girls [22, 23, 24]. These pipeline factors continue throughout the research life course as men and women become segregated in doctoral programs by field and prestige [25]. When these significant obstacles are overcome, women continue to experience inequality in academic work, including in publication. Publication is both the outcome of academic labor and the currency of academic careers. How many and the type of publications produced by a scientist often translates into tangible resources, like salary raises and the job security of tenure, and less tangible resources, like prestige. While significant gains have been made by women in academic work - numerous fields that were male dominated in the mid- to late-twentieth century are now majority female - publication remains a ~~site of persistent inequality [3]~~potential site of stratification in terms of both the number of publications scholars accrue over their careers and the quality of those publications [3, 33]. Early signs indicate that these forms of gender inequality may have increased due to the COVID-19 pandemic or a "pandemic penalty" [27].

Sociologists of science have spent decades trying to disentangle the factors associated with gender differences in publication, especially related to differences in the number of publications and citations. Nearly 40 years ago, Cole and Zuckerman referred to the ongoing male advantage in publication and citation counts as the "productivity puzzle" because the causes of this advantage remained difficult to pinpoint [28]. They conclude, "[S]ince gender differences in published productivity persist, the productivity puzzle has yet to be solved" (pg. 250). ~~While women have made significant gains in many academic fields, research~~ Research suggests that the productivity puzzle remains. For example, Erin Leahey's research shows how the level of research specialization interacts with gender to affect productivity, with consequences for earnings [29, 30]. More recent

work provides evidence that the productivity puzzle in STEM fields results from variation between men and women in career length and exit rates as productivity appears to be more equal across shorter time horizons [31]. Scholars have also turned to more complex mechanisms that may help perpetuate gender and racial hierarchies in academic work. For example, using data on US doctoral recipients, Hofstra ~~and coauthors~~ et. al. find that gender and racial minorities are more likely to generate innovative scientific work, but that this work is less likely to be adopted by future researchers with consequences for academic hiring [1].

A prestige puzzle may also coincide with the productivity puzzle as women may be less likely to publish in the ~~top or~~ most prestigious journals in their fields [32]. As top journals have higher impact factors, the prestige puzzle could have a significant impact on differences in publication metrics and career outcomes. This form of inequality may occur for a variety of reasons including a lack of mentorship, different family and work-based responsibilities between men and women, and differences in specialization similar to Leahey's work on specialization and productivity [29, 30, 33]. Drawing on literature on occupational segregation and identity, this work on elite publication shows how the prestige puzzle has changed over time within sociology [33]. Earlier cohorts of women sociologists were significantly less likely to publish in top sociology journals compared to men regardless of specialty areas. However, as more women entered sociology, occupational segregation occurred with subfields becoming sharply gender imbalanced. While baseline models of the prestige puzzle for more recent cohorts reveal the persistence of this form of inequality, more complete models that control for specialization, or occupational identity, show that the contemporary effect likely operates through these segregation processes. More recent work on sociology, economics, and political science shows a null effect of gender on citation when social scientists are situated in similar disciplinary and sub-field spaces, suggesting that teasing apart the contexts when a gender penalty persists and when it does not remains an important concern for those interested in inequality [34].

Collaboration also likely plays a role with significant historical differences in coauthorship networks based on gender [35]. Recent research on computer scientists by Jadidi and coauthors finds significant differences in collaboration between men and women. Men are more likely to have larger coauthorship networks and to play brokerage roles within them, while women's networks exhibit increasing gender homophily [36]. These differences relate to publication outcomes as the network factors positively affect measures of productivity, including the h-index. Jadidi and coauthors conclude that women "on average are less likely to adapt to the collaboration patterns that are related with success. However, those women who become successful computer scientists exhibit the same collaborative behavior as their successful male colleagues" (pg. 19). Research in both political science and sociology also identifies how team science affects gender and publication in these disciplines pointing to how the structure of scientific work can negatively impact women social scientists [37, 38].

Collaboration may impact individual-level inequalities beyond gender. Collaboration is broadly understood as a key aspect of epistemic culture. Epistemic culture consists of "those amalgams of arrangements and mechanisms... which, in a given field, make up how we know what we know" ([39], pg. 1). The questions that scholars ask and the strategies that they use to answer them differ by field with consequences for individuals embedded in specific cultures. Plainly, the structure of team science matters for publication outcomes. This suggests that within discipline differences may affect publication metrics, like the h-index, but also points to how inequality may occur between disciplines as pronounced differences exist between disciplines in terms of factors like gender composition and team versus sole authorship.

Inequality Between Disciplines and Fields

Disciplinary differences affect inequality along several dimensions. For example, differences in gender composition may have direct and indirect effects on how resources are distributed in universities. Disciplinary cultures also differ and these differences may affect inequality. For example, disciplines differ in terms of how work is evaluated [40] or even how emotions are expressed at work [41]. Moreover, disciplines differ in terms of how academic work is conducted [31]. Do scholars collaborate in teams or are they more likely to work alone? Less is known about how these between discipline-based inequalities differ from within discipline inequality. Research on collaboration and citation impact using the h-index shows disciplinary differences in the effect that collaboration has on impact with more collaboration having a stronger positive effect in physics and medicine, while having a smaller effect in the brain sciences or computer science [42]. Disciplinary differences occur regarding more tangible resources, like federal funding. Research describes the substantial differences between fields in terms of federal funding with implications - a "domino effect" - for future non-federal funding and a potential site of cumulative advantage stratifying disciplines [43, 44].

One of the ways that disciplines differ and also one of the ways that disciplines may be valued differently is how they perform on commonly used metrics. Time-worn debates in the philosophy of science have tried to identify the implications of and/or reconcile the so-called "two cultures" or the division between the arts and the sciences [45]. While this debated formulation may exaggerate differences, the two cultures perspective draws attention to the variation that occurs across fields both at a philosophical-level, but also as a more practical concern. In terms of the latter, prior work has suggested field normalizing the h-index to account for between field variation [46]. Questions remain about the extent of variation within and between fields in addition to the variation occurring within and between disciplines situated in fields. Do field-level variations in gender composition and team science result in apple-to-oranges comparisons when using common scholarly metrics, like the h-index?

The Risks of Quantification

Quantification has become a central feature of contemporary life as "[a]dministration, management, and even mundane daily activities are increasingly structured around performance measures, cost-benefit analysis, risk calculations, ratings, and rankings" [16, p. 224]. Critiques often focus on the risks of quantification as a central factor in determining worth, which occurs in a variety of fields from the law to business to education [18, p. 4]. This "metrics fixation" is part of the broader process of neoliberalization of education. Neoliberalization succinctly captures an effort to "economize everything" [47, p. 171], such that neoliberal reason becomes common sense or simply the default rationale people use to make decisions. Metrics reinforce the notion that individual performance at work can be easily calculated and compared; therefore, material rewards like promotions and raises can be fairly and transparently applied. Of course, metrics often hide as much as they reveal as they simplify a process by carving away essential components. In universities, the metrics that help reinforce and are reinforced by neoliberal reason summarize entire careers for administrators who likely have little to no understanding of the research that the metrics summarize. This disempowers colleagues within and outside of a scholar's university, while centralizing power in the hands of bureaucrats with little incentive to actually understand academic work with which they are inexpert.

University administrators' use of metrics to evaluate faculty output is a fairly recent phenomenon. Prior to the development of bibliometric indicators, evaluation of

scholarly research was performed primarily by disciplinary specialists who offered qualitative assessment of a research record. While peer assessment is still a central part of evaluation processes, metrics such as citation counts, journal impact factors, and the h-index are now commonly incorporated into hiring and promotion decisions [48] and are seen as more important to more vulnerable untenured scholars than tenured ones [49]. Qualitative expert assessment that involves hours of engagement with the work summarized by metrics may be devalued by administrators motivated by a logic that privileges competition, central decision-making, and market valuation [47].

Critics have raised concerns about how metrics transform scholarship into a capitalist-like market at the core of neoliberalization resulting in "perverse incentives" for researchers to publish shoddy or fraudulent work [17], while simultaneously resulting in mental health trauma for academic workers experiencing hypercompetitive markets and suspicious management [50]. Administrators' continued reliance upon metrics serves as a modern form of Taylorism, a production principle ~~used in the~~ developed in the late nineteenth and early twentieth century that pursued technological solutions to the "problem" of worker-related inefficiencies on the shop floor with little regard to employee satisfaction or wellness [51]. By using technology to set the nature and pace of production, owners and managers gain greater control over the labor process itself. Prioritizing metrics creates a demand for quantity over quality, and by following these demands academic laborers risk surrendering some degree of control over their own labor processes.

The Case of the h-index

One key metric used for evaluation purposes is the h-index or Hirsch Index. Physicist Jorge Hirsch proposed the h-index as a "useful index to characterize the scientific output of a researcher" in a 2005 article in the Proceedings of the National Academy of Sciences [52]. While acknowledging the "potentially distasteful" use of metrics for evaluation, he presents quantification as an economical means of evaluating impact. In this highly cited article, Hirsch defines the h-index as follows: "A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ papers have $\leq h$ citations each" (p. 16569). In other words, a scholar with 10 of their 100 publications with a citation count of 10 or higher will have an h-index of 10. He goes on to specify - again with some acknowledgement that metrics offer a "rough approximation" of a research portfolio - how and when the index could be put to use: "Based on typical ~~h and m~~ h and m values found, I suggest (with large error bars) that for faculty at major research universities, ~~h 12~~ $h \approx 12$ might be a typical value for advancement to tenure (associate professor) and that $h \approx 18$ might be a typical value for advancement to full professor" (p. 16571). In sum, this publication announced a simple means of evaluating research impact and permission to use the metric for evaluation purposes.

The immediate response to the h-index was largely positive with features in top scientific journals; however, some criticism of the index also quickly appeared [53]. Critics identified a range of issues from the relationship between the h-index and career length as well as the effect of self-citation ~~[see 54, 55, among others]~~ [54, 55, among others]. However, the h-index and variants have proven enormously popular both in the bibliometrics and science of science communities and among university administrators seeking quick and cheap ways to evaluate scholars, including universities and science funding agencies [53]. The h-index is included as a key quantitative metric for annual review and/or tenure and promotion in faculty handbooks in a range of departments and schools in the United States (c.f., handbooks from the Boston University School of Public Health [56], the Ohio State University Department of Surgery [57] or Oregon State University's College of Business [58]).

Survey research in Germany on whether and how scholars understand the importance of the h-index indicate that natural scientists widely understand the importance of the h-index to their careers, but scholars in the humanities and social sciences do not [59]. This variation is unfortunate as quantitative metrics are widely applied in German universities as elsewhere. In sum, despite some criticism, the h-index has been widely adopted although perhaps not widely understood. This analysis contributes to the broader literature on metrics and science by examining the factors contributing to within and between discipline variation in the h-index.

Hypotheses

We develop the following hypotheses to better understand within and between discipline differences in the h-index, or Hirsch Index, based on the prior literature. Consistent with work on gender inequality and science and particularly work on the productivity and prestige puzzles, we examine the following:

Within Discipline Hypotheses

In light of research on team science and its impact on academic careers, we examine the following:

> *Female Penalty Hypothesis (H1):* Authors with names more common among women will have lower h-index scores, on average.
>
> *Sole Author Hypothesis (H2):* Authors with ~~more~~ a higher share of sole-authored publications will have lower h-index scores, on average.

Between Disciplines Hypotheses

Although less well understood, based on the research describing disciplinary differences in culture, such as propensity to collaborate, and material differences in resources, we examine ~~three~~ two related hypotheses:

> ~~*Disciplinary Differences*~~ *Feminized Discipline Hypothesis (H3a):* ~~A substantial amount of the variation in the h-index likely occurs between disciplines.~~
>
> ~~*Feminized Discipline Hypothesis (H3b):*~~ Disciplines with a higher share of women will have lower h-index scores, on average.
>
> *Teamwork Variation Hypothesis (~~H3e~~H3b):* Disciplines with a higher share of sole authorship will have lower h-index scores, on average.

Field-level Hypothesis

Finally, a field-level view may help disentangle cultural and compositional effects and shed light on the question of "apples-to-oranges" comparison within and between fields. We, therefore, evaluate the following hypothesis:

> *Field Variation Hypothesis (H4):* Significant field differences will exist in ~~terms of within and between disciplinary differences in~~ the relationship of the h-index to gender and coauthorship at both the individual and disciplinary level.

Materials and Methods

To evaluate these hypotheses, we analyze data on high performing scholars according to well-known metrics. Ioannidis and coauthors [8] collected author-level bibliometrics on 100,000 high performing scholars from the Scopus database. They updated this data through 2019 and expanded to include the top 2% of authors overall across a wide range of disciplines [60]. ~~We use the most up-to-date data for our analyses.~~ These data suffer from several limitations, including coverage differences between disciplines within the Scopus database [61, 62]. Nonetheless, these well-curated data represent a unique opportunity to evaluate within and between disciplinary differences in h-index scores. We also see these data as offering a conservative test of such differences as variation is likely to widen when moving beyond scholars who are in the top 2% on these metrics.

We reduce the data along several dimensions to address some of the limitations of the Scopus database and for analytic purposes. First, we limit the analysis only to the 68,016 scholars in the United States to account for geographic variation in the Scopus database and geographic variation in how universities are structured. We then further restrict the analytical data to those who first published in 1960 or after and those who last published in 2017 or later to identify active scholars and reduce the impact of historical, rather than contemporary, patterns. These reductions, along with a small number of missing values on the variables below, leave us a final analytical sample size of 54,825 scholars nested in 174 disciplines.

The dependent variable for the analysis is the h-index. The key independent variables are the probability of female name and the percentage of sole authored publications. We also include additional control variables of the count of scholars from the same university in the same dataset to measure highly productive university environments, a specialization score measured as the proportion of total articles appearing in the main discipline for each author, and career length, measured by the years between ~~first and last~~ each scholar's first and most recent publication.

Estimating gender is problematic for numerous reasons including the typical reliance on government-provided data that often assumes and contributes to the gender binary [63], and these methods should be used with caution and only when necessary. In this case, names are our only means of estimating gender. Here, we draw on first ~~names~~ name data from the 1940-1990 US Social Security Administration to assign a probability of female name using the gender package in R [64]. We use this probability directly in our models, rather than assigning an arbitrary cutoff for a binary gender assignment. A substantial number of cases (18%) are missing on this variable, usually because they were identified by initials rather than a full first name. Rather than lose this many cases, we use multiple imputation to assign values to the probability of female name based on the respondent's other variables, including discipline. We impute five complete datasets and pool analyses across these datasets for all of the models presented here.

We also consider how the patterns we observe may differ within large fields among disciplines. To explore this issue, we divide the total 174 disciplines into five large fields of the humanities, medical (including public health), professional, social sciences (including policy), and ~~STEM~~science, technology, engineering, and math (STEM). These fields differ from the fields provided in the original data but more closely correspond to the division of disciplines within the American academy. The supplementary materials show a full list of disciplines, which disciplines were assigned to each field, and summary statistics for each discipline.

Table 1 presents the descriptive statistics for ~~the~~ all variables used in the analyses for the total sample and across the five different fields. ~~Differences across fields in both~~

Table 1. Mean and standard deviation on key variables for the whole sample and by field, based on first complete dataset

~~Variable All Humanities Medical Prof. Soc. Sci. STEM H-index 44.1 (20.9) 19.8 (10.4) 52.2 (21.7) 29.4 (10.7) 33.1 (15.0) 38.8 (17.6) Percent female 19.1 (37.7) 30.8 (45.1) 22.1 (40.1) 20.7 (39.1) 28.3 (43.8) 13.7 (32.3) Percent sole author 10.7 (14.4) 56.0 (31.4) 8.1 (9.4) 18.2 (16.8) 24.6 (22.9) 9.1 (11.5) Career length 35.3 (10.1) 33.6 (11.2) 36.2 (9.3) 32.2 (9.9) 34.7 (10.6) 34.7 (10.6) Specialization 0.73 (0.19) 0.64 (0.21) 0.76 (0.18) 0.75 (0.21) 0.72 (0.19) 0.70 (0.20) University count 198 (202) 255 (204) 185 (201) 166 (203) 246 (209) 204 (200) N 54,825 856 25,581 1,228 3,987 23,173 Note: standard deviations shown in parenthesis~~

		Fields				
Variable	All	Humanities	Medical	Prof.	Soc. Sci.	STEM
h-index	44.1 (20.9)	19.8 (10.4)	52.2 (21.7)	29.4 (10.7)	33.1 (15.0)	38.8 (17.6)
% female	19.0 (37.6)	30.6 (45.0)	22.0 (40.0)	21.4 (39.5)	27.8 (43.6)	13.5 (32.0)
% sole author	10.7 (14.4)	56.0 (31.4)	8.1 (9.4)	18.2 (16.8)	24.6 (22.9)	9.1 (11.5)
career length	35.3 (10.1)	33.6 (11.2)	36.2 (9.3)	32.2 (9.9)	34.7 (10.6)	34.7 (10.6)
specialization	0.73 (0.19)	0.64 (0.21)	0.76 (0.18)	0.75 (0.21)	0.72 (0.19)	0.70 (0.20)
uni. count	198 (202)	255 (204)	185 (201)	166 (203)	246 (209)	204 (200)
N	54,825	856	25,581	1,228	3,987	23,173

Note: standard deviations shown in parenthesis

~~the *h*-index and independent variables are notable. For example, scholars in the humanities have the highest average probability of a female name at 31.1%, while scholars in the STEM field have the lowest average probability at 13.5%. Scholars in the humanities also have the highest percentage of sole authored work at 57.8% while scholars in the medical fields have the lowest with only 8.1%. The much lower mean *h*-index among scholars in the humanities (19.1) relative to other disciplines is consistent with the expectation of gender differences and a penalty for sole-authored work. However, to more accurately gauge these relationships we turn to a modeling strategy that can estimate associations both within and between disciplines.~~

Analytic Strategy

We model variation in *h*~~-score indices~~ -index scores across scholars using random effects within-between (REWB) models [65]. ~~REWB~~ Random effects within-between (REWB) models are a variant of multilevel models that allows researchers to estimate the effect of a given predictor variable both within (as per a standard "fixed effects" model) and between the higher level clusters (in this case disciplines). In general, the structure of the REWB model is:

$$y_{ij} = \beta_{0j} + \beta_1(x_{ij} - \bar{x}_{.j}) + \beta_2(\bar{x}_{.j}) + \upsilon_{0j} + \epsilon_{ij}$$

Where y_{ij} is the outcome for the ith unit in the jth cluster and x_{ij} is the predictor variable for the ith unit in the jth cluster. υ_{0j} and ϵ_{ij} are cluster-level and individual-level random errors, respectively. Because the mean of ~~the~~ x for cluster j ($\bar{x}_{.j}$) is included in the model and ~~the~~ x_{ij} values are mean centered by cluster, the β_1 parameter is identical to that of a fixed-effects model in which all between variance is absorbed by cluster-level ~~dummies~~dummy variables. However, the REWB model also includes an estimate of the between cluster effect of x estimated in β_2, which is impossible in a fixed-effects model. This β_2 term is equivalent to the estimate obtained by aggregating data to the higher level and examining the relationship between the means of x and y. Thus, this model ~~combines~~ maintains the advantage of traditional fixed effects models by absorbing all cluster level differences when estimating ~~individual~~

~~(lr)3-7Grouping All Humanities Medical Prof. Soc. Sci STEM Between discipline 32.0% 53.6% 20.4% 16.2% 31.3% 20.1% Within discipline 68.0% 46.4% 79.6% 83.8% 68.7% 79.9%~~

Table 2. Partition of variance in h-index scores between and within disciplines

		Field				
Grouping	All (%)	Humanities (%)	Medical (%)	Prof. (%)	Soc. Sci. (%)	STEM (%)
Between discipline	32.0	53.6	20.4	16.2	31.3	20.1
Within discipline	68.0	46.4	79.6	83.8	68.7	79.9

~~level effects~~lower level parameters, while at the same time allowing for an analysis of the "contextual" ~~effects of disciplines themselves~~relationships at the higher level.

We use this feature to estimate both within and between effects of gender and sole authorship on a scholar's h-index. ~~We use 0-1 coding of~~ Both the probability of a female name variable and the sole authorship variable are mean centered by discipline at the individual level ~~to mimic the interpretation typical of a dummy variable on gender. We use the~~ so that the interpretation of their effect is solely among scholars within the same discipline. For the between discipline effects, we calculate the mean probability of a female name ~~in percentage terms (0-100)~~ and the mean sole authorship of each discipline. For probability of a female name, we use a 0 to 100 percentage scale at the disciplinary level ~~to provide an estimate of the feminization of a given discipline~~for ease of interpretation of coefficients as the expected change in the h-index for a one percentage point increase in the mean probability of a female name. For sole authorship, we use ~~the percent of a scholar's articles that are sole-authored at the individual level. At the disciplinary level, we use a~~ a natural log transformation of the mean percent sole-authorship variable at the disciplinary level because exploratory analysis indicated a negative diminishing returns relationship between sole-authorship and the h-index at this level. ~~Both of the individual level variables are mean centered by discipline so that the interpretation of their effect is solely among scholars within the same discipline.~~

The control variables of career length, specialization, and university publication count are only included as grand mean centered individual level variables. We also standardize university publication count to have a mean of zero and a standard deviation of one for ease of interpretation.

In addition to the models for the full data, we also run these same models separately for each of these fields to explore differences in patterns across fields.

Results

We begin by analyzing a partition of the variance in the h-index within and between disciplines in a null multilevel model. The percentage of the total variation in the h-index that occurs between disciplines is given by the intraclass correlation coefficient (ICC) of this model. Table 2 shows the ICC for the model across all observations as well as separately by field. In total, roughly a third of the variation in the h-index occurs between disciplines. ~~This substantial heterogeneity across disciplines indicates that comparing the h-index across disciplines is problematic because of overall differences in culture, productivity, and citation patterns across disciplines.~~

~~Parition of variance in h-indexscores between and within disciplines~~

The ICC is substantial across all fields as well, but also varies substantially. Slightly more than half the variation in the h-index among scholars in the humanities is between

Table 3. Multilevel models predicting h-index score across all disciplines with clustering at the disciplinary level.
~~Model 1 Model 2 Model 3 Model 4 Intercept 39.32* 69.64* 68.89* 68.95* (1.77) (2.12) (2.13) (2.06) Prob. 0-1of female name† −2.95* −3.64* −2.60* (0.23) (0.22) (0.23) Disc. mean percent female name −0.15* 0.10* 0.11* (0.07) (0.05) (0.05) Percent sole authored pubs† −0.41* −0.42* −0.47* (0.01) (0.01) (0.01) Disc. mean percent sole authored (log) −12.87* −13.36* −13.50* (0.79) (0.82) (0.80) Career length* 0.35* (0.01) Specialization 0-1* −2.66* (0.44) Highly productive uni. count (stdized)* 1.72* (0.07) Residual SD between discipline 11.86 7.41 7.34 7.06 Residual SD within discipline 17.41 16.83 16.78 16.34 N (discipline) 174 174 174 174 N (individual) 54825 54825 54825 54825~

	Model 1	Model 2	Model 3	Model 4
Intercept	39.32*	69.64*	68.86*	68.91*
	(1.80)	(2.12)	(2.13)	(2.06)
Prob. [0-1] of female name†	−3.00*		−3.67*	−2.63*
	(0.23)		(0.22)	(0.21)
Disc. mean percent female name	−0.15*		0.10*	0.11*
	(0.07)		(0.05)	(0.05)
Percent sole authored pubs†		−0.41*	−0.42*	−0.47*
		(0.01)	(0.01)	(0.01)
Disc. mean percent sole authored (log)		−12.87*	−13.39*	−13.53*
		(0.79)	(0.82)	(0.79)
Career length*				0.35*
				(0.01)
Specialization [0-1]*				−2.66*
				(0.44)
Highly productive uni. count (stdized)*				1.72*
				(0.07)
Residual SD between discipline	11.86	7.41	7.33	7.05
Residual SD within discipline	17.41	16.83	16.78	16.34
N (discipline)	174	174	174	174
N (individual)	54825	54825	54825	54825

$^*p < 0.05$. Standard errors shown in parenthesis. †=discipline mean centered; *=grand mean centered.

disciplines, while only one-fifth or less of the variation in the h-index is between disciplines among scholars in the medical, STEM, and professional fields. ~~Some fields, like the medical, STEM, and professional fields, seem to have more shared culture and practices in terms of publishing productivity, while in other fields, we observe substantial disciplinary heterogeneity. Scholars in fields with greater commonality may mispercieve the comparability of the h-index across disciplines more broadly.~~

Table 3 presents the multilevel models predicting the h-index across all disciplines. Model 1 predicts h-index scores by the individual and disciplinary variables for gender. Model 2 predicts h-index scores by the individual and disciplinary variables for sole authorship. Model 3 includes both sets of predictor variables from Models 1 and 2 together. Finally, Model 4 includes additional control variables for career length, specialization, and highly productive universities. The estimates for the key predictor variables of gender and sole-authorship are robust to these additional controls, although gender differences do decline in size somewhat. We use the results from Model 4 to describe overall patterns below, except where noted otherwise.

Consistent with the productivity puzzle, female scholars have an h-index approximately ~~2.6~~ 2.63 points lower than male scholars in the same discipline, on average. More frequent sole-authorship is also associated with a lower h-index. Within

the same discipline, a one percentage point increase in the percent of sole-authored publications for a given scholar is associated with a 0.47 lower h-index score, on average.

Additionally, we find substantial differences between disciplines in h-index scores based on the feminization of the discipline and sole-authorship norms. Sole authorship behaves as expected. A one percent increase in the mean percent sole-authored publications in a discipline is associated with a 0.135 decline in the mean h-index for that discipline. Thus, cultural norms of more sole-authorship within a discipline contribute to lower overall h-index scores for that discipline.

The feminization of a discipline, as indicated by the mean percent female name within the discipline, has a more complex relationship to h-index scores. Model 1 shows a slightly negative ~~relationship~~ association between percent female and mean h-index scores across disciplines. However, ~~once the sole-authored variables were added to the model, this relationship reverses direction and~~ this slightly negative association is a spurious byproduct of the underlying tendency for more feminized disciplines to also be more focused on sole-authorship ($r = 0.30$). Model 3 shows that, once we hold constant the disciplinary tendency for sole-authorship, the association between feminization and h-index scores becomes slightly positive. In the final model, we estimate that a one percentage point increase in the percent female name within a discipline is associated with a 0.11 increase in the mean h-index score for ~~a discipline. The underlying issue is that highly feminized disciplines also tend to be more focused on~~ that discipline. Thus, when we compare disciplines with similar sole-authorship ~~($r = 0.30$). After accounting for this disadvantage~~tendencies, feminized disciplines actually have a slight advantage ~~, although~~ in mean h-index scores, even though, within disciplines, women are still disadvantaged relative to men.

Table 4 presents models equivalent to Model 4 in Table 3, but separated by field. The most notable change in these sets of models is that there is no observable effect of disciplinary feminization within ~~these fields, whereas~~ all fields, except for the Social Sciences. In contrast, we observed a moderate positive effect of disciplinary feminization ~~across all disciplines~~when we pool all fields. This finding implies that the slight positive effect of disciplinary feminization was driven by compositional issues between fields and differences in overall field productivity. The remaining variables are consistent in direction, but vary substantially in magnitude across fields. The within-discipline differences by gender are smallest in the humanities and STEM fields and largest in the medical field. Similarly, the within-discipline differences by sole-authorship are smallest in the humanities and largest in the medical field. The between-discipline effect of sole-authorship is largest in the medical field and smallest in the professional field where it is only a third as large. ~~Overall, the heterogeneity of these effects across fields indicates another problem of comparability when attempting to compare scholars across disciplines and fields.~~

Discussion

The substantial heterogeneity across disciplines (roughly a third of all variation) indicates that comparing the h-index ~~, or Hirsch Index, is a widely used metric used for performance evaluation or quality valuation in the sciences and across the academy. This research aimed to contribute to the literature on bibliometrics and inequality in science by examining both within and between discipline differences in the~~ across disciplines is problematic because of overall differences in culture, productivity, and citation patterns across disciplines. Additionally, we found that fields vary in the level of disciplinary heterogeneity. Some fields, like the medical, STEM, and professional

Table 4. Multilevel models predicting h-~~indexscore~~ index score within each field. Clustering at the disciplinary level.

~~Humanities Medical Prof. Soc. Sci. STEM Intercept 79.11* 93.03* 47.21* 77.62* 60.05* (12.61) (6.37) (8.33) (4.57) (5.04) Prob. 0-1of female name† −1.35* −3.67* −1.68* −2.00* −1.33* (0.51) (0.33) (0.72) (0.44) (0.38) Disc. mean percent female name −0.00 0.06 0.01 −0.13* 0.06 (0.08) (0.08) (0.21) (0.05) (0.17) Percent sole authored pubs† −0.16* −0.66* −0.26* −0.24* −0.49* (0.01) (0.01) (0.02) (0.01) (0.01) Disc. mean percent sole authored (log) −15.40* −21.73* −6.94* −13.62* −10.79* (3.09) (3.03) (3.46) (1.20) (1.81) Career length* 0.13* 0.45* 0.29* 0.31* 0.30* (0.02) (0.01) (0.03) (0.02) (0.01) Specialization 0-1* −3.16* 0.03 −0.49 −6.32* −4.37* (1.10) (0.79) (1.55) (1.03) (0.57) Highly productive uni. count (stdized)* 0.55* 1.51* 0.60* 0.37* 2.16* (0.21) (0.12) (0.29) (0.18) (0.10) Residual SD between discipline 4.52 6.71 4.27 3.14 6.23 Residual SD within discipline 5.98 18.69 9.48 11.31 14.41 N (discipline) 15 52 12 28 67 N (individual) 856 25581 1228 3987 23173~~

	Humanities	Medical	Prof.	Soc. Sci.	STEM
Intercept	79.45*	93.04*	47.21*	77.81*	58.73*
	(12.68)	(6.37)	(8.23)	(4.57)	(5.18)
Prob. [0-1] of female name†	−1.30*	−3.73*	−1.57*	−2.11*	−1.25*
	(0.50)	(0.31)	(0.76)	(0.43)	(0.34)
Disc. mean percent female name	−0.02	0.07	−0.01	−0.13*	0.12
	(0.09)	(0.08)	(0.21)	(0.05)	(0.17)
Percent sole authored pubs†	−0.16*	−0.66*	−0.26*	−0.24*	−0.49*
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
Disc. mean percent sole authored (log)	−15.38*	−21.79*	−6.84*	−13.66*	−10.57*
	(3.08)	(3.02)	(3.34)	(1.19)	(1.83)
Career length*	0.13*	0.45*	0.29*	0.31*	0.30*
	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)
Specialization [0-1]*	−3.13*	0.03	−0.46	−6.31*	−4.36*
	(1.10)	(0.79)	(1.55)	(1.03)	(0.57)
Highly productive uni. count (stdized)*	0.55*	1.51*	0.60*	0.38*	2.16*
	(0.21)	(0.12)	(0.29)	(0.18)	(0.10)
Residual SD between discipline	4.51	6.70	4.27	3.13	6.22
Residual SD within discipline	5.98	18.69	9.48	11.31	14.41
N (discipline)	15	52	12	28	67
N (individual)	856	25581	1228	3987	23173

*$p < 0.05$. Standard errors shown in parenthesis. †=discipline mean centered; *=grand mean centered.

fields, seem to have more shared culture and practices in terms of publishing productivity, while in other fields, we observe substantial disciplinary heterogeneity. Scholars in fields with greater commonality may misperceive the comparability of the h-index. We used REWB models to predict the h-index for high-performing scholars in 174 disciplines . across disciplines more broadly.

Results indicate that, within disciplines, gender and sole authorship affect the h-indexproviding support for our female penalty (. Consistent with H1) and sole authorship hypotheses (and H2). , we find that women have lower h-index scores than men and that more sole-authorship is associated with lower h-index scores, respectively. Thus, among scholars in the same discipline, variation in both ascriptive characteristics such as gender and research practices can generate substantial inequalities in h-index scores.

We find evidence of between discipline differences as well. Results provide support for the disciplinary differences hypothesis (H3a), although the extent of disciplinary differences varies by field. Significant differences also exist between disciplines in terms of sole authorship, supporting the Disciplines with more overall sole-authorship have substantially lower mean h-index scores overall, consistent with the teamwork variation hypothesis (H3cH3b). Support for the feminized discipline hypothesis (H3b) is mixed. Results indicate that a small negative feminized discipline effect reverses when including sole authorship and that H3a) is more mixed. In the full models, results vary by whether disciplinary sole-authorship practice is included in the model because feminized disciplines are more likely to be high sole-authorship disciplines as well. In field specific models, we find a negative feminization effect on h-index scores only for the Social Sciences, and on effect for the remaining disciplines, even when controlling for disciplinary sole-authorship.Thus, to the extent, that the feminization effect exists, it appears limited to certain fields.

We also observe that both the within discipline and between discipline effects likely result from compositional differences across fieldsconfirming vary substantially across fields, consistent with the field variation hypothesis (H4). In sum, important differences exist This heterogeneity in models across fields implies an additional apples-to-oranges comparison problem when comparing scholars across fields. Even if some attention is paid to the overall predictors of productivity within and between disciplines in the h-index. Moreover, this analysis provides evidence that across field evaluation with the h-index risks apples-to-oranges comparisonwhen evaluating scholars, field heterogeneity in the effects of these predictors will complicate the ability to make accurate comparisons.

Data limitations suggest several avenues for future research. First, the data select on high-performing scholars, and, therefore do not generalize to the broader population of academics. While we believe that this limitation likely results in a conservative estimate of within and between discipline inequality, more data on academia writ large are required to verify this claim. Second, the estimation of gender suffers from well-known limitations. Data linking self-reported gender beyond the gender binary to citation data would be a welcome resource. Gender estimation also required reducing the dataset by country. Third, and along similar lines, most scholarship on publication and inequality focuses on gender as because it is possible , despite these weaknesses to infer gender from namesand other , despite some weaknesses in the approach. Other aspects of inequality, like race and class, remain under-studied. Future research would benefit from data linking self-reported race and class to large bibliometric data to better understand the broad effects of inequality in academic work.

In conclusion, this

Conclusion

This analysis provides further evidence that metrics in performance evaluation in academia should be used with caution, if at all. Metrics are subject to within and between discipline biases that severely hinder their value in making both intradisciplinary and interdisciplinary comparisons. Of course, these comparisons are exactly what the quantification of scholarly work proposes to facilitate. Like other forms of quantification, academic metrics simplify complex processes at a cost. This cost can reinforce existing inequalities and even generate new "automated inequalities" [66]. Those using metrics for evaluation, as Koopman and Galton [67] ~~writes~~ write about data usage generally, "need to be fervently attentive to the ways in which inequalities may be designed into their data" (pg. 16). In this vein, this analysis ~~provided~~ provides evidence of within and between discipline differences in the h-index with a focus on gender ~~, but there is reason to believe that the bias embedded~~ and sole authorship, but variation in scholarly metrics ~~extends~~ may extend to other dimensions of power, including race, class, and sexual orientation, or be related to other academic practices, like research funding, teaching loads, or service responsibilities.

Acknowledgements

~~the authors have no acknowledgements to provide~~We thank Nicholas Theis and jimi adams for their helpful comments on previous drafts of this article.

References

1. Hofstra B, Kulkarni VV, Munoz-Najar Galvez S, He B, Jurafsky D, McFarland DA. The Diversity~~–~~Innovation Paradox in Science. Proceedings of the National Academy of Sciences. 2020;117(17):9284–9291.
2. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR. Bibliometrics: Global Gender Disparities in Science. Nature. 2013;504(7479):211–213.
3. Xie Y, Shauman KA. Sex Differences in Research Productivity: New Evidence about an Old Puzzle. American Sociological Review. 1998;63(6):847–870. doi:10.2307/2657505.
4. Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, et al. Science of Science. Science. 2018;359(6379):eaao0185. doi:10.1126/science.aao0185.
5. Gardner SK. Conceptualizing Success in Doctoral Education: Perspectives of Faculty in Seven Disciplines. The Review of Higher Education. 2009;32(3):383–406. doi:10.1353/rhe.0.0075.
6. Stephan P. How Economics Shapes Science. Harvard University Press; 2012.
7. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: The Leiden Manifesto for Research Metrics. Nature. 2015;520(7548):429–431. doi:10.1038/520429a.
8. Ioannidis JPA, Baas J, Klavans R, Boyack KW. A Standardized Citation Metrics Author Database Annotated for Scientific Field. PLOS Biology. 2019;17(8):e3000384. doi:10.1371/journal.pbio.3000384.
9. Fox MF. Gender, Family Characteristics, and Publication Productivity among Scientists. Social Studies of Science. 2005;35(1):131–150.

10. Fox MF, Whittington K, Linkova M. Gender,(in) Equity, and the Scientific Workforce. Handbook of science and technology studies. 2017; p. 701–731.

11. ~~Long JS, Fox MF. Scientific Careers: Universalismand Particularism. Annual review of sociology. 1995;21(1):45–71.~~
~~Fox MF. Gender, Family Characteristics, and Publication Productivity among Scientists. Social Studies of Science. 2005;35(1):131–150.~~

12. Grant L, Kennelly I, Ward KB. Revisiting the Gender, Marriage, and Parenthood Puzzle in Scientific Careers. Women's Studies Quarterly. 2000;28(1/2):62–85.

13. Long JS, Fox MF. Scientific Careers: Universalism and Particularism. Annual review of sociology. 1995;21(1):45–71.

14. Morgan AC, Way SF, Hoefer MJD, Larremore DB, Galesic M, Clauset A. The Unequal Impact of Parenthood in Academia. Science Advances. 2021;7(9):eabd1996. doi:10.1126/sciadv.abd1996.

15. Katz Y, Matter U. Metrics of Inequality: The Concentration of Resources in the U.S. Biomedical Elite. Science as Culture. 2020;29(4):475–502. doi:10.1080/09505431.2019.1694882.

16. Mennicken A, Espeland WN. What's New with Numbers? Sociological Approaches to the Study of Quantification. Annual Review of Sociology. 2019;45:223–245.

17. Edwards MA, Roy S. Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. Environmental Engineering Science. 2017;34(1):51–61. doi:10.1089/ees.2016.0223.

18. Muller JZ. The Tyranny of Metrics. Princeton University Press; 2019.

19. Siler K, Larivière V. Who Games Metrics and Rankings? Institutional Niches and Journal Impact Factor Inflation. Research Policy. 2022;51(10):104608.

20. Bihari A, Tripathi S, Deepak A. A Review on H-Index and Its Alternative Indices. Journal of Information Science. 2023;49(3):624–665. doi:10.1177/01655515211014478.

21. ~~Fox MF. Gender, Science, and Academic Rank: KeyIssues and Approaches. Quantitative Science Studies. 2020;1(3):1001–1006.~~

22. Cech E, Rubineau B, Silbey S, Seron C. Professional Role Confidence and Gendered Persistence in Engineering. American Sociological Review. 2011;76(5):641–666. doi:10.1177/0003122411420815.

23. Ma Y, Xiao S. Math and Science Identity Change and Paths into and out of STEM: Gender and Racial Disparities. Socius. 2021;7:23780231211001978. doi:10.1177/23780231211001978.

24. Penner AM, Willer R. Men's Overpersistence and the Gender Gap in Science and Mathematics. Socius. 2019;5:2378023118821836. doi:10.1177/2378023118821836.

25. Weeden KA, Thébaud S, Gelbgiser D. Degrees of Difference: Gender Segregation of US Doctorates by Field and Program Prestige. Sociological Science. 2017;4:123–150.

26. Light R. Gender Stratification and Publication in American Science: Turning the Tools of Science Inward. Sociology Compass. 2009;3(4):721–733.

27. King MM, Frederickson ME. The Pandemic Penalty: The Gendered Effects of COVID-19 on Scientific Productivity. Socius. 2021;7:23780231211006977. doi:10.1177/23780231211006977.

28. Cole J, Zuckerman H. The Productivity Puzzle. Advances in Motivation and Achievement. 1984; p. 217–258.

29. Leahey E. Gender Differences in Productivity: Research Specialization as a Missing Link. Gender & Society. 2006;20(6):754–780. doi:10.1177/0891243206293030.

30. Leahey E. Not by Productivity Alone: How Visibility and Specialization Contribute to Academic Earnings. American Sociological Review. 2007;72(4):533–561. doi:10.1177/000312240707200403.

31. Huang J, Gates AJ, Sinatra R, Barabási AL. Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines. Proceedings of the National Academy of Sciences. 2020;117(9):4609–4616. doi:10.1073/pnas.1914221117.

32. Light R. Gender Inequality and the Structure of Occupational Identity: The Case of Elite Sociological Publication. In: Networks, Work and Inequality. vol. 24. Emerald Group Publishing Limited; 2013. p. 239–268.

33. ~~Light R. Gender Stratification and Publication in AmericanScience: Turningthe Tools of Science Inward. Sociology Compass. 2009;3(4):721–733.~~

34. Lynn FB, Noonan MC, Sauder M, Andersson MA. A Rare Case of Gender Parity in Academia. Social forces. 2019;98(2):518–547.

35. Moody J. The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999. American sociological review. 2004;69(2):213–238.

36. Jadidi M, Karimi F, Lietz H, Wagner C. Gender Disparities in Science? Dropout, Productivity, Collaborations and Success of Male and Female Computer Scientists. Advances in Complex Systems. 2018;21(03n04):1750011.

37. Akbaritabar A, Squazzoni F. Gender Patterns of Publication in Top Sociological Journals. Science, Technology, & Human Values. 2021;46(3):555–576. doi:10.1177/0162243920941588.

38. Teele DL, Thelen K. Gender in the Journals: Publication Patterns in Political Science. PS: Political Science & Politics. 2017;50(2):433–447.

39. Cetina KK. Epistemic Cultures: How the Sciences Make Knowledge. Harvard University Press; 1999.

40. Lamont M. How Professors Think: Inside the Curious World of Academic Judgment. Harvard University Press; 2009.

41. Koppman S, Cain CL, Leahey E. The Joy of Science: Disciplinary Diversity in Emotional Accounts. Science, Technology, & Human Values. 2015;40(1):30–70.

42. Parish AJ, Boyack KW, Ioannidis JPA. Dynamics of Co-Authorship and Productivity across Different Fields of Scientific Research. PLOS ONE. 2018;13(1):e0189742. doi:10.1371/journal.pone.0189742.

43. Lanahan L, Graddy-Reed A, Feldman MP. The Domino Effects of Federal Research Funding. PLOS ONE. 2016;11(6):e0157325. doi:10.1371/journal.pone.0157325.

44. Lynn FB, Espy HW. 15. Cumulative Advantage. Research Handbook on Analytical Sociology. 2021; p. 286.

45. Snow CP. The Two Cultures. Cambridge University Press; 2012.

46. Bi HH. Four Problems of the H-Index for Assessing the Research Productivity and Impact of Individual Authors. Scientometrics. 2023;128(5):2677–2691. doi:10.1007/s11192-022-04323-8.

47. Berg LD, Huijbens EH, Larsen HG. Producing Anxiety in the Neoliberal University. The Canadian Geographer/le ~~géographe~~ géographe canadien. 2016;60(2):168–180.

48. McKiernan EC, Schimanski LA, Nieves CM, Matthias L, Niles MT, Alperin JP. Meta-Research: Use of the Journal Impact Factor in Academic Review, Promotion, and Tenure Evaluations. Elife. 2019;8:e47338.

49. Niles MT, Schimanski LA, McKiernan EC, Alperin JP. Why We Publish Where We Do: Faculty Publishing Values and Their Relationship to Review, Promotion and Tenure Expectations. PLOS ONE. 2020;15(3):e0228914. doi:10.1371/journal.pone.0228914.

50. Forrester N. Mental Health of Graduate Students Sorely Overlooked. Nature. 2021;595(7865):135–137. doi:10.1038/d41586-021-01751-z.

51. Braverman H. Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century. nyu Press; 1998.

52. Hirsch JE. An Index to Quantify an Individual's Scientific Research Output. Proceedings of the National Academy of Sciences. 2005;102(46):16569–16572. doi:10.1073/pnas.0507655102.

53. Barnes C. The H-Index Debate: An Introduction for Librarians. The Journal of Academic Librarianship. 2017;43(6):487–494. doi:10.1016/j.acalib.2017.08.013.

54. Kelly CD, Jennions MD. The h Index and Career Assessment by Numbers. Trends in Ecology & Evolution. 2006;21(4):167–170. doi:10.1016/j.tree.2006.01.005.

55. Purvis A. The h Index: Playing the Numbers Game. Trends in Ecology & Evolution. 2006;21(8):422. doi:10.1016/j.tree.2006.05.014.

56. University B. Boston University School of Public Health Faculty Handbook; 2018.

57. University OS. Department of Surgergy, Appointments, Promotion, and Tenure; 2014.

58. University OS. Oregon State University College of Business Faculty Handbook; 2020.

59. Kamrani P, Dorsch I, Stock WG. Do Researchers Know What the H-Index Is? And How Do They Estimate Its Importance? Scientometrics. 2021;126(7):5489–5508. doi:10.1007/s11192-021-03968-1.

60. Ioannidis JPA, Boyack KW, Baas J. Updated Science-Wide Author Databases of Standardized Citation Indicators. PLOS Biology. 2020;18(10):e3000918. doi:10.1371/journal.pbio.3000918.

61. Mongeon P, Paul-Hus A. The Journal Coverage of Web of Science and Scopus: A Comparative Analysis. Scientometrics. 2016;106(1):213–228. doi:10.1007/s11192-015-1765-5.

62. Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The Journal Coverage of Web of Science, Scopus and Dimensions: A Comparative Analysis. Scientometrics. 2021;126(6):5113–5142.

63. Mihaljević H, Tullney M, Santamaría L, Steinfeldt C. Reflections on Gender Analyses of Bibliographic Corpora. Frontiers in big Data. 2019;2:29.

64. Blevins C, Mullen L. Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction. DHQ: Digital Humanities Quarterly. 2015;9(3).

65. Bell A, Fairbrother M, Jones K. Fixed and Random Effects Models: Making an Informed Choice. Quality & quantity. 2019;53:1051–1074.

66. Eubanks V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press; 2018.

67. Koopman C. From Galton's Pride to Du Bois's Pursuit: The Formats of Data-Driven Inequality. Theory, Culture & Society. 2023; p. 02632764231162251.

Supporting Information

S1 Table Descriptive statistics by discipline.