



# BMJ Open Authorship inequality: a bibliometric study of the concentration of authorship among a diminishing number of individuals in high-impact medical journals, 2008–2019

Kamber L Hart <sup>1,2</sup>, Roy H Perlis <sup>1,2</sup>

**To cite:** Hart KL, Perlis RH.

Authorship inequality: a bibliometric study of the concentration of authorship among a diminishing number of individuals in high-impact medical journals, 2008–2019. *BMJ Open* 2021;**11**:e046002. doi:10.1136/bmjopen-2020-046002

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-046002>).

Received 30 October 2020  
Revised 06 December 2020  
Accepted 15 December 2020



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA

<sup>2</sup>Center for Quantitative Health, Massachusetts General Hospital, Boston, MA, USA

## Correspondence to

Professor Roy H Perlis;  
[rperlis@partners.org](mailto:rperlis@partners.org)

## ABSTRACT

**Objective** Authorship and number of publications are important criteria used for making decisions about promotions and research funding awards. Given the increase in the number of author positions over the last few decades, this study sought to determine if there had been a shift in the distribution of authorship among those publishing in high-impact academic medical journals over the last 12 years.

**Design** This study analysed the distribution of authorship across 312 222 original articles published in 134 medium-impact to high-impact academic medical journals between 1 January 2008 and 31 December 2019. Additionally, this study compared the trends in author distributions across nine medical specialties and a collection of cross-specialty high-impact journal articles.

**Primary outcome measures** The distribution of authorship was assessed using the Gini coefficient (GC), a widely used measure of economic inequality.

**Results** The overall GC for all articles sampled across the 12-year study period was 0.49, and the GCs for the first and last authorship positions were 0.30 and 0.44, respectively. Since 2008, there was a significant positive correlation between year and GC for the overall authorship position ( $r=0.99$ ,  $p<0.001$ ) the first author position ( $r=0.75$ ,  $p=0.007$ ) and the last author position ( $r=0.85$ ,  $p<0.001$ ) indicating increasingly uneven distribution in authorship over time. The cross-specialty high-impact journals exhibited the greatest rate of increase in GC over the study period for the first and last author position of any specialty analysed.

**Conclusion** Overall, these data suggest a growing inequality in authorship across authors publishing in high-impact academic medical journals, especially among the highest impact journals. These findings may have implications for processes such as promotions and allocation of research funding that use authorship metrics as key criteria for making decisions.

## INTRODUCTION

Authorship and number of publications are important criteria used for making decisions about promotions and research funding awards. However, over the last few

## Strengths and limitations of this study

- This bibliometric study analyses 312 222 original research articles published in 134 medium-impact to high-impact academic medicine journals from 2008 to 2019.
- The distribution of authorship can be easily captured and understood by using the Gini coefficient, commonly used to measure economic inequality.
- Additional sensitivity analyses were conducted to assess the impact of the number of available author positions, consortia publications, number of articles published per year and number of articles sampled per specialty on the distribution of authorship.
- Automated sampling of articles and author names makes it challenging to correctly link authors across articles.
- Further work is needed to understand the effects of disparity in publication, as it may be an indicator of differences in productivity or lack of equity in access to funding resources.

decades there has been a well-documented increase in the size of the author list on peer-reviewed medical journal articles.<sup>1</sup> Referred to as ‘authorship inflation’, this increase in the number of authors may be due to the growing complexity of medical research and an increased number of collaborations and consortium teams.<sup>1</sup> Alternatively, authorship inflation may also be due to honorary authorship. This could include the desire to include more senior, well-established authors in the hope that they will help facilitate the acceptance and publication of a manuscript.<sup>2</sup> Additionally, and potentially more insidiously, senior authors could demand inclusion on manuscripts even without significant contribution, a concept known as the White Bull effect.<sup>3</sup>

In 2003, the International Committee of Medical Journals (ICMJE) implemented

stricter guidelines to distinguish authorship from other forms of contribution to research, and has further adjusted these recommendations in the subsequent years.<sup>4</sup> Yet even with these recommendations, the number of author positions continues to grow at a rate that is inadequately explained by the growing complexity of biomedical research.<sup>1,4,5</sup>

Prior work has not determined if this increase in the number of author positions has shifted the distribution of authorship across individuals. In other words, has the increase in author positions increased the number of publications across all authors, or has it increased the number of publications associated with a few individuals?

Given the importance of authorship for academic promotion and obtaining research funding, a better understanding of trends in author distributions would help to contextualise and inform decisions that rely on authorship as a metric. More generally, the extent to which authorship is concentrated among a small number of individuals may inform us about the distribution of authority in a given specialty—that is, to what extent can a modestly sized group of individuals impact an area of medicine?

The distribution of authorship can be easily captured and understood by using the Gini coefficient (GC).<sup>6</sup> Commonly used to measure economic inequality, the GC measures the distribution of a certain resource across a given population.<sup>7</sup> The coefficient ranges from 0 to 1 with 0 indicating complete equality across individuals sampled, and 1 indicating complete inequality, or 1 individual holding all of the resources within the sampled cohort. Therefore, a smaller GC indicates a more equal distribution of resources across individuals, and a larger GC indicates a less equal distribution of resources among individuals.

While, to the best of our knowledge, the GC has not been previously applied to authorship, GCs have been used to assess distribution of resources outside of economics. Within healthcare, this includes assessments of disparities in geographical distributions of healthcare resources as well as the distribution of health-related quality of life across time and different cohorts within the USA.<sup>8,9</sup> In this manuscript, we extend the use of the GC to summarise the distribution of authorship within academic medicine, assessing if a few authors are publishing more frequently (a high GC) or if authorship positions are spread more equally across individuals (a low GC). These analyses also sought to better understand changes in the distribution of authorship among high-impact academic medical journals over the last 12 years. Therefore, we aimed to (1) assess the overall distribution of authorship within high-impact academic medical journals, especially for the meaningful first and last author positions, (2) evaluate trends in author distributions over the last 12 years, and (3) examine and compare the trends in author distributions across medical specialties.

## METHODS

### Article selection

This study used author information from 312 222 original articles published between 1 January 2008 and 31 December 2019. The methods for selecting journals and compiling these articles have been previously published.<sup>10</sup> To briefly summarise, we identified the top 15 journals by impact factor for 9 medical specialties (paediatrics, radiology, anaesthesiology, obstetrics and gynaecology, neurology, general medicine, dermatology, psychiatry and oncology) along with 4 additional high-impact cross-specialty journals (*New England Journal of Medicine*, *Journal of the American Medical Association*, *British Medical Journal*, and *The Lancet*; online supplemental table S1).<sup>11</sup> Metadata for articles published during the above time period were scraped from PubMed, and then the article list was limited to original research articles. Other article types, including comments, editorials, review articles, guidelines and retracted articles, were excluded. Article and author information for original research articles were compiled by scraping the extensible markup language (XML) code for each article using the RISmed package (V.2.1.7) and the EasyPubMed package (V.2.5) in R (V.3.5.1).<sup>12</sup>

Five journals (*Ultrasound in Obstetrics & Gynecology*; *Seminars in Perinatology*; *Clinics in Perinatology*; *Journal of the American Academy of Child and Adolescent Psychiatry*; *Journal of Neurology, Neurosurgery, and Psychiatry*) were included among the top 15 journals by impact factor for more than one specialty. The author lists from articles published in these journals were duplicated only for the analyses of author distribution separated by specialty.

### Data cleaning

Author names were cleaned in order to standardise presentation across articles. The cleaning procedures are outlined in the online supplemental file 1. Briefly, the data were cleaned to remove special characters, and authors with blank first or last names were excluded. Each author's name was defined as the concatenated string of their first and last name separated by a space.

### Primary analyses

To assess the distribution of authorship across individuals, the GC was calculated using the DescTools package in R (V.0.99.30). Please refer to the online supplemental file 1 for an example of the dataset and the calculation of the GC. GCs were calculated for the entire author list as well as for authors listed in first author position or in the last author position to assess equality across these important positions.

Next, the change in GC over time for each authorship position (overall authorship, first authorship and last authorship) was assessed using Spearman rank-order correlation. Finally, to compare the equality of authorship between specialties, the GC for each of the nine medical specialties and the high-impact cross-specialty journal was calculated, and the change in GC over time

for each specialty was assessed using Spearman rank-order correlations.

## Sensitivity analyses

### ORCID IDs

Since the methodology for matching authors by name is potentially limited by authors with the same name—a challenge in name disambiguation studies known as name homography—we attempted to quantify the rate of authors with the same name by comparing names across ORCID IDs. To do so, we downloaded 2019 ORCID Public Data File which is freely available data containing public information for all records in the ORCID database at the time the snapshot was taken (October 2019).<sup>13</sup> We extracted the given (first) and last (family) name for each record in this database and cleaned the database using the same methods used to clean the authorship database used in the main analysis. Then, we calculated the number of authors with matching names, assuming that each ORCID ID referred to a unique author.

Next, in an attempt to account for the extent to which initials identified distinct authors, we also cleaned the ORCID database to remove initials, and recalculated the number of ORCID records with matching names. The method used to identify and remove initials is outlined in the online supplemental file 1.

### Controlling for author list length, consortia and number of articles published per year

We conducted additional sensitivity analysis to exclude the possibility that changes in GC were driven by three trends in medicine publishing that directly impact the author list: publishing articles with longer author lists, publishing via consortia and the increase in number of articles published. Therefore, we conducted a sensitivity analysis to determine if the extreme right tail of author list lengths influenced the measurement of GC by repeating the primary analyses solely among articles with 20 or fewer authors. This threshold was chosen a priori to exclude articles with the largest author lists, while also keeping at least 95% of articles for analysis. The second sensitivity analysis assessed the change in GC over time in articles without a consortium listed as the last author. Finally, the third sensitivity analysis repeated the primary analysis while randomly sampling 20 000 articles from each year without replacement and calculated the GC for each author position in each year. This sampling was repeated 10 times. The limit of 20 000 was determined a priori by assessing the smallest number of articles published in any 1 year (n=24 144) and selecting a limit below that minimum. The change in GC over time for each model was assessed using Spearman rank-order correlations.

### Controlling for variance across specialties

Given the substantial increase in GC within the high-impact cross-specialty journals over the study period, a

fourth and fifth sensitivity analyses were completed. The first assessed the change in GC for the entire collection of articles excluding articles published in the four cross-specialty journals. This analysis aims to determine if the finding that the overall increase in GC over time looking across specialties was driven by the articles published in the cross-specialty journals, or if articles published in non-cross specialty journals also demonstrated an increase in inequality.

Additionally, the cross-specialty journal category included a smaller number of journals and therefore articles than the other specialties. To control for the potential impact of different number of articles (and therefore author positions), we repeated the analyses comparing GC between specialties after sampling for 1000 articles per specialty per year. This limit was determined by assessing the smallest number of articles published in one specialty in any 1 year (n=1362 articles) and selecting a limit below that minimum.

### Gender and authorship distribution

Demographic factors including gender are likely to impact the distribution of authorship. To assess the impact of gender on the authorship distribution, we used a well-validated software (Genderize.io (<https://genderize.io>)) to predict the gender of each author using their first name using 60% as the confidence threshold for assigning a gender to an author. Additional information on this method has been published previously.<sup>10 14</sup> Then we assessed the distribution of authorship within each gender both in the overall cohort and over the study period.

All analyses were conducted in R (3.5.1).<sup>12</sup>

### Patient and public involvement

There was no patient or public involvement in this study.

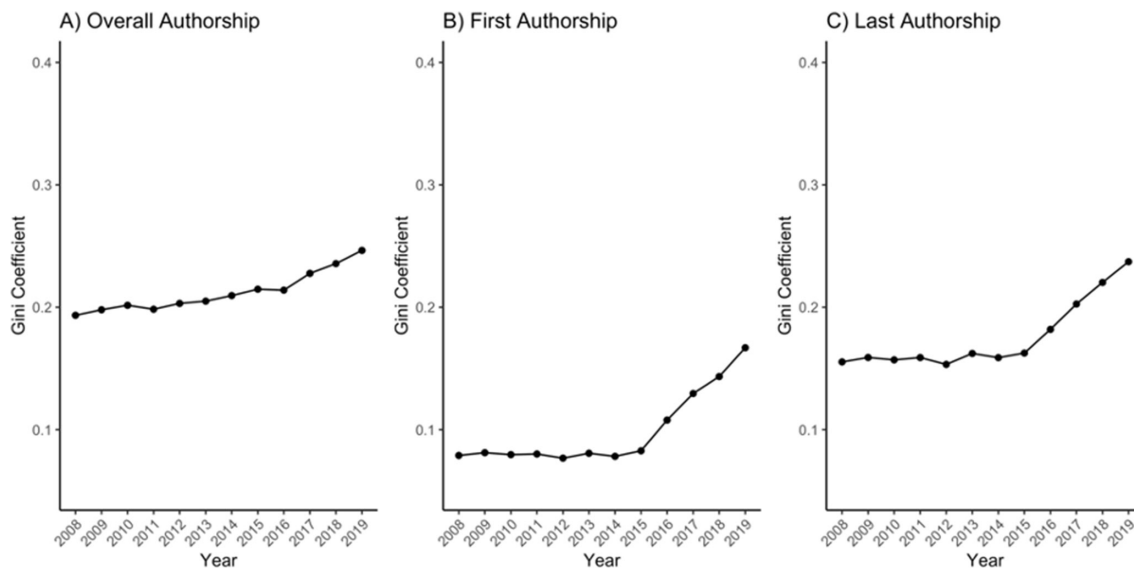
## RESULTS

Our data contained 2 254 432 author positions and 936 574 unique authors. Only 0.017% of identified authors (156/936 574) had published more than 100 articles during the 12-year study period. The maximum number of articles attributed to one author during the entire study period was 337 articles (online supplemental figure S1). These data demonstrate an increase in the number of articles published each year, and an increase in the number of author positions available each year (online supplemental figure S2).

### Primary analyses

#### Whole cohort

To reorient readers, the GC ranges from 0 to 1 with 0 indicating complete equality across individuals sampled and 1 indicating complete inequality or 1 individual holding all of the resources within the sampled cohort. A smaller GC indicates a more equal distribution of resources across



**Figure 1** Change in Gini coefficient by authorship position (A. overall authorship, B. first authorship, and C. last authorship) over time.

individuals, and a larger GC indicates a less equal distribution of resources among individuals.

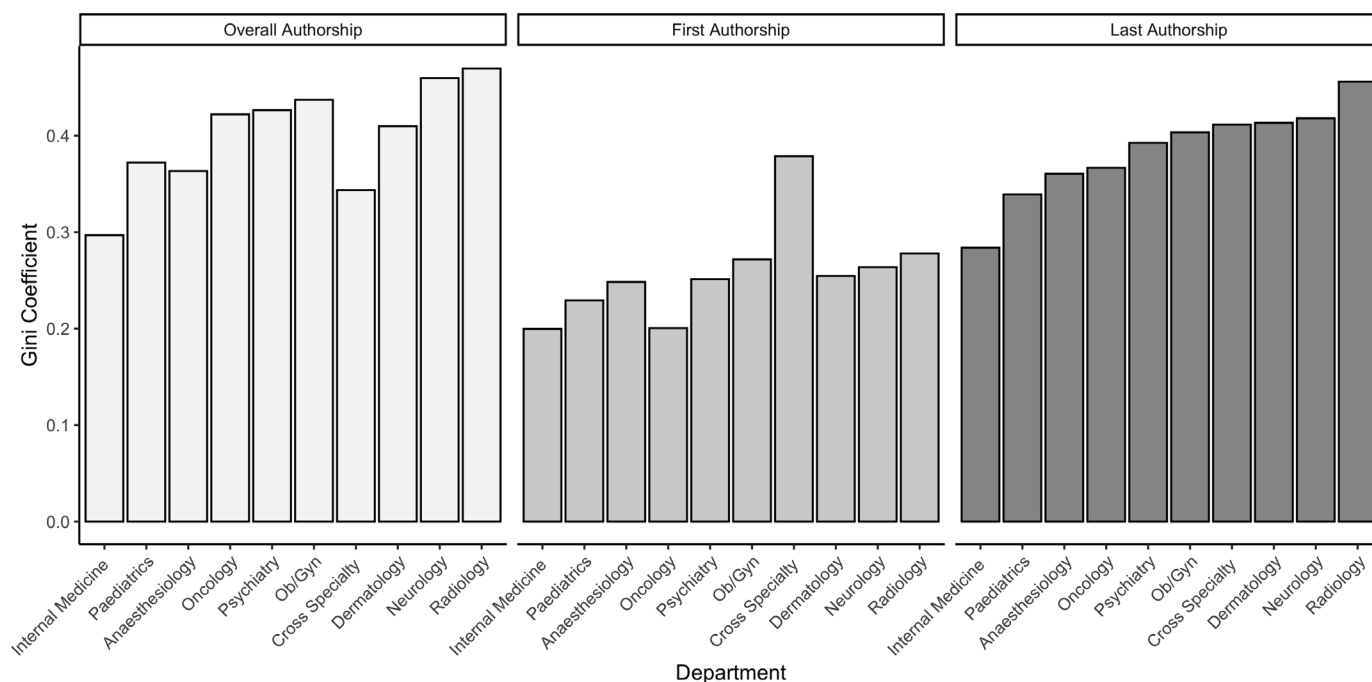
The overall GC for all articles sampled across the 12-year study period was 0.49, and the GCs for the first and last authorship positions were 0.30 and 0.44, respectively. This indicates that there was more inequality of authorship among overall authorship and last authors than among first authors.

Looking across time, [figure 1](#) shows the change in GC from 2008 to 2019 for each authorship position. There was a significant positive correlation between year and GC for the overall authorship position ( $r=0.99$ ,  $p<0.001$ ), the first author position ( $r=0.75$ ,  $p=0.007$ ) and the last author

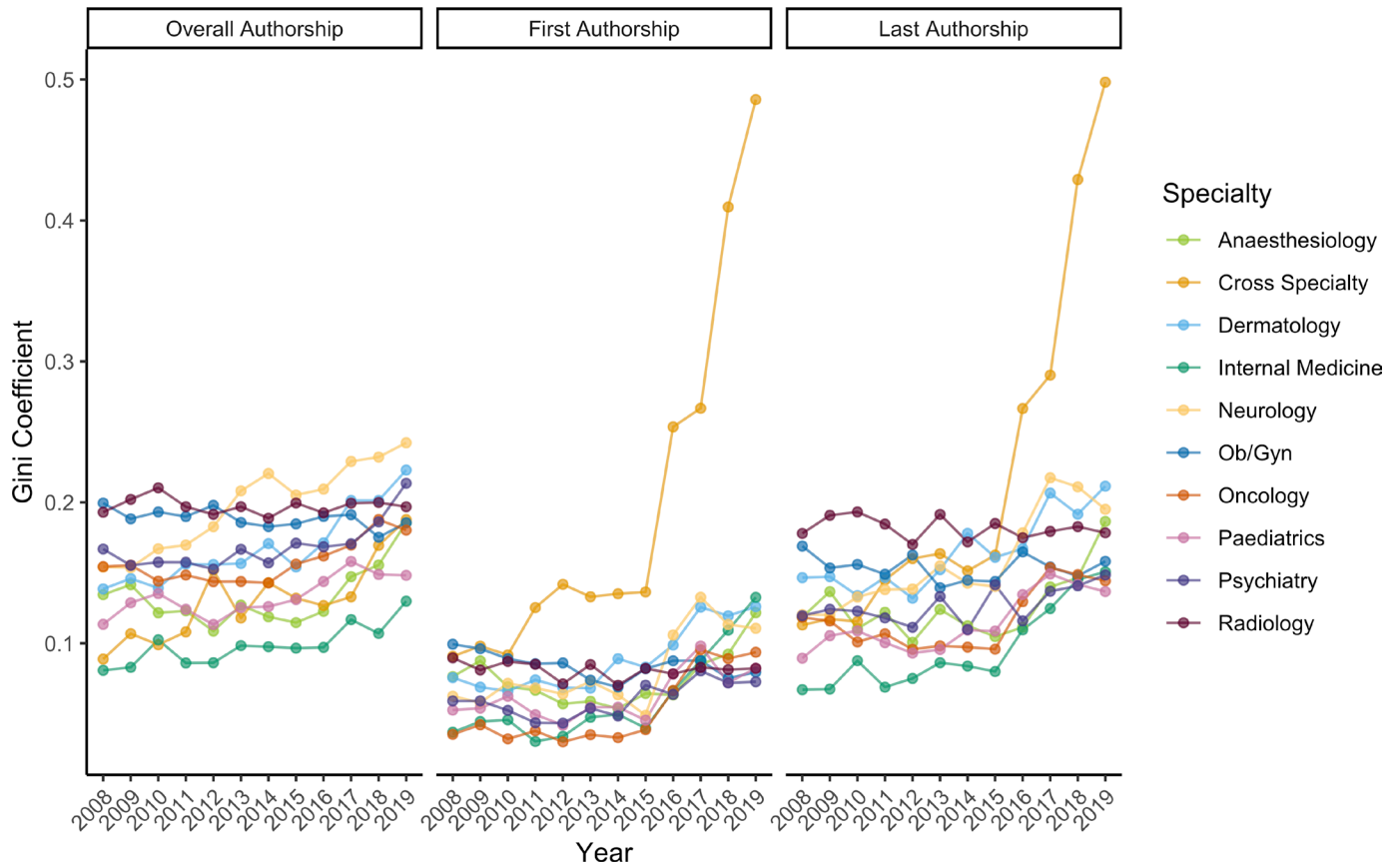
position ( $r=0.85$ ,  $p<0.001$ ). This positive increase in GC over time indicates that the distribution of authorship was becoming more unequal over the study period for each of the author positions analysed.

#### Specialty analysis

Online supplemental table S2 and [figure 2](#) present the GC by authorship position by specialty for the entire collection of articles during the study period. The lowest GC for overall authorship (ie, the most widely distributed among individuals) was observed in internal medicine (0.30), cross-specialty journals (0.34) and anaesthesiology (0.36); the highest GC (ie,



**Figure 2** Gini coefficient by authorship position separated by specialty.



**Figure 3** Change in Gini coefficient by authorship position over time separated by specialty.

highest inequality of authorship) was observed among radiology (0.47), neurology (0.46), and obstetrics and gynaecology (0.44). Among first authors, oncology (0.20), internal medicine (0.20) and paediatrics (0.23) had the lowest GCs. Cross-specialty journals (0.38), radiology (0.28), and obstetrics and gynaecology (0.27) had the highest GCs. Among last authors, the specialties with the lowest GC for overall authorship were internal medicine (0.28), paediatrics (0.34) and anaesthesiology (0.36). The specialties with the highest GCs were radiology (0.46) and neurology (0.42).

Figure 3 presents the change in GC by specialty over time and table 1 presents the Spearman rank-order correlations for each specialty over time. Of note, the cross-specialty high-impact journals have the greatest rate of increase in GC for the first and last author position, indicating increasingly uneven distribution in authorship over time.

**Sensitivity analyses**

The 2019 ORCID Public Data File contained 7 314 172 entries. After cleaning to remove special characters, 7 015 154 entries met criteria for inclusion. Of those records, 1 763

**Table 1** Spearman rank-order correlations by author position by specialty

| Specialty                  | Overall authorship |         | First authorship |         | Last authorship |         |
|----------------------------|--------------------|---------|------------------|---------|-----------------|---------|
|                            | R                  | P value | R                | P value | R               | P value |
| Anaesthesiology            | 0.34               | 0.28    | 0.22             | 0.48    | 0.41            | 0.18    |
| Cross-specialty            | 0.84               | 0.001   | 0.95             | <0.001  | 0.96            | <0.001  |
| Dermatology                | 0.92               | <0.001  | 0.80             | 0.003   | 0.87            | <0.001  |
| Internal medicine          | 0.78               | 0.004   | 0.80             | 0.003   | 0.87            | <0.001  |
| Neurology                  | 0.96               | <0.001  | 0.67             | 0.02    | 0.94            | <0.001  |
| Obstetrics and gynaecology | -0.57              | 0.06    | -0.57            | 0.06    | -0.17           | 0.59    |
| Oncology                   | 0.62               | 0.03    | 0.65             | 0.03    | 0.43            | 0.16    |
| Paediatrics                | 0.73               | 0.01    | 0.63             | 0.03    | 0.80            | 0.003   |
| Psychiatry                 | 0.73               | 0.01    | 0.66             | 0.02    | 0.49            | 0.11    |
| Radiology                  | -0.09              | 0.78    | -0.34            | 0.28    | -0.24           | 0.46    |

106 (25.13%) contained a name that was identical to at least one other author. After processing to remove initials, 1 759 008 records out of 6 904 982 (25.47%) records contained a name that was an identical match to at least one other record.

The sensitivity analyses assessing the impact of author list length, consortia and number of articles published per year did not differ meaningfully from the primary analysis. They are outlined in detail within the online supplemental figures S3–S5 and supplemental tables S3 and S4.

The sensitivity analysis demonstrated that the increase in GC within the high-impact cross-specialty journals was not driven by the smaller number of articles published in the cross-specialty journals (online supplemental figures S6–S8). Additionally, after excluding articles published in cross-specialty journals, the first and last author positions still demonstrated a significant positive correlation between year and GC for overall authorship ( $r=0.92$ ,  $p<0.001$ ) and the last author position ( $r=0.69$ ,  $p=0.02$ ). However, the first author position did not demonstrate a positive correlation between year and GC ( $r=0.48$ ,  $p=0.12$ ).

### Gender analyses

Looking across the entire cohort the distribution of authorship among authors identified as men was more unequal (higher GCs) than among authors identified as women (online supplemental table S5). However, looking across time, there was a greater increase in the GC among first and last authors identified as women than first and last authors identified as men (figure 4; online supplemental table S6).

## DISCUSSION

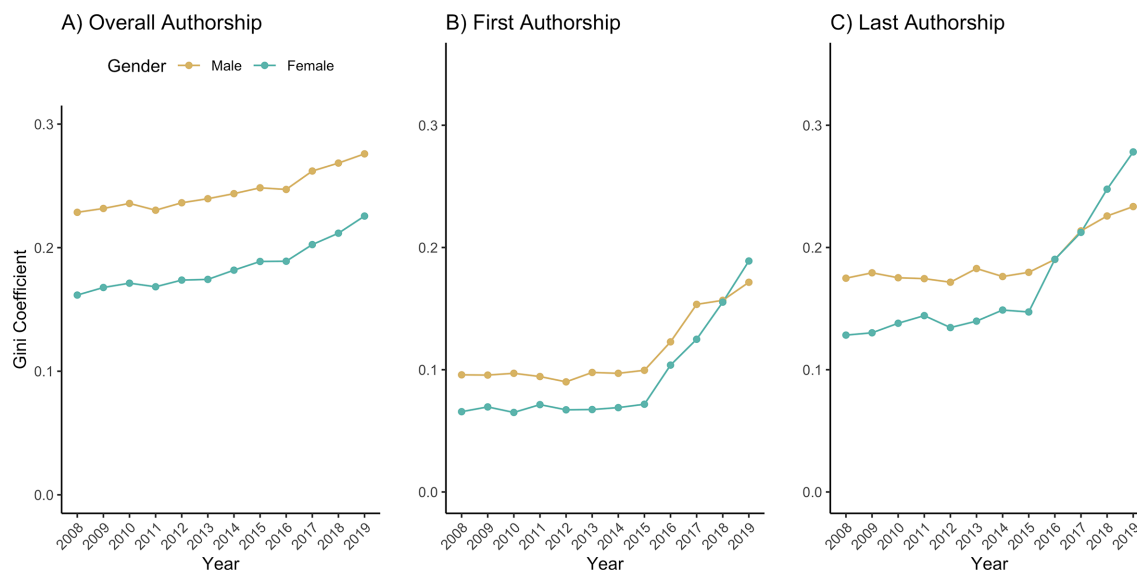
In this bibliometric study assessing the distribution of authorship across individuals publishing 312 222 original research articles in 134 medium-impact to high-impact academic medical journals, we find moderate and inequality in authorship distribution. The first author position exhibited a more even distribution of authorship,

indicated by a lower GC of 0.30 compared with 0.49 for all authors and 0.44 for senior authors. Looking over time, we see an increase in the GC for the entire author block as well as for the first and last author positions indicating an increase in the inequality of authorship distribution for all three authorship metrics. Sensitivity analyses demonstrated that the shift in the distribution of authorship is not being driven by especially large author lists, consortium collaborations or the number of articles published each year.

The skew in author distribution is particularly apparent when examining the distribution in authorship within high-impact cross-specialty journals over time. These journals showed the greatest increase in GC, as compared with the other specialty groups for the first and last author position. However, it is important to note that we also observed an increase in author inequality for overall authorship and last authorship among articles not published in the high-impact cross-specialty journals. This indicates that authorship inequality is increasing across academic medicine not just in the high-impact cross-specialty journals.

Although these effects are robust, their meaning will require further analysis. In economics, disparity in wealth has been associated with a range of negative outcomes.<sup>15</sup> Disparity in publication, however, may not be intrinsically bad—some researchers will inevitably be more productive than others, so it is not reasonable to expect a uniform distribution of publications. Indeed, prior work demonstrated that hyperproductive authors vary in the types of articles produced with some producing primarily letters and review articles and others producing primarily reports of clinical trial data.<sup>16</sup>

On the other hand, where authorship distribution is particularly uneven, it may reflect an underlying lack of equity in access to resources—that is, a small number of investigators control data and funding, to



**Figure 4** Change in Gini coefficient by authorship position (A. overall authorship, B. first authorship, and C. last authorship) over time separated by gender.

the exclusion of others. This concern for authorship concentration among a small number of authors is not new, with differential credit accruing to more famous authors referred to as ‘the Matthew effect’ in a 1968 *Science* paper.<sup>17</sup> This concern parallels those relating to federal grant funding, which precipitated efforts to reduce the concentration of competitive grant support among senior researchers.<sup>18</sup> Such concentration may also play a role in perpetuating other forms of inequity in academic medicine.<sup>10 14 19</sup> Another potential consequence is academic monoculture—that is, inhibition of new or less widely accepted hypotheses that are not supported by a small number of powerful investigators.<sup>20</sup> By slowing the emergence of initially unpopular ideas, like the association of *Helicobacter pylori* infection with gastric ulcers,<sup>21</sup> the concentration of authorship may inadvertently delay the development of new treatments.

Additionally, while there has been an increase in the representation of women within academic medicine, our results indicate that publications are becoming more unequally distributed across women, especially for the first and last author position.<sup>10</sup> Therefore, even though there has been an increase in the number of publications with a woman as the last author, this finding may be somewhat driven by the increased productivity of a small group of individuals (indicated by substantial increased GC in the last author position over time). Therefore, we must be careful to not mistake the success of a few as individuals as an indicator of equity of access and resources for the larger cohort. Furthermore, additional research is needed to examine how other demographic factors including age, race, ethnicity and nationality influence the distribution of authorship.

The present study cannot address these questions but may help to motivate their investigation. By demonstrating that the distribution of authorship is becoming increasingly unequal, especially among high-impact journals, our work provides further evidence for the need to conceptualise academic success as a multifaceted concept that does not solely rely on publication statistics. Over the last 10 years, there has been increased awareness of the pitfalls of using publication-based indicators (including number of publications, h/g-index or impact factor of the journals in which an author publishes) as metrics of academic success. While these indicators represent succinct, easily calculated measurements, their quantitative nature can fail to capture important, more qualitative, aspects of an individual’s performance or prowess.<sup>22 23</sup> Additionally, these metrics may be skewed by the well-documented trend of authorship inflation. For example, a recent report indicated that manuscripts with hyperauthorship (more than 100 authors or representation of authors from more than 30 countries) have aberrant citation patterns leading to elevated citation rates.<sup>24</sup>

In addition to efforts by the ICMJE and individual journals to ensure all authors meet the threshold for inclusion, multiple publications and initiatives have

focused on improving research assessment and evaluation, including the San Francisco Declaration on Research Assessment (2012) and the Leiden Manifesto (2015) which was recently adopted by Elsevier.<sup>22 25 26</sup> These works emphasise the importance of contextualising these quantitative metrics alongside other indicators of success.<sup>22 23 25</sup>

We note a few limitations in interpreting these results. The automated sampling of articles and author names allows for the analysis of an extensive number of articles and journals. While this is a substantial advantage of this method, this automated sampling may lead to the combination of authors with the same first and last name, a problem known as name homography.<sup>27</sup> Without manually checking all authors included, it would be impossible to identify these combinations. However, in our sensitivity analysis of names compiled from ORCID IDs, we noted that 25.13% of the names matched the name for another ORCID ID. This rate of name homography aligns with prior work on author disambiguation which showed that between 4% and 30% of authors had a name matching another author, depending on the database.<sup>27</sup> ORCID IDs may show a higher rate of name homography because people with a name similar to other authors may be more likely to create an ORCID ID to distinguish their own work.

Another potential limitation is that automatic sampling makes it challenging to check that an author’s name is consistently formatted across articles. Indeed, our prior work has noted inaccurate indexing of author names in PubMed, (ie, only including author initials instead of author full name).<sup>10</sup> Therefore, we may be failing to combine all articles for a given author. This is a problem known as name variability. Prior work on author name disambiguation in non-medical databases identified rates of name variability between 1% and 23% depending on the level of detail included in the database (initials vs full names).<sup>27</sup> Our analysis removing initials from ORCID IDs demonstrated that removal of initials only accounted for a small increase in name homography potentially indicating an increased importance of name homography over name variability in this space. The inclusion of ORCID identifiers should help to ameliorate this problem, once they are universally adopted.

Another potential limitation is our choice to analyse first and last author as the key positions within the author list. Recent work has indicated that listing authors in descending order by contribution is almost unanimously accepted. However, survey respondents were split on the importance of the second author versus the last author.<sup>28</sup> While further guidelines on this distinction are still needed, our choice to analyse the last author position is based off of the distinction that last authorship is more widely regarded as the prominent position within medical research.

While not necessarily a limitation, it is important to note that the GCs calculated for each specialty are precise measurements quantifying the Lorenz curve. Therefore,



CIs cannot be calculated from the GC and all comparisons of GC between specialties are qualitative in nature. However, this could be viewed as a strength of the methodology as programmatic scraping makes it feasible to analyse a large number of articles and journals. Additionally, we chose not to correct for multiple comparisons in our correlations between GC and year nor in our sensitivity analyses due to the large sample of the data and the novelty of this approach to examining authorship distributions.

Another limitation of this study is that we are unable to examine factors that may influence the decision to submit manuscripts to a high-impact journal or factors that impact the acceptance of a manuscript for publication within a journal. Prior work has demonstrated that double-blind peer review increases both the representation of women and marginalised groups as authors, and may decrease any unfair advantage given to well-known authors or individuals working at prestigious institutions.<sup>29–31</sup> Future work should continue to examine factors that may influence submission and acceptance rates by under-represented cohorts.

Finally, these data do not assess the relationship between authorship distributions and funding. While funding information is available on PubMed and other sources, it is not easily scrapped or processed. Therefore, it is outside of the scope of the current project. Future work is needed to examine the relationship between funding and authorship inequalities, especially in high-impact journals.

Overall, these data suggest a growing inequality in authorship across authors publishing in high-impact academic medical journals, especially among the highest impact journals. These findings may have implications for processes such as promotions and allocation of research funding that use authorship metrics as key criteria for making decisions. More broadly, they point to the importance of understanding how such shifts in distribution may impact the medical literature, and whether interventions to moderate these shifts may be needed.

**Contributors** RHP was responsible for the conceptualisation of the project, supervision of data collection and manuscript preparation. KLH collected and analysed the data, and drafted the manuscript. Both authors read and approved the final manuscript.

**Funding** This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** RHP reports personal fees from Genomind, Outermost Therapeutics, RID Ventures and Takeda. He reports equity in Psy Therapeutics and Outermost Therapeutics. All are outside the submitted work. KLH has nothing to disclose.

**Patient consent for publication** Not required.

**Ethics approval** This study was not submitted for ethics approval as it does not analyse data collected from human or animal subjects. The analysis plan was not preregistered as this was a secondary analysis of existing data.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available at <https://github.com/kamberhart/GC-academic-medicine>.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iDs

Kamber L Hart <http://orcid.org/0000-0002-0168-645X>

Roy H Perlis <http://orcid.org/0000-0002-5862-6757>

#### REFERENCES

- 1 Tilak G, Prasad V, Jena AB. Authorship inflation in medical publications. *Inquiry* 2015;52. doi:10.1177/0046958015598311. [Epub ahead of print: 29 07 2015].
- 2 Eisenberg RL, Ngo L, Boiselle PM, et al. Honorary authorship in radiologic research articles: assessment of frequency and associated factors. *Radiology* 2011;259:479–86.
- 3 Kwok LS. The white bull effect: abusive coauthorship and publication parasitism. *J Med Ethics* 2005;31:554–6.
- 4 Huth EJ, Case K. *The URM: twenty-five years old*. Science Editor, 2004: 27. 17–21.
- 5 International Committee of Medical Journal Editors. Defining the role of authors and contributors. Available: <http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>
- 6 Gini C. Measurement of inequality of incomes. *Econ J* 1921;31:124–6.
- 7 Chappelou J. Gini Coefficient. *Investopedia*, 2020. Available: <https://www.investopedia.com/terms/g/gini-index.asp>
- 8 Asada Y. Assessment of the health of Americans: the average health-related quality of life and its inequality across individuals and groups. *Popul Health Metr* 2005;3:7.
- 9 Jin J, Wang J, Ma X. Equality of medical health resource allocation in China based on the Gini coefficient method. *Iran J Public Health* 2015;44:445–57.
- 10 Hart KL, Perlis RH. Trends in proportion of women as authors of medical Journal articles, 2008–2018. *JAMA Intern Med* 2019;179:1285.
- 11 Clarivate Analytics. Data Citation Index. Available: <https://clarivate.com/products/web-of-science/web-science-form/data-citation-index/>
- 12 R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, 2019.
- 13 Blackburn Ret al. *ORCID public data file 2019*, 2019.
- 14 Hart KL, Frangou S, Perlis RH. Gender trends in authorship in psychiatry journals from 2008 to 2018. *Biol Psychiatry* 2019;86:639–46.
- 15 Piketty T, Goldhammer A. *Capital in the twenty-first century*. Belknap Press, 2017.
- 16 Wager E, Singhvi S, Kleinert S. Too much of a good thing? An observational study of prolific authors. *PeerJ* 2015;3:e1154.
- 17 Merton RK. The Matthew effect in science. *Science* 1968;159.
- 18 Wahls WP. The NIH must reduce disparities in funding to maximize its return on investments from taxpayers. *Elife* 2018;7:e34965.
- 19 Kaplan SE, Raj A, Carr PL, et al. Race/Ethnicity and success in academic medicine: findings from a longitudinal multi-institutional study. *Acad Med* 2018;93:616–22.
- 20 Kuhn TS, Hacking I. *The structure of scientific revolutions*. The University of Chicago Press, 2012.
- 21 Weintraub P. The doctor who drank infectious broth, gave himself an ulcer, and solved a medical mystery. *Discover* 2010.
- 22 Hicks D, Wouters P, Waltman L, et al. Bibliometrics: the Leiden manifesto for research metrics. *Nature* 2015;520:429–31.
- 23 Wilsdon Jet al. *The metric tide: report of the independent review of the role of metrics in research assessment and management 2015*.



- 24 Adams J, Potter R, Szomszor M. *Multi-authorship and research analytics*, 2019.
- 25 San Francisco Declaration on Research Assessment, 2012. Available: <https://sfdora.org/read/>
- 26 Elsevier endorses Leiden manifesto to guide its development of improved research evaluation, 2020. Available: <https://www.elsevier.com/about/press-releases/corporate/elsevier-endorses-leiden-manifesto-to-guide-its-development-of-improved-research-evaluation>
- 27 Müller M-C, Reitz F, Roy N. Data sets for author name disambiguation: an empirical analysis and a new resource. *Scientometrics* 2017;111:1467–500.
- 28 Mavis B, Durning SJ, Uijtdehaage S. Authorship order in medical education publications: in search of practical guidance for the community. *Teach Learn Med* 2019;31:288–97.
- 29 Tomkins A, Zhang M, Heavlin WD. Reviewer bias in single- versus double-blind peer review. *Proc Natl Acad Sci USA* 2017;114:12708–13.
- 30 Hirshfield LE. A case for double-blind review. *Acad Med* 2020;95:1627–8.
- 31 Budden AE, Tregenza T, Aarssen LW, *et al*. Double-Blind review favours increased representation of female authors. *Trends Ecol Evol* 2008;23:4–6.