Representation In The Endocrinology Literature And The Effect Of Double-Blinded Review

Gal Hodish
gal.hodish@yale.edu

Follow this and additional works at: https://elischolar.library.yale.edu/ysphtdl

Recommended Citation
Hodish, Gal, "Representation In The Endocrinology Literature And The Effect Of Double-Blinded Review" (2024). Public Health Theses. 2402.
https://elischolar.library.yale.edu/ysphtdl/2402

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Representation in The Endocrinology Literature and the Effect of Double-Blinded Review

Hodish G¹, Wininger M², Moretti B³, Chen D³, Esfandiari N³, Hoh J¹, Haymart M³, Papaleontiou M³

Authors’ affiliations:

¹Yale School of Public Health, Department Chronic Disease Epidemiology, New Haven, CT, US.

²Yale School of Public Health, Department of Biostatistics, New Haven, CT, US.

³University of Michigan, Division of Metabolism, Endocrinology and Diabetes, Ann Arbor MI, US.

Corresponding author: Gal Hodish, Yale School of Public Health, Department Chronic Disease Epidemiology, 60 College st, New Haven, CT, US.

Telephone number: 7347301089 email: gal.hodish@yale.edu

Reprint requests should be addressed to Gal Hodish

Running title: Representation in The Endocrinology Literature

Key words: Publication bias; Authorship bias; Representation; Gender bias; Diversity; Race Bias

Contributorship: GH – design, data collection, data interpretation and writing the manuscript; MW – data collection, data interpretation and writing the manuscript

Declaration of Competing Interests: GH declares no competing interest.
Abstract

Background:

There is an open question as to whether authorship in peer-reviewed scientific journal articles is representational. Part of this discrimination has been attributed to bias in the review process of manuscripts, with the suggestion that a double-blind review process ensures equitable outcomes due to complete deidentification of both authors and reviewers.

Methods:

We performed an automated review of 4,191 manuscripts published at three endocrinology journals with comparable impact factor (two single-blind journals; one double-blind journal) in the years 2010-2023; manuscripts were included in the analysis if they were published from institutions based in the United States. Gender and race/ethnicity associated with authors’ names were predicted using publicly available algorithms (GenderAPI and REthnicity, respectively). Univariate and multivariate analyses were conducted to determine whether relationships are present between peer-reviewed journal blindness and author characteristics.

Findings:

There was a statistically significant difference in proportion of female first-authors by journal (P=0.00055), but no significant difference among last-authors (P=0.1838). We observed the opposite trend in race-ethnicity: statistically significant difference between journals among last-authors (P=0.04518), but not among first-authors (P=0.1465). Female representation among these authors was less than expected (absolute difference: 11.5%, P<0.05), but this trend appease to be ameliorating over time (increase of 1.5% female representation per year, P<0.05). Secondary and exploratory analyses revealed complex trends in gender and race-ethnicity with respect to review double-blind.
Interpretation:

Historical trends in under-representation of female investigators among published authors in endocrinology appears to be substantial, but trending toward equality. Nuanced findings in both gender and race-ethnicity provide compelling opportunities for further investigation.

Introduction

Proportional representation of gender and race/ethnicity among clinical investigators is an increasingly compelling topic in various medical sub-specialties and in the medical community at-large. Indeed, the notion of representation among authors has provided an unprecedented view into the potentially differential impacts of disruptive events on career scientists, viz. the COVID-19 pandemic. And inquiry into this proportionality is important: representation and visibility are integral to the notion of belonging; it's hard to be what you can't see.

Within the field of Endocrinology, there is unique opportunity to draw inferences about the proportionality of representation, particularly as pertains to author gender. There is lately consistent evidence of a demographic shift toward a female majority among clinical practitioners which appears to be driven by an especially large representation of young clinicians entering fellowship. Furthermore, contemporaneous to this trend of increasingly female representation in the clinical ranks, one of the field’s flagship journals revised their operational policy to incorporate a double-blind review. Whereas elsewhere the decision to embrace a double-bind review structure has been explained as a measure to self-protect against possible bias in the review process, re-configuring a journal as double-blind could reasonably be expected to yield a measurable change in that journal’s published works.
In this study, we seek to estimate the distribution of manuscript authorship by gender and by racial-ethnic group in a representative sample of endocrinology journals. We assess these trends before and after the time in which one of these journals converted to a double-blind review format.
Material and Methods

Sample of articles

Article accessioning was performed via PubMed for three journals, which for the purpose of anonymity, we will refer to as Journal A, Journal B, and Journal C, within years 2010 – 2023. Given that PubMed archives by publication date, and reposit articles ahead of publication, some manuscripts in our analysis have publication dates of 2024. Article retrieval was accomplished via the easyPubMed package in R v4.2.1. An overview of the journals is provided in Table 1.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Blind</th>
<th>Publisher IF</th>
<th>SCIMagoJR H-Index</th>
<th>Single-Blind</th>
<th>Double-Blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Single</td>
<td>3.4</td>
<td>165</td>
<td>290</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>Single</td>
<td>5.8</td>
<td>6</td>
<td>3049</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>Single (&lt;2016); Double (≥2017)</td>
<td>4.2</td>
<td>103</td>
<td>502</td>
<td>353</td>
</tr>
</tbody>
</table>

Table 1: Summary of journals included in this study. Impact Factor retrieved in April, 2024. Number of articles reflects number of articles surviving data-conditioning steps described in Sample of Articles. Pre- and Post-Blind reflect the transition to double-blind review at Journal C. IF = Impact Factors provided by publisher retrieved from journal homepage for latest year available; SCIMagoJR data reflects year 2023.

Dataset

A full description of our dataset preparation is available as an appendix/supplement. In brief summary: we performed a batch retrieval of all manuscripts published in three journals (Table 1), yielding a preliminary catchment of 17,350 manuscripts with 131,504 total author listings. Manuscripts were removed from the analysis stream if they appeared to be associated with an address outside of the United States. An authorship index was created tallying appearances within the dataset (1 for first appearance in the dataset, 2 for second appearance, and so on) yielding a numerical proxy measure of author seniority. Race-Ethnicity was
estimated using the Rethnicity package in R, Gender was estimated using the web-based GenderAPI, and manuscript article type was obtained from NCBI using their structured categorical field information. Authors with race-ethnicity values with less than 70% assignment probability were categorized as Unknown. Following all data conditioning steps, the final dataset contained 4,194 manuscripts.

Analysis

Our primary objective was to characterize the trend of gender representation amongst first- and last-authors over time (so-called “bookend authors”). We tested this with a multi-variable regression with a response variable of proportion of manuscripts with female investigators with bookend authorship, and an interaction between two predictor terms: Journal and Year. For this aim, we took as precept that the distribution of potential manuscript authors was approximately 50-50 by gender. And as a corollary, we conjectured that if the pool of potential eligible authors was 50-50, then among any two bookend authors on a manuscript, there was a \( P_0 = \binom{2}{0} \cdot \frac{1}{2}^2 = \frac{2!}{2! \cdot 0!} \cdot \left(\frac{1}{2}\right)^2 - 0^2 = 0.25 \) probability of encountering a manuscript with no female representation (binomial probability), and thus a 75% probability of encountering a manuscript with some female representation among its booked authors. Thus, we took on this analysis with two null hypotheses: 1) the proportion of female representation in these manuscripts is not significantly different from 75%, and 2) the slope of female representation in these manuscripts is not different from zero (i.e. there has been no significant change over time).

As a secondary objective we sought to ascertain whether journal-blinding associated with a significant change in the representation of females among bookend authors. To test this, we devised a logistic regression with female authorship as a dichotomous response variable
(Yes = one or both bookend authors female; No = no bookend authors female), and double-blind review as a dichotomous predictor. In order to account for possible bias among non-blinded journals due to author seniority or subject-matter expertise, we included author index as a continuous co-variate here. Our Null Hypothesis is that change in blind did not change the representation of females among authors.

As an additional objective, we sought to characterize the racial-ethnic distribution of bookend authors over-time, and as a variable with potential sensitivity to single- versus double-blinded review. Because we were unable to find credible demographical information related to the racial-ethnic background of clinical practitioners or career scientists within the field, we were unable to rigorously test hypotheses against a benchmark. As a result, this analysis is somewhat exploratory and descriptive.
Results

Descriptive Statistics
Of the 4,191 papers analyzed, 3,869 (92.3%) yielded two authors with identifiable gender. Among these manuscripts, 2,542 (65.7%) featured one or two females among the bookend authors. A further 1,515 manuscripts yielded two authors with estimable race-ethnicity. Among these manuscripts, 75.6% featured one or two non-White authors among the booked authors. Detailed summaries are provided in Table 2.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>290</td>
<td>3049</td>
<td>502</td>
<td>353</td>
</tr>
<tr>
<td>First/Last</td>
<td>First/Last</td>
<td>First/Last</td>
<td>First/Last</td>
<td>First/Last</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>167/113</td>
<td>1519/1100</td>
<td>247/158</td>
<td>163/141</td>
</tr>
<tr>
<td>Male</td>
<td>99/155</td>
<td>1370/1781</td>
<td>246/330</td>
<td>178/201</td>
</tr>
<tr>
<td>Unknown</td>
<td>23/22</td>
<td>159/166</td>
<td>9/14</td>
<td>12/11</td>
</tr>
<tr>
<td>Race-Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>99/54</td>
<td>888/564</td>
<td>158/99</td>
<td>139/98</td>
</tr>
<tr>
<td>Black</td>
<td>5/10</td>
<td>90/90</td>
<td>14/22</td>
<td>10/12</td>
</tr>
<tr>
<td>Hispanic</td>
<td>21/24</td>
<td>157/156</td>
<td>25/16</td>
<td>14/17</td>
</tr>
<tr>
<td>White</td>
<td>58/83</td>
<td>647/864</td>
<td>114/186</td>
<td>64/74</td>
</tr>
<tr>
<td>Unknown</td>
<td>107/119</td>
<td>1267/1375</td>
<td>1911/79</td>
<td>126/152</td>
</tr>
</tbody>
</table>

Table 2: Summary of gender and race-ethnicity among bookend authors.

First authorship was approximately evenly distributed between males and females (52.5% female), but females were substantially under-represented among last-authors (38%). There was a preponderance of non-white first-authors (64.7%), and an approximately equal distribution of white- versus non-white last-authors (49.1%).
In univariate analysis of gender-by-journal, there was a statistically significant difference in proportion of female first-authors by journal (P=0.00055), but no significant difference among last-authors (P=0.1838). We observed the opposite trend in race-ethnicity: statistically significant difference between journals among last-authors (P=0.04518), but not among first-authors (P=0.1465).

Primary objective: Female representation over time

In our linear regression model, we found that starting in year 2010, female representation in bookend authorship roles was approximately 64%, which was significantly different from our baseline assumption of 75% (i.e. the random-chance benchmark; P=0.0359; Figure 1). We found a statistically significant increase in female representation over time (a slope of 1.5% percent increase per-year over the span 2010-2023; P=0.04). We did not find a statistically significant difference in representation between journals, nor did we see a statistically significant interaction between journals and time.
Figure 1: Summary of regression of female representation over time among bookend authors.

These trends can be visualized as seen in Figure 2:
Figure 2: Trend in authorship by-journal over time.

**Secondary objective: Female representation as a function of blind**

In our logistic regression model, we found that double-blinding did not change the proportion manuscripts with female representation: OR = 0.81 (0.65 – 1.03), P=0.0858.
Tertiary objective: Non-White author representation as a function of blinding.

In our exploration of race-ethnicity, we found that among manuscripts with both authors’ race-ethnicity being identifiable, 1028 out of 1376 (75%) of single-blind manuscripts featured non-white authors; in double-blind review, 118 out of 142 (83%) manuscripts contained non-white authors. This difference was statistically significant in both conventional...
chi-square analysis (P=0.03) and in logistic regression accounting for author index (P=0.03; Figure 4).

Figure 4: Summary of regression of Non-White representation as a function of double-blind status.
Discussion

Assumptions and Context

Concerning the endocrinology literature, gender inequality is well documented. As of 2021, 53% of endocrinologists in the U.S. were female, which was an increase from 44% as of 2014.\textsuperscript{1,2} In contrast, the rates of female authorship in endocrinology journals were 29.6% for first authors and 22.2% for last authors.\textsuperscript{3} Regarding the distribution of race/ethnicity, as of 2018, of all female endocrinologists, 44.5% were White, while in the same year of all Male endocrinologists, 53% were White.\textsuperscript{5,6} Given this trajectory of increasing female presence among field professionals, and a contemporaneous change in review practice at a major field journal, there is now a worthy opportunity to assess trends in authorship in the endocrine community.

Though incomplete, the preliminary evidence appears to suggest that the field transition to a female-majority practitioners is somewhat recent\textsuperscript{2}. Given this, we should expect that in a proportional evolution among clinical investigators, there would be a trend toward more female first-authors, and that any shift toward female senior-authors may still take years to manifest.

There are some noteworthy reports of authorship trends in endocrinology, including a recent report reviewing a span of seven decades \textsuperscript{3}. This report found an upward trend in female representation; however, the results suggested a persistent predominance toward male authorship, even in later years, and the analysis was performed in ten-year cross-sectional view, and without consideration of review blind.

Foremost, we interpreted First Authorship as corresponding to the investigator more junior in their career, and Last Author as the more the senior. This is a reductionistic view, that
ignores the complex subtleties of each study group that cannot be known or easily inferred, e.g. effort (time investment), contributions (creativity versus technical), or political or interpersonal considerations. We based this interpretation on common convention in science and medicine, but it is an assumption none-the-less, and therefore may not be uniformly applicable across all study groups and journals.

Separately, we motivated our research question by alluding to the relationship between diverse representation among clinicians, which is documented through institutional means (medical schools, employment firms, etcetera), and investigators, which is rather more diffuse and untracked. Indeed, studies like ours are perhaps the only way to estimate demography among authors, as few journals collect this information and -to our knowledge- none make any data publicly available. So while the observable trend toward female-majority cadre of clinicians is contextual, it is not per se a useful proxy of what could credibly be expected among a different group of professionals, i.e. researchers: it is entirely possible that the personal, societal, and infrastructural forces that influence an individual's path into one career are different-in nature or in outcome to those that influence the pathway into the other career. Alas, this point is likely more of an academic consideration: it is constructive to leverage the trends among clinicians as a presumptive template for what might be observed among researchers, but it isn't necessarily a co-dependent process, and doesn't need to be, in order to embrace this line of inquiry.

**Design Considerations**

The so-called bookend authors (i.e. first- and last authors of a publication) were chosen as representative of the writing team, since they are typically those who contributed most to the publication and are the most senior researcher on the project, respectively. We assumed that these are the authors who may potentially be more vulnerable to reviewer bias.
The selection of year 2010 as the starting point for this analysis was partly serendipitous: there were multiple sources of field demographics starting in this year, and this study group’s mainline efforts in this line of inquiry were taken up in year 2024, i.e. seven years after the conversion of Journal A to double-blind review. Thus, there was an attraction to the year 2010 as an approximately equal catchment of manuscript review in support of a before-and-after analysis. The two comparator journals (Journal A and B) were selected on the basis of their comparable scope, reputation, and impact factor. We favored PubMed as the sole access point for authorship information given its comprehensive repository and its ease of retrieval of structured article information through automated approaches.

While we found many impressive tools for ascertaining gender and race-ethnicity using just name information, we believed that GenderAPI and Rethnicity promised a well-suited combination of accuracy and convenience for use within the open-source R programming environment. While we recognize that any of these design elements could influence the outcomes of a study, our choices were made with an interest in delivering a robust analysis that was openly scrutable and supported replicability.

Limitations

Surely, any inquiry into representation in authorship would be benefitted from having some insight into who was submitting articles. Without this context, it’s impossible to draw a conclusive inference about a possible review bias. Here, we are limited to reporting on prevalence, and are not able to make any claims about propensity. Indeed, these data do not exist: medical journals do not routinely collect demographic information about the authors who
submit manuscripts for peer review, so there is simply no way to fully contextualize our findings.

Naturally, there is opportunity to increase the search scope of this work by assessing publication trends across a larger catchment of journals, or possibly longer time-span. And there are potentially fruitful comparisons to be made against other medical sub-disciplines, or by executing a deep-dive on the demographic distribution of the field of endocrinology investigators. But we feel that there is much to be learnt from even a modest inquiry, and potentially a strategic advantage in terms of clarity of interpretation. Our interest in this research was to ascertain the specific effect of double-blinding within the specialty of endocrinology; comparing one journal against two comparable counterparts is a worthy design with appropriate scope.

We recognize that there is much to be gained through the addition of thoughtful co-variates. For instance, we drafted our own Author Index heuristic, based on each investigator’s presence within our dataset. It is unknown whether a more holistic biographical measure (e.g. author h-Index, years in the field, number of publications or citations, academic rank, etcetera) would be better-suited. We chose an alternative measure due to the complexities of obtaining this information for each author. While we suppose that there could be some benefit to the model in using a more formal measure, we believe that the novel author index constructed here strikes a suitable balance between validity and feasibility. Conclusion

In this study we found that female authors are under-represented relative to their putative eligibility to participate in the authorship process, but that this trend is abating over time, and does not appear to be a trend that is specific to one journal, but rather is observed with consistency across the three journals observed. We also found that non-White authors were seemingly well-represented among authors of endocrinology papers, and that there has been a modest increase in this representation in association with the double-blind review process.
Further work could extend these findings by increasing the scope of analysis or through obtaining information beyond what is publicly available through web retrieval.

**Bibliography**


**Supplementary material**

Appendix: Dataset Preparation

*Article Accessioning and Aggregation*
Articles were retrieved from PubMed via automated retrieval through the easyPubMed package for R (v4.2.1). A two-criterion query containing journal title (field TA) and publication date (field PDAT, range: 2010-01-01 to 2023-12-31) were passed to NCBI, yielding a series of batch files containing structured article data in XML format. These batch results were interrogated for their contents and aggregated into a single dataset in delimited format (comma-separated values; .csv). The fields obtained were as follows:

- PubMed ID, DOI, journal article Title, Abstract, publication Year + Month + Day, Journal Abbreviation, Journal Title, Keywords, Author First- and Last Name, and Address of the corresponding Author.

Email data was omitted prima facie due to lack of relevance.

**Dataset Conditioning**

The raw retrieval yielded 131,504 articles. Author First- and Last-Name, and Corresponding Author address were inspected for special-characters and replaced with the nearest ASCII character equivalent. A novel author identifier was created as first-initial plus last name. Date was converted to YYYYMMDD format.

Author location being as within the United States was ascertained as follows: corresponding author address was searched for any string segment that matched one of three patterns:

Pattern 1: any two-letter state abbreviation (Alabama = AL, Alaska = AK, etcetera) followed potentially by an arbitrary number of dot-characters, comma, or a whitespace, and then any non-word character.

Pattern 2: any fully spelled state name, followed potentially by an arbitrary number of dot-characters, comma, or a whitespace.

Pattern 3: the three letters USA, followed potentially by an arbitrary number of dot-characters, comma, or a whitespace.

...And then a list of 289 country names in English language format (exclusive of “United States”) were obtained from the R countrycode::codelist utility, and utilized as a final filter for international authorships.

**Race-Ethnicity**

Author race-ethnicity was predicted via the rethnicity::predict_ethnicity function, utilizing authors’ full names (full first- and last names). This utility yields a probability score for the author being Asian, Black, Hispanic, or White. These probabilities sum to unity, and a the race-ethnicity corresponding to the maximum probability is categorically designated as the final determination. This software has an accuracy rate of less than 75% for classifying a name as White or Black. For classifications of Hispanic or Asian, this software has over 80%. We designated a 70%
probability threshold for race-ethnicity ascertainment. Any author for whom no single probability estimate exceeded 70% was designated as Unknown.

**Gender**

We used genderAPI ([https://gender-api.com/en/](https://gender-api.com/en/)) as our repository for estimating author gender, streamed through genderAPI::get_gender accessioning utility in R, using the US-based name library. The GenderApi software is made up of a database of over 6 million names from 190 countries. The accuracy of the gender prediction is based on the number of records of that name they have in the database. Following a web-retrieval in January, 2024, a crosswalk file was created listing author first-name, gender, and prediction accuracy, and this crosswalk was utilized thereafter, as a stable source of gender information.

**Article Type**

In a separate retrieval, we obtained PubMed Article type from NCBI, yielding an additional crosswalk of PubMed ID and Article Type.

**Analyte Preparation**

The dataset was converted from long-format (authors-as-rows) to wide-format (manuscripts-as-rows) with two authors (first- and last) listed per article. Manuscripts with both authors flagged as United States-based were retained; any manuscript with an international address was eliminated from the analysis stream. Superfluous variables were removed and remaining variables were renamed for convenience.

**Final Analyte**

Following the retrieval and processing steps described above, the final analyte contained the following columns:

<table>
<thead>
<tr>
<th>PubMedID</th>
<th>Article Title</th>
<th>Journal Abbreviation</th>
<th>Year</th>
<th>First Author:</th>
<th>-First Name</th>
<th>-Last Name</th>
<th>-Race Ethnicity</th>
<th>-Author Index</th>
<th>-First name as entered into genderAPI</th>
<th>-Gender as predicted by genderAPI</th>
<th>-Gender estimation accuracy, as per genderAPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last Author:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Authorship Index was tallied without consideration to international papers. For instance, we held as assumption that an author JSmith, co-authoring a paper with a corresponding address in the United States would be a different investigator from, say, JSmith, co-authoring a paper with a United Kingdom address. This decision was concession to the complexity of the task of ascertaining author identities in a high-throughput way. We thus made two assumptions in this step 1) no authors collaborated with international corresponding authors, and 2) no authors shared initnames.

The other opportunity for information-loss imposed on the authorship index is that the tally considers only the catchment of the article accessioning (i.e. starting in year 2010 and limited to the three journals). Without considering an author’s corpus outside of this pool, we are surely under-estimating the seniority of many authors, particularly those early in the dataset.

These are unfortunate concessions made in the name of feasibility and scope, with an interest in serving the grander design, i.e. obtaining a continuous co-variate to approximate author seniority. While these design decisions will ultimately limit the accuracy of this co-variate, they are pragmatic and scrutable, i.e. a “good faith effort.” We do note that this approach will happily consolidate authors whose listings variously do- or do-not use full first-name (i.e. an author listed variously as Jane Smith versus J. Smith, or even Jane. Q. Smith or J. Q. Smith). We cannot realistically know whether these decisions introduce any systematic bias with regard to the variables of interest, i.e. gender or race-ethnicity.

There was no consideration given to whether a manuscript carried a first- and last author with identical names, i.e. solo-authorship. These articles will appear in both analyses.