

Tools for Determining Equitable Representation of Women in LIS Publications

Amalia Monroe-Gulick and Marla Schleuder
University of Kansas, USA

Introduction

Librarianship has long been viewed as a “pink collar” profession, meaning a predominantly female profession. Such a gendered distinction still holds true when it is broken down into subfields such as academic and public librarianship. A 2017 article “Gender in the Journals: Publication Patterns in Political Science” by Teele and Thelen inspired this current study because we wanted to know if there was a similar gender gap in publication in a profession dominated by women as there is a profession dominated by men.¹ The article also inspired the current study because of the utilization of tools that automate the collection of data and the prediction of author gender for a large dataset.

Studying gender representation in published research is not a new area of inquiry for library and information science (LIS). There is a long history of looking at the gender gap in publishing patterns in journals within a profession that has been made up predominantly of women. However, with the emergence of new tools, different methodologies can be employed to look at larger samples that cover longer time periods to get a potentially more accurate picture of gender parity in the LIS research landscape regarding author representation.

Literature Review

While the current study was informed by a political science-focused study, there is a record of research in the LIS field as well other social science disciplines. Studies often look at different variables such as the rate of collaboration and might also happen to include gender. However, if gender is the main variable being considered, rates of citation are more likely to be looked at rather than rates of publication. Some LIS studies have found the rates of female-authored publications have increased over time, even becoming the majority (over 50%).² However, few studies have compared this to the composition of the profession as a whole.

Nisonger studies authorship patterns in LIS in order to understand the scholarly communications practices within the discipline, the subjects being presented in a given journal, and the ability to compare journals through the study of one LIS journal from 1977–1995.³ While many variables were studied, Nisonger finds the gender breakdown of the journal was 53.9% male and 46.1% female.

In a 1996 article, Terry updates an earlier study that looked at both gender and collaboration.⁴ The author examines articles from *College & Research Libraries* between 1989 and 1994 looking at gender, affiliation, and primary authorship. Terry concludes 51.7% of all authors are female, with 50% of primary authors and 51.7% of co-authors female. The results show that there is a higher rate of collaboration among women, but Terry offers no explanation for why this might be the case. Additionally, while the study finds the majority of articles are female-authored publications, it does not compare results to the overall gender composition of the profession.

A 2010 study by Reece-Evans analyzes two electronic-only LIS journals for gender and citations.⁵ The author finds that 52% of the articles are written by male authors and 46% written by female authors. The same study also looked at references within articles, and found both genders reference men at higher rates than women. When using citation analysis, the author found articles written by women authors in the

sample did receive more citations. Reece-Evans concludes “if a gendered system of accumulated advantage indeed exists, then women are at a disadvantage. Instead, female LIS professionals should enjoy the increased status, productivity, and opportunities for promotion that an equitable scholarly publishing system would bring.”⁶

One study that considers the rate of gender authorship against the profession as a whole looks at LIS monographs. In a 2019 ACRL conference paper, Tran and Nevius analyze the gender gap in LIS monograph publishing, including how gender affects the number of authors and the topic of the publication.⁷ They use eight publishers from 2015 to 2017 and end up with a sample of 431 titles. The authors compare the results of their study with the profession using the results of an unpublished ACRL membership survey. While the survey showed women comprise 77% of the profession, the average of female authorship during this time period was 63%. Tran and Nevius also find 67.2% of the collaborators were women, and women are more likely to work with other women. Finally, the authors conclude that gender “could impact the published topics in LIS books,” but it can be difficult to have definite conclusions because of “publishers’ concern for the marketplace.”⁸

Literature exists that is critical of the methods we have used in our study, with good reason. However few, if any, of these critiques detail feasible alternative methods.⁹ The literature generally falls into one of two categories: either a data collection- or data analysis-focused study that discusses a project done using similar methodology while ignoring non-binary genders, or an equity and inclusion-focused study that discusses the shortcomings of this methodology.¹⁰

Methodology

Scope

This research focuses on 10 journals that cover the major subfields of academic librarianship over a 15-year period: 2004–2019. These journals are: *College and Research Libraries (CRL)*, *Information Technology and Libraries (ITL)*, *Journal of Academic Librarianship (JAL)*, *Journal of Documentation (JOD)*, *Library & Information Science Research (LISR)*, *Library Hi Tech (LHT)*, *Library Quarterly (LQ)*, *Library Resources & Technical Services (LRTS)*, *Library Trends (LT)*, and *portal: Libraries and the Academy (PLA)*.

Process

EBSCOhost Integration Toolkit was used because the databases the journals and articles appear in are accessible through EBSCO and our institution’s subscriptions. These databases are *Library, Information Science & Technology Abstracts* and *Academic Search Complete*.

Using The University of Kansas’ *EIT Profile ID* and *Profile password* provided access to the citation data in XML format for the same articles available through EBSCO in the library’s discovery search tool. Once the syntax and forming requests through URLs became more familiar, these query URLs were stored in a CSV file, and ultimately OpenRefine was used to pull in the results. From there, the retrieved XML data was parsed out into separate columns and cleaned to remove remnants of XML code. Articles from more than 90% of the issues required were obtained using this method. There were a few journal issues that were missing, and this information was acquired with web scraping methods.

Automated Methods

Automated methods were employed to obtain data and predict a secondary variable, gender. After article metadata was obtained and organized into a usable format, the results were around 9000 articles. Spot checking was done to ensure the API had retrieved articles for each issue. The number of articles for each issue was checked against the number of articles listed for each issue on the publisher website. The

dataset was then pared down to remove non-scholarly works such as editorials, reviews, correction notices, letters to the editor, speeches, book reviews, obituaries, opinions, and table of contents to focus the study on abstracts, articles, biographies, case studies, proceedings, and reports. Articles lacking author names were looked up manually. Articles where author names could not be retrieved were removed from the dataset. This resulted in a dataset of 6191 articles (see Table 1).

In order to determine author genders and the composition of women authors for these articles, first the citation metadata for all of the articles published in these journals over the 15-year period had to be obtained. An automated method was necessary in order to do this quickly, maintain the accuracy of the data, and keep the results organized in a useful format. Each of the articles has between one and 20 authors, for a total of 11,847 authors (see Table 1). While obtaining the article metadata manually would have been overly time consuming, the same was true for author genders; an automated method was necessary to determine this variable.

Table 1: Summary of Articles and Women Authors

OVERALL	All Articles	At least 1 US-Affiliated Author	Only US-Affiliated Authors
Authors	11,847	6,338	5,836
Articles	6,191	3,386	3,224
Women authors	6,620	3,873	3,626
PER ARTICLE			
Average number of authors per article	1.91	1.87	1.81
ARTICLES WITH ONE AUTHOR			
Articles with one author	2,933	1,670	1,666
Articles with one woman author	1,587	975	973
ARTICLES WITH MULTIPLE AUTHORS			
Articles with multiple authors	3,258	1,716	1,558
Articles with multiple authors, primary author is woman	1,902	1,072	987

International data on gender in the library profession was not available, so this study focuses on data for the United States (US). One of the variables included in the article metadata is “author affiliations” (i.e., universities and institutions). Looking at author affiliations made it possible to derive the country or countries each article is associated with. In this way, the dataset was narrowed down to articles where at least one author hails from a US institution, or articles where the author(s) are only associated with US institutions. This was mainly done to compare against gender distribution in the profession in the US, since the US is one of the few countries that shares this data.

genderize and ssa methods

In order to predict gender for the authors, R's gender package was used, which has several methods available, including *genderize* and *ssa*. Each method works similarly, but stems from different sources. For example, *ssa* uses data from the US Social Security Administration from those born between 1880 and 2012. For each person recorded by the Social Security Administration, there is a birth year, a first name, and sex. For each first name and birth year or range of possible birth years, the *ssa* dataset predicts a 'gender' based on the sex most often associated with that name as well as columns that tell the percentage or proportion of individuals with that name who are associated with a particular sex. The *genderize* method uses data based on "user profiles across major social networks."¹¹

Teele and Thelen's study uses the *genderize* method through Python. While the *genderize* method is based on gender identities listed on social media profiles from all over the world, the *ssa* method uses data from real people in the United States. Initially, it appeared the *ssa* method utilized a more legitimate source, and focused on the US, making it a more suitable tool for the project than *genderize*. Either of these methods is considered more accurate than using lists of male first names and female first names, since these automated methods are based on real-world usage.

After running the package using the *ssa* method, several columns were added to the data frame: "gender" which is the predicted gender, "proportion male" and "proportion female." The *ssa* method successfully predicted genders for most of the first names in the dataset. Those that could not be identified or had a "proportion" of less than 70% were looked up by hand from online biographies through authors' affiliated institutions. Approximately 20% of names required hand coding (see Table 2).

Table 2: Hand-Coded Gender

Hand-coding determination	Number
'female-h'	781
'male-h'	1,692
TOTAL	2,473

Comparing Methods

Later it was recognized that the *genderize* method had some advantages which were initially overlooked, such as flexibility in self-reporting or altering gender, and potentially predicting genders for non-Western first names more easily. Therefore, the *genderize* method was run and results were compared to findings for the *ssa* method.

Eighty-one percent of authors had the same predicted gender value from either method (including 'female,' 'male,' and 'unknown'). This rose to approximately 90% when hand-coded findings from the *ssa* method, 'female-h' and 'male-h,' were matched against 'female' and 'male' predictions from the *genderize* method, respectively (see Table 3).

Table 3: Comparing SSA Method to genderize Method

SSA Method	genderize Method	Number of Author Names	% of all Author Names
MATCHES:			
Gender prediction values: female, male, unknown, female-h, male-h	Gender prediction values: female, male, unknown	10,639	89.80%
DISCREPANCIES:			
'unknown'	'female'	103	0.87%
'unknown'	'male'	113	0.95%
'female-h'	'unknown'	202	1.71%
'male-h'	'unknown'	266	2.25%
'male'	'female'	28	0.24%
'male-h'	'female'	66	0.56%
'female'	'male'	279	2.36%
'female-h'	'male'	144	1.22%
(other non-match)	(other non-match)	7	0.06%

While the *genderize* method appears to have predicted the gender for more names than were identified using the *ssa* method, there is a slight discrepancy between hand-coded names and what *genderize* assigned. *genderize* had a slight tendency to incorrectly predict 'male' more than it incorrectly predicted 'female' in our dataset.

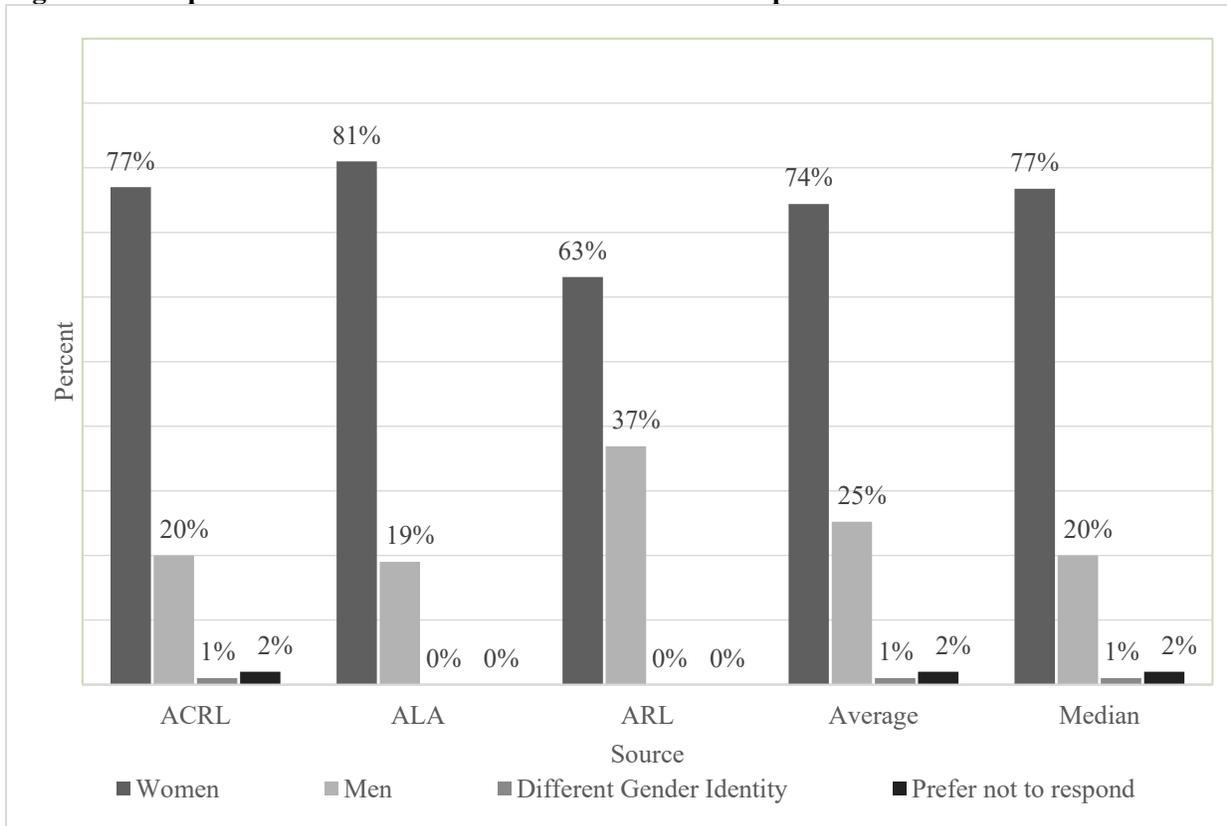
Determining Gender Distribution in the Profession

One of the driving forces when considering the question of equity is not looking to see if there is an equal number of authors for each gender. Rather, it is looking to see if there is equitable representation when considering the proportion of genders within the profession to proportion of genders of authors represented in the literature. While this cannot be a perfect one-to-one comparison, it can be one measure of equity in academic librarianship. ARL statistics are often used because they are produced annually and provide detailed demographic information (see Figure 1).¹² However, we would argue that they are not necessarily representative of the profession because they only include research libraries; compared with both ACRL and ALA, there is a potential for over-representation of men.¹³

The use of first names to predict gender is fairly common, but it treats gender as binary, posing the question: what about non-binary individuals in academic libraries? The only information available on non-binary gendered individuals working in academic libraries is from an unpublished ACRL survey done in 2018 (see Figure 1).¹⁴ It shows individuals who identify with a non-binary gender make up 1% of

the overall pool of respondents, and those who prefer not to respond regarding gender make up 2%. Because this study is focused on the number of women in the field, a majority figure, the method was not altered for the reported 3% value. This is one of the limitations of the study.

Figure 1: Comparison of Gender Distribution in Librarianship



Limitations

The documentation for R’s package ‘gender’ lists several limitations. First, it relies on data generated by government entities that do not recognize non-binary genders and therefore do not collect this information.¹⁵ Second, the package should be used to study aggregated data, rather than to study individuals. Third, this method should be used only when “a more nuanced and justifiable approach” cannot be used and the alternative would be to not do the study at all.¹⁶

The datasets used from the gender package were based on populations with a different proportion of women as compared to our dataset. Because we know that women make up approximately half of the population of the United States but (on average) 74% of the US academic library community, there is likely a slight tendency for women authors in our dataset to have their gender incorrectly predicted as ‘male.’

There are additional variables that cannot be accounted for, such as non-librarians publishing in the journals. According to a 2015 Luo and McKinny study that analyzed 10 years of JAL articles, 74.5% of single-author articles were written by librarians, and multiple-author articles had 52% all librarians and 25.2% librarian and non-librarian authors.¹⁷ The same study also determined that 17.4% of the single-author papers were written by LIS faculty. According to the 2017 membership report from the Association for Library Information Science Education (ALISE) (which treats gender as binary), representation is 50/50 female-to-male for full-time faculty. This represents a slight decrease in

representation for women over a ten-year period, which is apparent when comparing results to the 2006 survey which reported full-time female faculty at 52.3% and full-time male faculty at 47.7%.¹⁸

Results

Overall Authorship by Gender

As described in the methods sections, the sample was divided into three sub-samples because of the lack of international comparative professional data: the overall sample, a sample with at least one author from a US-affiliated institution, and a sample with only US-affiliated authors.

The overall sample of all articles is comprised of authors from institutions all over the world. This sample shows a low rate of women authors in general (56%), but increases for articles with at least one US-affiliated author (61%), and further increases for articles where all authors are only affiliated with US institutions (62%; see Figure 2). This suggests women who hail from US institutions or co-author with authors from US institutions are more likely to get work published than international peers without these affiliations. A similar pattern exists for articles with female primary authors across the three samples. Articles with women as primary authors represent the largest grouping for each sample, meaning that regardless of geographic affiliation, women are more likely to be published as primary authors for a co-authored article than published for articles written on their own. Since this figure is also higher when compared to the overall percentage of women authors in each sample, it shows women are more likely to be primary authors on a group project than they are to be published in general. Inversely, this may suggest men are more likely to participate in co-authoring papers, but are less inclined to act as primary author.

Figure 2: Women Authored Articles by Country Affiliation

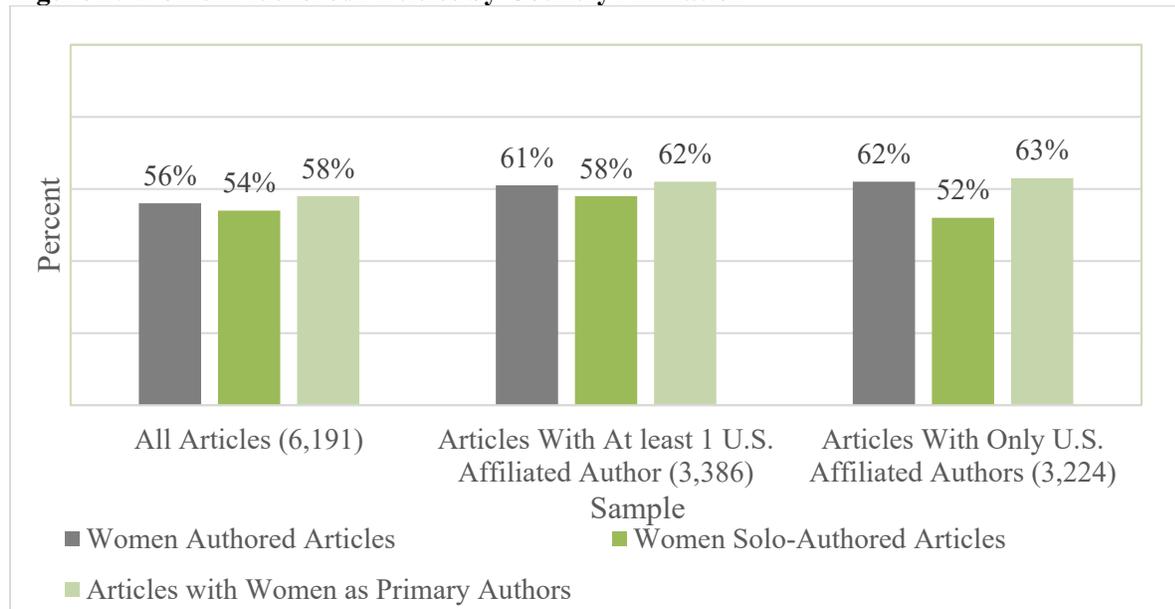


Figure 3 shows the overall results for all three samplings, highlighting the author affiliation with the additional data of the average of women in the profession. What the data also highlights is that solo-authored publications, no matter the sample, are lower than the overall authorship. The results also demonstrate that the US affiliation of authors does have an impact on the percent of women authorship. This can be due to a number of factors we have not isolated, but could include, depending on the geography, the non-US-affiliated authors coming from LIS programs rather than libraries. These programs could have a more even distribution of men and women if they are at all similar to US

institutions. No matter the breakdown or sample, there is no authorship combination that comes close to the average of women in the profession, which is 74%.

Figure 3: Women Authored Articles by Number of Authors



Gender and Journals

Over the past 16 years, four of the journals in the current study have published female authors at about the same rate in which women are represented within the profession according to ARL, 63% (see Figure 4). However, none of the journals come within 13% of ACRL’s reported proportion of women in the profession, 77%, or near our average proportion of women in the profession from various sources, 74% (see Figure 1). Three journals published women authors at an overall rate of less than 50%, and two journals published women authors at a rate of less than 60%.

Figure 4: Women Authored Articles by Journal, 2004–2019

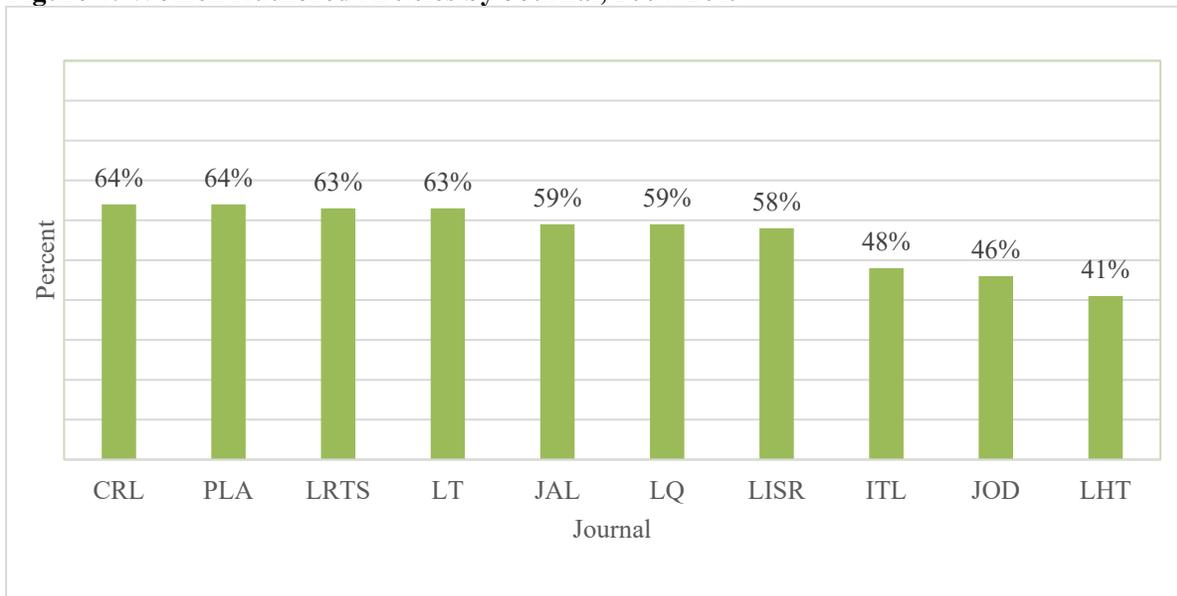
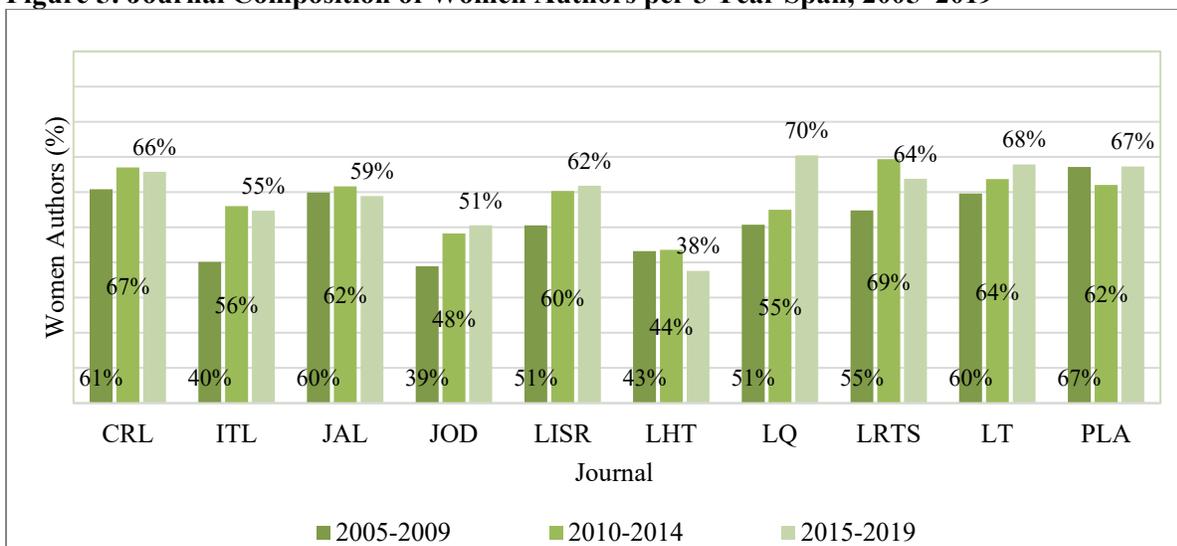


Figure 5 shows the results broken out into five-year time spans to show changes over time of gender representation in the sample with all author affiliations. Five of the 10 journals had continual increases over 15 years. Four had initial increases, then decreased, and one decreased over the five-year breakdown. *Library Hi-Tech* (LHT) never reached 50% representation of women authors in our study. The highest LHT reached was 44% in 2010–2014, and then decreased by 6% in 2015–2019 to 38%. When looking at smaller time spans, there are more indications of improvement than just the overall results might show. In fact, *Library Quarterly* (LQ) showed an increase of 15% over the last five years, from 55% to 70%, after an increase of only 4% in the previous four years. However, none of the 10 journals reached the threshold of 74%. There was an increase of four to five journals that met ARL’s threshold of proportionality within the last five years.

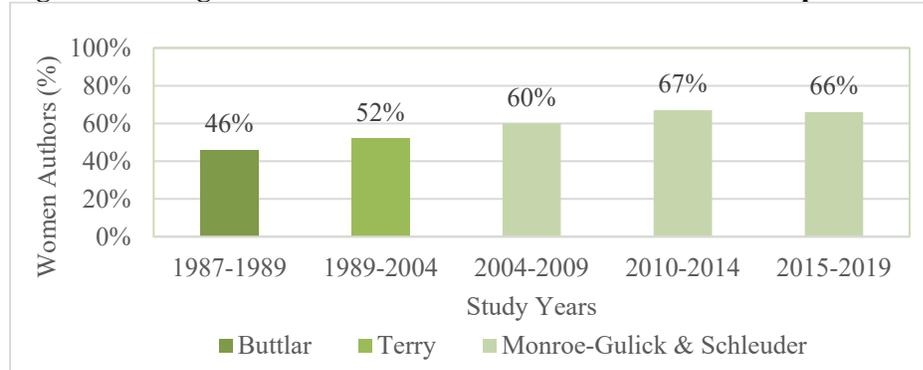
Figure 5: Journal Composition of Women Authors per 5 Year Span, 2005–2019



Finally, comparing the results of studies of one journal over time is useful in demonstrating a pattern of increasing female authorship. Based on previous studies,¹⁹ *College & Research Libraries* (CRL) has one

of the highest representations of women authors, though, if we go by the profession, it is still not entirely representative (see Figure 6). After small continuous gains in representation between 1987 and 2014, representation then stalled or dropped slightly. This is a trend we saw with four of the journals in our study: initial increases followed by decreases.

Figure 6: College & Research Libraries: Gender and Authorship Studies 1987–2019



Discussion, Conclusions, and Next Steps

Researching gender equity in published literature within a discipline is challenging. Different methods can be applied, but limitations exist with each option. Within LIS, one of the largest obstacles is deciding what determines equity because of the nature of a profession that contains wide variances, including public and academic librarians, as well as faculty within schools of library and information sciences.

If publication data alone is considered, women authors represent more than half of the 11,847 authors in the sample. However, the results also show female authorship is not proportional to the gender make-up of the profession as whole. So if equity in representation is the goal, more women need to be published in these journals, and potentially LIS research publishing as a whole.

This project required us to find a middle ground between conceptual ideals of equity and inclusion we wanted to support, and the practicality of using precedent methods and available tools and datasets. Future gender studies need to be inclusive of non-binary-gendered individuals. Organizations and institutions that collect first name and gender data should include and account for non-binary genders, but share this information only in such a way that protects the privacy of individuals. The understanding of gender as a fluid variable has become accepted in the LIS community and beyond; this understanding must also be acted upon and inform how this data is collected and managed.

What would happen if packages that predict gender based on first names were able to include ‘non-binary’ as an additional gender? Even if no first names in the near future were predicted as ‘non-binary’ with a proportion of more than 50%, records of non-binary gendered individuals would still lower the proportions for female and male predictions accordingly.

Alternatively, if a gender study is focused on a sample population with a higher percentage of non-binary individuals than the wider population from which the data for the package was gathered, it could skew the results and underrepresent the number of non-binary individuals in the focus group. This would increase the importance of aggregate data showing the gender makeup of the profession, which would provide a basis for comparison.

These methods rely on association analysis (records of multiple people that have the same first name and gender). If we reach a point where the number of non-binary-gendered individuals has increased such that

first name analysis no longer yields mostly accurate results for gender, methods like this should be abandoned altogether.

There is more to be learned from the large sample we have. We are in the process of coding the articles by subject to determine publishing trends, including whether women are publishing on certain topics within certain journals more than others. We hope this will help get to some of the “why’s” of the gender gap. The article that inspired this study also had a series of journal-specific articles with additional methods for further understanding publishing patterns. The next steps will also build on older LIS research in this area that focuses on specific journals.²⁰

There is also the consideration regarding the impact of COVID-19 on faculty productivity specifically among women. Preliminary studies at the time of this writing show submissions from women are down across several disciplines.²¹ An update of the project in approximately two years could be important when looking at not just questions of equity, but also at the potential lasting impacts of the pandemic on academic librarians’ programs of research. Publishing has long-term impacts on careers, which includes not only tenure and promotion decisions, but financial and career advancement implications as well.

—Copyright 2021 Amalia Monroe-Gulick and Marla Schleuder

Endnotes

- ¹ Dawn Langan Teele and Kathleen Thelen, "Gender in the journals: Publication patterns in political science," *PS: Political Science & Politics* 50, 2 (2017): 433–447, doi:10.1017/S1049096516002985.
- ² Peter Hemon, Allen Smith, and Mary Bailey Croxen, "Publication in College & Research Libraries: Accepted, Rejected and Published Papers, 1980–1991," *College & Research Libraries* 54, no. 4 (1993): 303–321; James L. Terry, "Authorship in 'College & Research Libraries' Revisited: Gender, Institutional Affiliation, Collaboration," *College & Research Libraries* 57, no. 4 (1996), doi:10.5860/crl_57_04_377; Ngoc-Yen Tran and Erin Nevius, "Bias in Publishing? Gender Trends in Academic Library and Information Science Monograph Publications," in *Recasting the Narrative: The Proceedings of the ACRL 2019 Conference, April 10–13, 2019*, Cleveland, Ohio (2019): 518–528.
- ³ Thomas E. Nisonger, "Authorship in library acquisitions: Practice & theory," *Library Acquisitions: Practice & Theory* 20, no 4 (1996): 395–419, doi:10.1016/S0364-6408(96)00080-4.
- ⁴ Terry, "Authorship," 379.
- ⁵ Linsay Reece-Evans, "Gender and citation in two LIS e-journals: A bibliometric analysis of LIBRES and Information Research," *LIBRES: Library and Information Science Research Electronic Journal* 20, no. 1 (2010): 1–18.
- ⁶ Reece-Evans, "Gender and citation in two LIS e-journals," 14.
- ⁷ Tran and Nevius, "Bias in Publishing," 523.
- ⁸ Tran and Nevius, "Bias in Publishing," 526.
- ⁹ Cameron Blevins and Lincoln Mullen, "Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction," *DHQ: Digital Humanities Quarterly* 9, no. 3 (2015).
- ¹⁰ Kamil Wais, "Gender Prediction Methods Based on First Names with genderizeR," *R J* 8, no. 1 (2016): 17–35.; Helena Mihaljević, Marco Tullney, Lucía Santamaría, and Christian Steinfeldt, "Reflections on Gender Analyses of Bibliographic Corpora," *Frontiers in Big Data* 2, no. 29 (2019), doi:10.3389/fdata.2019.00029.
- ¹¹ Lincoln Mullen, "gender: Predict Gender from Names Using Historical Data," *R package version 0.5.4.*, 4 (May 12, 2020): 1–7, <https://github.com/ropensci/gender>.
- ¹² Shaneka Morris, *ARL Annual Salary Survey 2018–2019* (Washington, DC: Association of Research Libraries, 2019), doi:10.29242/salary.2018-2019.
- ¹³ Kathy Rosa and Kelsey Henke, *2017 ALA Demographic Study* (Chicago: American Library Association Office for Research and Statistics, 2017).
- ¹⁴ Association of College and Research Libraries, unpublished membership survey data, 2017.
- ¹⁵ Mullen, "gender," 2.
- ¹⁶ Mullen, "gender," 2.
- ¹⁷ Lili Luo and Margaret McKinney, "JAL in the past decade: A comprehensive analysis of academic library research." *The Journal of Academic Librarianship* 41, no. 2 (2015): 123–129. doi:10.1016/j.acalib.2015.01.003.
- ¹⁸ ALISE, 2017 Statistical Reports (Westfort, MA: ALISE, 2017).
- ¹⁹ Lois Buttlar, "Analyzing the library periodical literature: Content and authorship," *College & Research Libraries* 52, no1 (1991): 38–53, doi:10.5860/crl_52_01_38.; Terry, "Authorship," 377–383.
- ²⁰ Thomas König and Guido Ropers, "Gender and editorial outcomes at the American Political Science Review," *PS: Political Science & Politics* 51, no. 4 (2018): 849–853, doi:10.1017/S1049096518000604; Dani K. Nedal and Daniel H. Nexon, "Gender in the international studies quarterly review process," *PS: Political Science & Politics* 51, no. 4 (2018): 859–865, doi:10.1017/S1049096518000628; David AM Peterson, "Author gender and editorial outcomes at political behavior," *PS: Political Science & Politics* 51, no. 4 (2018): 866–869, doi:10.1017/S104909651800063X.; David Samuels, "Gender and editorial outcomes at Comparative Political Studies," *PS: Political Science & Politics* 51, no. 4 (2018): 854–858, doi:10.1017/S1049096518000616; Carissa L. Tudor and Deborah J. Yashar, "Gender and the editorial process:

World Politics, 2007–2017,” *PS: Political Science & Politics* 51, no. 4 (2018): 870–880, doi:10.1017/S1049096518000641.

- ²¹Merin Oleschuk, “Gender Equity Considerations for Tenure and Promotion during COVID-19,” *Canadian Review of Sociology* 57 (2020):502–515, doi:10.1111/cars.12295; Jessica L. Malisch, Breanna N. Harris, Shanen M. Sherrer, Kristy A. Lewis, Stephanie L. Shepherd, Puntiwitt C. McCarthy, Jessica L. Spott, Elizabeth P. Karam, Naima Moustaid-Moussa, Jessica McCrory Calarco, Latha Ramalingam, Amelia E. Talley, Jaclyn E. Cañas-Carrell, Karin Ardon-Dryer, Dana A. Weiser, Ximena E. Bernal, and Jennifer Deitloff, “Opinion: In the wake of COVID-19, academia needs new solutions to ensure gender equity,” *Proceedings of the National Academy of Sciences of the United States of America* 117, no. (July 7, 2020): 15378–15381, doi:10.1073/pnas.2010636117.