The Impact of the COVID-19 Pandemic on Faculty Productivity and Gender Inequalities in STEM Disciplines

Raag Patel 1, Roslyn Smith 1, Satvik Ajmera 2, Dr. Monnie McGee 3

1 Master of Science in Data Science, Southern Methodist University,  
Dallas, TX 75275 USA

raagp@smu.edu

roslyns@smu.edu

ajmerasatvik@gmail.com

mmcgee@mail.smu.edu

**Abstract.** Women in STEM have faced increased prejudice regarding being published in academic journals, but the problem has grown during COVID-19. During the pandemic, women experienced comparatively lower publication rates while taking on more responsibilities. This paper examines the effect of the COVID-19 pandemic on the academic productivity metrics of women and minorities in STEM. Data from top-tier journals are analyzed to identify disparities in publication time, authorship roles, and journal prestige. The focus is on pre and post-pandemic periods. [Insert Main Result] Key findings include... The research aims to guide institutions towards equitable promotion and tenure policies. [Insert Main Conclusion] Conclusively, the study exposes practices potentially marginalizing underrepresented scholars, contributing to the equality discourse in academia.

1 Introduction

Women and minorities often encounter significant challenges in accessing, maintaining, and advancing in STEM fields. Research has shown that women face higher attrition rates in STEM fields, particularly after the birth of their first child, compared to men (Else, 2019). This indicates a structural problem within STEM fields, where professionals are discouraged from prioritizing personal lives and face penalties. Furthermore, societal expectations that place a greater burden on women to make sacrifices for family responsibilities result in slower progress and advancement of women in academia. The COVID-19 pandemic has aggravated these challenges, as women were disproportionately involved in childcare duties and educational support for their children during school closures (Anders et. al., 2020).

The pandemic has deepened structural inequities, particularly in the lab sciences (Newire, 2021). This study will investigate how the pandemic has influenced traditional measures of faculty productivity leveraged for promotion and tenure, specifically focusing on publication and submission frequency, citations of published work, editorial board membership, and scholarly organization leadership.

The research will gather information on authors, including names, citation counts, submission dates, acceptance dates, and publication dates, for the prominent journals in the statistical sciences, mathematics, physics, earth sciences, biology, and chemistry, with a possible extension to social sciences and humanities journals and books if time permits.

The analysis will use Natural Language Processing to perform web scraping, extracting information from the various journals. This approach will also enable the identification of gender-related information. It is important to note that gender will be inferred from the authors' names, which might not accurately reflect the authors' gender identity if it differs from the name-derived assumption. Once the necessary data is compiled, a time series analysis will be conducted. This analysis aims to uncover historical trends and potentially forecast future impacts if proactive measures are not taken. The supplementary analysis will delve into trends in citation volume and authorship positions to identify additional disparities that may be prevalent within the data.

The research aims to investigate potential inequities exist regarding time to publication, the proportion of first/senior authorship for women, and the prestige of journal authorship. It also explores how these dynamics have changed over time, particularly before and after the COVID-19 pandemic. Identifying these inequities exist can provide valuable insights for SMU and other institutions. This information can help guide the development of suitable strategies to redefine promotion and tenure guidelines. This can ensure that all scholarly merit is recognized and rewarded rather than overlooked. Additionally, it seeks to shed light on any invisible institutional or disciplinary practices that may exclude women are made visible.

2 Literature Review

**2.1 Women in STEM**

Women are under-represented in STEM fields. These fields consist of science, technology, engineering, and mathematics. Education and workplace environments discourage women from entering these fields, and those that do make it harder for women to succeed (Bostwick et al., 2022). Toxic workplace environments that aren't conducive to a gender-neutral working environment make it hard for a minority to have a foothold in the group. Archaic practices and a lack of support for women in the workforce benefit men more than women. However, that doesn't mean that women are not entering those fields. The number of women entering those fields at undergraduate levels has increased over the last few decades. However, still, women make up a minority of senior staff (Holman et al., 2017).

Making sure that people of all backgrounds are welcome in academia is critical to the existence of society. As we move forward, we need to make sure that all viewpoints from different backgrounds across the planet are able to be heard and broadcast to all. Focusing on helping women advance at the highest academic levels has been researched and documented for decades. This paper aims to study these effects and how they impact publishing rates for women in STEM in the highest academic field. Focusing on publication rates and understanding past studies done looking at mixed gender doctoral cohorts.

Previous work that has been done studying graduation rates for doctoral cohorts found that financial support is highly correlated with PhD completion rates when looking solely at completion rates (Abedi et.al, 1987; Ehrenberg, et.al., 1995). Other work focusing on gender finds many mixed results looking at same-gendered mentors. Female advisors did not affect female doctoral students, specifically in economics (Neumark, et.al., 1998). However, the same paper found that having a female mentor influences the time spent in graduate school (Pezzoni, et.al, 2016). Some papers found that female doctoral students with female advisors have higher publishing rates than that of female doctoral students with male advisors (Pezzoni, et.al, 2016).

The first four years of a career are traditionally important. It dictates a lot about their future regarding wages, fields, and location. It is no different in doctoral cohorts, where once awarded a doctorate, the first four years of their career dictate their future paths. An article published by Jonas Lindahl in 2018, found that those that get published in a top journal are significantly more likely to continue to get published in top journals over the next four years. He also found that those who did not get published to a top journal within the first four years of their career were significantly less likely to get published in a top journal within the 4 following years. The article then found similar results in a larger multiple logistic regression model, where being published in a top journal was the most significant predictor when looking at the success of a career.

**2.2 Post-Pandemic Gendered Effects**

In the first half of 2020, journals submitted to the top specialist journals across all STEM fields, about 36% of those being from women. In the first half of 2022, an article written by Jocalyn Clark, found that percentage dropped to about 23%. During COVID-19 (this paper refers to this era from 2020-2022), articles related to COVID-19 received an average of 20 times more citations compared to non-COVID-19 related articles. This spike in citations in expected give the global crisis. However, among the top 45 academic authors, each with a minimum of 60 papers published in the first 18 months (1 and a half years) of COVID-19, only 5 (11%) of those were women (Clark, 2023). Clark further disclosed that, of the top medical journals in the U.K., the top academic authors comprised no higher than 40% women authorship. Some journals barely had 1 out of 10 top academic authorships be women.

COVID-19's effect on the world led to mass deaths and fear amongst the population and shut down many places where people gathered. Social networking places such as conferences were closed, and schools were moved remotely. It is well documented that women faced more adverse challenges than their male counterparts during the move to remote schools. Traditional families found women taking on the household and children caring roles. With conferences shut down, collaborations ended, as well as limitations in accessing people. Female academic sponsors and mentors were harder to access, and as this paper cited earlier, there may be some significance to female doctors mentoring female doctoral students.

Misty Heggeness found that women in states that shut down in the beginning stages of the pandemic were 31% more likely to take leave than men. Heggeness continued to find that women with children attending K-12 (elementary to high school) education were 36% more likely to stay at home than men who also had children attending K-12 education.

There is a significant disproportion amongst academic leadership as well. Organizational work and tenure policies make it harder for women to succeed, as they don't account for some women's desire for a family (Higginbotham, 2021). Universities struggle to highlight and understand the importance of a good work-life balance, even though randomized controlled studies have shown that it completely, helps the university produce more papers that are published in top journals (Hammer, et.al., 2011). Women who hold high academic positions, such as presidency or provost, are less likely than men to be married and even less likely to have children (ACE, 2017).

**2.3 Web Scraping for Journal Authorship**

Capturing author names from journals can be challenging due to the varying formats and structures of articles across different publications. Common approaches for sourcing this data include using APIs, capturing metadata from journal websites and databases, and leveraging open-access datasets.

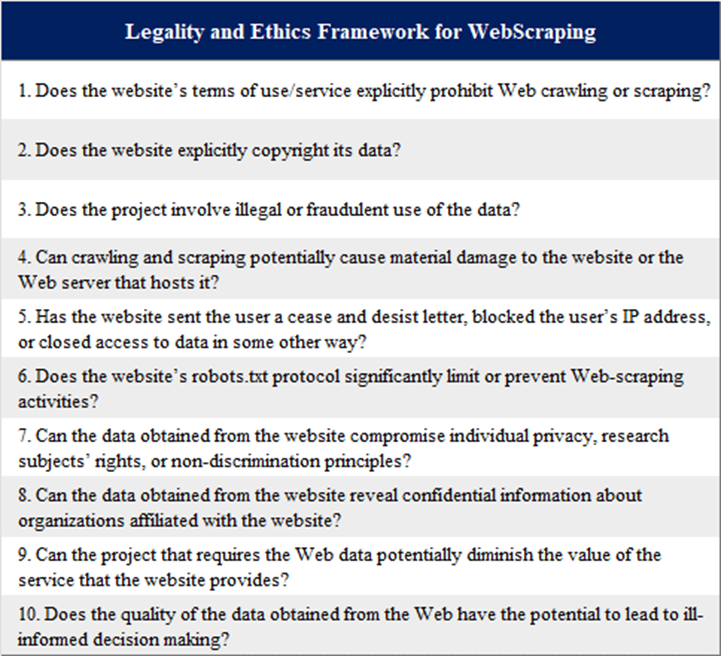
However, there are scenarios where APIs may not fully support the requirements, such as when a site lacks API support or when API usage is difficult or expensive. In such cases, web scraping, the process of automatically extracting data from websites, becomes a valuable tool for data extraction. Web scraping enables data collection from multiple web pages, automates data collection, and extracts specific details of interest. Setting up web scraping can be relatively simple with minimal programming effort (Glez-Pena et al., 2014). Programming languages such as R and Python provide web scraping libraries that facilitate the crawling and parsing of web data. Additionally, user-friendly tools, available both as cloud-based platforms and desktop applications, simplify the process of capturing data.

Sometimes, the available libraries or web scraping tools may not capture the exact information needed without manual intervention. Researchers have explored optimization algorithms to further enhance the accuracy of sourcing article citation information. One proposed algorithm, the "Firefly Optimization Algorithm based Web Scraping," combines principles from the Firefly Optimization Algorithm with web crawling and web scraping techniques (Suganya et al., 2021). This algorithm adapts the behavior of fireflies to guide the extraction of author information from web citation databases. By leveraging concepts from Particle Swarm Optimization and the Hidden Markov Model, the algorithm aims to improve the accuracy and efficiency of extracting citations. Evaluations have shown that the proposed algorithm outperforms existing methods.

There is also a lack of a complete and freely accessible catalogue of all scientific publishers and their journals. Existing databases could have biases or omissions, which must be considered (Nishikawa-Pacher, 2022). Nishikawa-Pacher aimed to identify the 100 largest scientific publishers by the number of journals published. Using Scopus, Publons, DOA, and Sherpa Romeo to capture a comprehensive list of the publishers and their journals, Nishikawa-Pacher then gathered data at a journal level to find publishers with 15 titles or more. This methodology resulted in a more comprehensive list of journals and articles than sourcing from one database.

It is essential to note that legal and ethical considerations must be considered when performing web scraping. Respecting website terms of service, avoiding overloading servers with excessive requests, and being aware of any restrictions on data usage or scraping imposed by the website are fundamental in this process. Neglecting these aspects of web scraping can lead to ethical controversies and even lawsuits (Krotov et al., 2020).

Krotov highlights that court cases involving disputes over web data often revolve around legal frameworks such as data access, copyright infringement, and trade secrets. They also argue that using web data can unintentionally compromise individuals' privacy, violate rights, contribute to bias, or reveal confidential information. Additionally, certain data usage methods may diminish a website's value. To proactively avoid legal issues, protect reputations, and address these concerns, Krotov proposes the "Legality and Ethics Framework for Web Scraping." This framework considers the legal and ethical dimensions of web scraping and includes ten questions that help researchers identify potential legal or ethical controversies associated with their web scraping projects.



**Fig. 1.** Legality and Ethics Framework for WebScraping as proposed in Tutorial: Legality and Ethics of Web Scraping (Krotov et al., 2020). This framework shows that between illegal and unethical activities lies a grey area of web scraping that must be carefully considered to ensure legality and ethics.

**2.4 Name Based Gender Classification**

The ability to accurately classify the gender of authors based on their names is crucial for the success of this research. However, several challenges are associated with identifying gender from names, including cultural variations, unisex names, multiple spelling variations, personal preferences, and evolving name trends. It is important to note that capturing the complexity and diversity of gender identities may be beyond the scope of this research.

Various methodologies and programs exist for determining gender from names. Databases like the one maintained by the Social Security Administration offer statistical information on the gender distribution of names based on historical data. Gender API services such as Gender-API and Genderize utilize algorithms and statistical models to predict gender based on names. Machine learning models can also be trained using large datasets with labeled names and genders.

A study investigating gender classification using machine learning models found that a linear model with feature engineering performs comparably to more complex models like neural network-based models or the language model BERT (Hu et al., 2021). The study also explored the effectiveness of incorporating first and last names, particularly in cultural variations in the gender connotations of first names.

Another method for inferring gender based on statistical characteristics of names achieved an 80% accuracy in predicting the gender of users (Mueller et al., 2016). The study leveraged features from 1 million Twitter users, including name length, number of vowels, number of consonants, number of syllables, and frequency of occurrence of the name in the dataset. This approach addressed issues with ill-formed names, fictitious names, or nicknames. Additional features such as user bio, location, user following count, and user follower count were incorporated to improve accuracy.

The growing interest in examining and explaining gender inequalities across various fields has led to the development of several services that offer accurate methods for inferring the gender from names. These services use extensive databases of names enriched with sociolinguistic information, culture-specific rules, and insights from social media profiles. However, it’s important to note that underlying data sources are often closed, raising concerns about their reliability and verifiability.

In a study by Suganaya et al. (2021), five name-to-gender inference services were analyzed to compare and benchmark their performance. The evaluated services were Gender API, genderize.io, NameAPI, NamSor and the Python package gender-guesser. The testing was conducted on a manually labeled dataset containing 7,076 names, focusing on misclassifications (assigning the incorrect gender to a name) and non-classifications (cases where prediction gender is impossible).

Through repeated, cross-validated, randomized searches with thresholds set at a maximum of 5% misclassifications and a requirement of at least 75% of all names to be assigned to a gender, Gender API demonstrated the best performance in the benchmarks, followed by NamSor. The study also found that names of Asian origin tended to have less confident predictions.

Considering the evolution of large language models and conversational A.I. tools, the study suggests exploring additional tools and methodologies to further enhance the characterization of gender from names.

**3 Methods**

In this study, the aim was to gather relevant data to analyze the disparities between men and women as authors in academic journals at the Ph.D. level. This required obtaining a substantive dataset encompassing varied articles from numerous journals across a select range of years.

Crossref, which was utilized to extract the desired data, offers an API that provides a myriad of metadata about academic articles, including the names of authors, publication dates, journal titles, and more. The API is not only easily accessible but also free to use, which significantly enhances its applicability for research purposes without imposing financial constraints. It operates on a broad network, accumulating metadata from numerous sources, which gives researchers access to a large, diverse, and rich dataset. This wide range of available data is crucial for conducting comprehensive and detailed analyses in academic research, such as exploring gender disparities in journal authorships.

Distinctive in its extensive subject matter coverage, Crossref stands out amidst other platforms and databases such as PubMed, Scopus, or Web of Science. The latter databases focus on specific subject domains, limiting their scope of content to particular fields of study. For example, PubMed is notably biomedical and life sciences-oriented, while others might be constrained to their particular realms of specialty. Contrastingly, Crossref gives access to metadata from over 200 diverse subject areas. This extensive subject matter range enables researchers to access and explore data from a multitude of disciplines, facilitating analyses and cross-disciplinary research. The inclusivity of various academic fields in a single platform enhances its utility.

The employed Python code, detailed in the appendix, serves to illustrate the methodology deployed in the data extraction process from Crossref. This study aimed to amass data concerning academic articles, employing the requests library in Python to interact with the Crossref API. Queries were formulated to yield results corresponding to predetermined criteria, notably the publication year and subject. The data obtained was subsequently processed and formatted into a structured format utilizing the pandas library, facilitating subsequent data analysis.

The initial step undertaken in the code aimed to procure a broad understanding of available subjects within the Crossref database. This was accomplished by extracting unique subjects from a subset of 1,000 articles. Following this, the main data retrieval function, get\_articles\_by\_subject(subject, year, rows=rows), was devised to extract articles corresponding to a particular subject and publication year, thereby enabling the code to filter and retrieve articles based on these delineated parameters. The function employs pagination, interfacing with the Crossref API, to ensure thorough retrieval of available articles and thus, ensuring a comprehensive dataset.

Notably, the publication years specified in the queries were selectively chosen to afford this study a temporally stratified view of the disparities between male and female authorship in academic journals at the Ph.D. level across distinct periods: 2017-2019 were chosen to provide a pre-COVID metric, 2020/2021 to encapsulate data from the COVID period, and 2022/2023 to potentially illuminate any post-COVID shifts in the gender authorship dynamic. This deliberate temporal segmentation was designed to enable an analysis that could identify and explore any potential impacts or shifts resulting from the global pandemic on gender disparities in journal authorship.

Upon gathering articles pertinent to the specified subject and year, the metadata is extracted, formatted, and organized into a DataFrame for each subject. The extracted information includes DOI, Title, Container Title (Journal Name), Publisher, Publish Date, Author First Name, Author Last Name, Author Order, and the number of times the article has been referenced. It also includes a 'Year' column to demarcate the publication year of each article, facilitating subsequent analysis.

A notable strength of this methodology is the systematic and scalable approach to data retrieval, which facilitates customization and replication for varied analyses in future research. Nevertheless, the study is not devoid of potential limitations and challenges, which will be deliberated upon in the succeeding sections of the discussion.

In the subsequent phase of the analysis, after retrieval of the data, this study adhered to a categorized approach to streamline and coherently structure the diverse array of subjects. The categorization was inspired by the established groupings utilized by the National Center for Education Statistics (NCES), which is a federal entity in the United States that collects, analyzes, and disseminates statistical information related to education in the U.S. and other nations. Using these predefined groupings ensures a systematic and comprehensible organization of subjects which aligns with recognized educational and research classifications.

The subjects retrieved from Crossref were systematically distributed into several broad thematic categories, resembling those utilized by the NCES, to facilitate an organized, hierarchical analysis. These thematic entities encompassed 'Life Sciences', 'Physical and Earth Sciences', 'Mathematics and Computer Sciences', 'Psychology and Social Sciences', 'Engineering', 'Education', 'Humanities and Arts', and a residual category termed 'Other'. This approach enables the establishment of a macro-view, thereby offering a structured framework through which disparities in authorship can be analyzed across broader research and academic themes.

The organized thematic areas were meticulously defined by allocating each individual subject, extracted from the Crossref data, to the most relevant category. The categorization process was executed to ensure that every subject was precisely mapped to one overarching theme, thus providing a coherent foundation for subsequent analyses. Each subject was accordingly associated with a respective category, establishing a systematic mapping that would facilitate a meaningful, categorized exploration of the data.

In instances where subjects did not naturally align with the predefined categories — for instance, subjects that did not concur with the classification provided by the NCES — they were assigned to the 'Other' category. This categorization provides a foundational basis for analyzing disparities in authorship across various scientific domains, offering insights into gender disparities within and across these broad thematic categories.

To determine the gender associated with each name, two primary methodologies were employed. First, a gender library named "gender guesser" was used. Gender guesser is a Python library designed to predict gender based on first names. It uses datasets from multiple countries to derive probable gender outcomes. While it's quite efficient for common names, its accuracy can wane when confronted with less prevalent or culturally diverse names.

Subsequent to the "gender guesser" output, a tool named "namesex" was employed. Namesex is designed to ascertain the likely gender of a name based on historical and cultural associations. However, like many such tools, it may not always be comprehensive in its recognition of global names, especially those that are less conventional or region-specific.

Moreover, technical challenges may arise due to names with special characters. For instance, names from Nordic languages might have characters like 'å', 'ä', or 'ö'. When these characters are not processed correctly, or if the library isn't equipped to recognize them, they can lead to misclassification or a complete lack of gender assignment.

Further complicating matters is the financial feasibility of using more sophisticated and comprehensive tools. While there are numerous APIs available that can provide gender predictions with higher accuracy, they come at a cost. Many of these APIs use a token-based charging system. Essentially, every word or piece of data processed by the API consumes a certain number of tokens. For instance, a text of about 750 words might utilize approximately 1000 tokens. Given the sheer volume of names in larger datasets, the costs associated with using these APIs can escalate rapidly. In addition to cost, many APIs impose limits on the volume of data that can be sourced through it’s services within a certain timeframe. These considerations make them impractical for extensive datasets.

While tools like "gender guesser" and "namesex" offer valuable insights for gender prediction, they are not without limitations. Factors like international name diversity, special characters, and the prohibitive costs of advanced APIs contribute to challenges in obtaining an accurate and comprehensive gender breakdown from datasets. As we move forward, it underscores the importance of integrating more inclusive databases and cost-effective methodologies to address these challenges.

**4 Results**

A graph of blue rectangular objects

Description automatically generated with medium confidence

The bar chart illustrates the distribution of unique titles across various academic disciplines. The "Life Sciences" category stands out with the highest number, nearing 140,000 titles. Both "Engineering" and "Physical and Earth Sciences" follow closely, each contributing over 100,000 titles. "Humanities and Arts" has roughly half the count of "Engineering", while "Mathematics and Computer Sciences" and "Psychology and Social Sciences" exhibit comparable figures, both slightly below 100,000. This visualization emphasizes the robust research activity in "Life Sciences", with "Engineering" and "Physical and Earth Sciences" also demonstrating substantial academic contributions.

|  |  |  |
| --- | --- | --- |
| **Gender** | **Count** | **Percentage** |
| Male | 5,322,905 | 34.28% |
| Female | 2,987,054 | 19.24% |
| Unknown | 7,218,833 | 46.49% |

The table above provides a breakdown of the demographic composition based on gender, inferred from a dataset of names. Specifically, the data reveals the count and corresponding percentage for each gender category, including "Male", "Female", and "Unknown".

The table reveals an intriguing point: nearly 46.49% of the names were categorized as "Unknown". This substantial percentage raises several considerations. International names, particularly those outside Western nomenclature, can be a stumbling block for many gender prediction tools. Names from various regions in Asia, Africa, or indigenous communities might not be well-represented in the databases these tools rely on.

**Placeholder for Chart with the volume of articles published in each year.**

Examining the volume of article publications within this timeframe (source), a consistent trend is found, marked by a conspicuous dip in 2021. This decline might be attributed to the significant effects of the global pandemic, influencing the ability of various industries to maintain regular operations. The headwinds posed by the pandemic, such as disruptions to research activities and shifts in priorities, likely contributed to a temporary decrease in published articles during that year. However, a notable surge in 2022 suggests a potential compensatory effect, indicating that articles initially scheduled for publication in 2021 might have been postponed to 2022. This surge could be indicative of a backlog of research output finding its way into publication, potentially reflecting adjustments and adaptations within the academic and research communities.

**Placeholder chart for actual numbers – Need to replace with actual numbers**

A graph showing the number of covid-19

Description automatically generated

Authorship Rates per Gender by year in Pre-COVID, COVID and Post-COVID timeframes

After completing the gender inference based on first names the data indicates a consistent trend of increased gender diversity among authors over the years. In 2017, male authors constituted 68% of the total, while female authors represented 32%. Subsequent years witnessed a gradual shift, with the proportion of male authors decreasing to 65% and conversely, the percentage of female authors experiencing a steady rise, reaching 35% by 2023.

**Some charts and write up of breakouts by subject?**

In reviewing the trends by subject area, we saw similar trends except in Computer Science. In this area, we saw male authors continue to increase in the proportion of authors and conversely, female authors decline. The underrepresentation of women in Computer Science has been an ongoing issue for decades with multiple factors contributing to this disparity. There are cultural and societal stereotypes that have played a significant role in creating and perpetuating the notion that men are more suited for technical roles which could deter women from pursuing or persisting in the field. In addition, some women who enter the Computer Science academia or tech industry report experiencing discrimination, microaggressions, or feelings of isolation due to being in the minority.

**5 Discussion**

The research aimed to show the disparity in journal authorship between males and females and the impact of the COVID-19 global pandemic in this dynamic. The examination of article publication trends revealed a noticeable dip in 2021, potentially attributed to the disruptive impact of the global pandemic. This aligns with expectations considering the challenges faced by various industries during this period. However, the surge in 2022 suggests a corrective response, indicating that articles initially scheduled for 2021 publication may have experienced delays.

The gender breakdown of authors over the years portrays a positive shift towards gender parity. The decreasing proportion of male authors and corresponding increase in female authors suggest and improving gender diversity trend. The shift is consistent with the broader societal conversation about gender equity indicating progress in the representation of women as authors in academic journals at the Ph.D. level.

While we see this trend based on this data, there are additional headwinds that have come up that could prove detrimental to gender parity. Affirmative action policies in educational institutions were implemented to address and correct historical inequalities. The end of such policies might reduce the number of women, admitted to STEM programs at institutions that have utilized these policies in the past.

The end of American Rescue Plan Act signed into law in 2021 ended on September 30th 2023. This plan provided funding for childcare centers across the country and was key to making child care more affordable for many parents. There are approximately 600,000 child care centers in the US of which over 220,000 child care programs received funds from the law (Office of Administration for Children & Families). The expiration of this funding is expected to result in the closing of 70,000 child care programs (The Century Foundation). This is projected to impact millions of parents with many leaving the workforce or reducing their hours which could further slowdown the path to gender equality.

Organizations such as Mom’s First was founded in response to the challenges the pandemic unveiled about motherhood in America. Central to the strategy of this group is a plan targeting three main areas: the transformation of workplaces by collaborating with both employees and executives to make them more mother-friendly; cultural shifts through creative campaigns and thought leadership to change the way motherhood is seen and valued; and governmental advocacy in partnership with other organizations to champion crucial policies like paid leave, childcare and direct financial support for mothers.

It will be important to continue monitoring key indicators showing gender inequality on the path to parity. As more time passes since the global pandemic, additional research should be conducted in this area to determine if the trends in the volume of publication for 2022 were just a correction or a new standard. Additionally, new data will need to be evaluated to see if the increase in female authorship continues in 2023 and beyond. Having visibility into these topics will be crucial for driving strategy and policy changes that create an equitable space for all genders to contribute.

**6 Conclusion**

2 paragraphs max on the overall findings and summary of the research.

In conclusion, this study undertakes a multi-dimensional examination of gender disparities in STEM academia, leveraging a wide range of data analysis methodologies and comprehensive data from six prominent databases. By dissecting these disparities through the lenses of authorship rates, publication rates, citation counts, and journal prestige, we aim to uncover the depth and breadth of gender inequalities in these fields. Furthermore, our study provides a unique perspective by factoring in the possible impact of the COVID-19 pandemic on these disparities.

Ultimately, we hope that our findings will not only shed light on the challenges faced by underrepresented groups in STEM academia but also inform institutional policies and practices. By unveiling potential biases and inequities in the current system, we can contribute to the ongoing dialogue about diversity and inclusivity in STEM. We believe that understanding and acknowledging these disparities is the first step towards creating an academic environment that truly values and supports all its members.

**Acknowledgments.** The heading should be treated as a 3rd level heading and should not be assigned a number.

**References**

1. Bostwick, V. K., & Weinberg, B. A. (2022). Nevertheless she persisted? gender peer effects in doctoral STEM programs. Journal of Labor Economics; J Labor Econ, 40(2), 397-436. https://doi.org/10.1086/714921
2. Holman, L., Stuart-Fox, D., & Hauser, C. E. (2018). The gender gap in science: How long until women are equally represented? PLOS Biology, 16(4). https://doi.org/10.1371/journal.pbio.2004956
3. Glez-Pena, D., Lourenco, A., Lopez-Fernandez, H., Reboiro-Jato, M., & Fdez-Riverola, F. (2014). Web scraping technologies in an API world. Briefings in Bioinformatics; Brief Bioinform, 15(5), 788-797. https://doi.org/10.1093/bib/bbt026
4. Hu, Y., Hu, C., Tran, T., Kasturi, T., Joseph, E., & Gillingham, M. (2021). What's in a name? – gender classification of names with character based machine learning models. Data Mining and Knowledge Discovery, 35(4), 1537-1563. https://doi.org/10.1007/s10618-021-00748-6
5. Krotov, V., Johnson, L. R., & Silva, L. (2020a). Tutorial: Legality and ethics of web scraping. Communications of the Association for Information Systems, 47, 539. https://doi.org/10.17705/1CAIS.04724
6. Mueller, J., & Stumme, G. (2016). Gender inference using statistical name characteristics in twitter. Paper presented at the https://doi.org/10.1145/2955129.2955182
7. Santamaria, L., & Mihaljevic, H. (2018). Comparison and benchmark of name-to-gender inference services. PeerJ.Computer Science; PeerJ Comput Sci, 4, e156. https://doi.org/10.7717/peerj-cs.156
8. Suganya, E., & Vijayarani, S. (2021). Firefly optimization algorithm based web scraping for web citation extraction. Wireless Personal Communications, 118(2), 1481-1505. https://doi.org/10.1007/s11277-021-08093-z
9. Clark, J. (2023). How pandemic publishing struck a blow to the visibility of women's expertise. BMJ, 381, p788. https://doi.org/10.1136/bmj.p788
10. National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Women in Science, Engineering, and Medicine; Committee on Investigating the Potential Impacts of COVID-19 on the Careers of Women in Academic Science, Engineering, and Medicine; Dahlberg ML, Higginbotham E, editors. The Impact of COVID-19 on the Careers of Women in Academic Sciences, Engineering, and Medicine. Washington (D.C.): National Academies Press (U.S.); 2021 Mar 9. Summary. Available from: https://www.ncbi.nlm.nih.gov/books/NBK570960/
11. Heggeness, M. L. (2020). Estimating the immediate impact of the COVID-19 shock on parental attachment to the labor market and the double bind of mothers. Review of Economics of the Household. https://doi.org/10.1007/s11150-020-09514-x
12. Hammer, L. B., Kossek, E. E., Anger, W. K., Bodner, T., & Zimmerman, K. L. (2011). Clarifying work–family intervention processes: The roles of work–family conflict and family-supportive supervisor behaviors. Journal of Applied Psychology, 96(1), 134–150. <https://doi.org/10.1037/a0020927>
13. Nishikawa-Pacher, A. (2022). Who are the 100 largest scientific publishers by journal count? A webscraping approach. Journal of Documentation, 78(7), 450–463. <https://doi.org/10.1108/JD-04-2022-0083>
14. Feldon, D. F., Peugh, J., Maher, M. A., Roksa, J., & Tofel-Grehl, C. (2017). Time-to-credit gender inequities of first-year PhD students in the biological sciences. CBE Life Sciences Education; CBE Life Sci Educ, 16(1), ar4. https://doi.org/10.1187/cbe.16-08-0237
15. King, M. M., & Frederickson, M. E. (2021). The pandemic penalty: The gendered effects of covid-19 on scientific productivity. Socius: Sociological Research for a Dynamic World, 7, 237802312110069. <https://doi.org/10.1177/23780231211006977>
16. Lundine, J., Bourgeault, I. L., Clark, J., Heidari, S., & Balabanova, D. (2018). The gendered system of Academic Publishing. The Lancet, 391(10132), 1754–1756. https://doi.org/10.1016/s0140-6736(18)30950-4
17. Deutsch, C., & Paraboni, I. (2023). Authorship attribution using author profiling classifiers. Natural Language Engineering; Nat.Lang.Eng, 29(1), 110-137. <https://doi.org/10.1017/S1351324921000383>
18. Filardo, G., da Graca, B., Sass, D. M., Pollock, B. D., Smith, E. B., & Martinez, M. A. (2016). Trends and comparison of female first authorship in high impact medical journals: Observational study (1994-2014). BMJ (Online); BMJ, 352, i847. <https://doi.org/10.1136/bmj.i847>
19. Yu, M., Krehbiel, M., Thompson, S., & Miljkovic, T. (2020). An exploration of gender gap using advanced data science tools: Actuarial research community. Scientometrics, 123(2), 767-789. <https://doi.org/10.1007/s11192-020-03412-w>
20. Lindahl, J. (2018). Predicting research excellence at the individual level: The importance of publication rate, top journal publications, and top 10% publications in the case of early career mathematicians. Journal of Informetrics, 12(2), 518-533. <https://doi.org/10.1016/j.joi.2018.04.002>