



RESEARCH ARTICLE

Gender differences in citation impact for 27 fields and six English-speaking countries 1996–2014

Mike Thelwall 

Statistical Cybermetrics Research Group, University of Wolverhampton, UK

an open access  journal



Citation: Thelwall, M. (2020). Gender differences in citation impact for 27 fields and six English-speaking countries 1996–2014. *Quantitative Science Studies*, 1(2), 599–617. https://doi.org/10.1162/qss_a_00038

DOI:
https://doi.org/10.1162/qss_a_00038

Received: 15 October 2019
Accepted: 11 January 2020

Corresponding Author:
Mike Thelwall
m.thelwall@wlv.ac.uk

Handling Editor:
Ludo Waltman

Copyright: © 2020 Mike Thelwall.
Published under a Creative Commons
Attribution 4.0 International (CC BY 4.0)
license.



Keywords: academic careers, citation impact, field differences, gender differences, research evaluation

ABSTRACT

Initiatives addressing the lack of women in many academic fields, and the general lack of senior women, need to be informed about the causes of any gender differences that may affect career progression, including citation impact. Previous research about gender differences in journal article citation impact has found the direction of any difference to vary by country and field, but has usually avoided discussions of the magnitude and wider significance of any differences and has not been systematic in terms of fields and/or time. This study investigates differences in citation impact between male and female first-authored research for 27 broad fields and six large English-speaking countries (Australia, Canada, Ireland, New Zealand, the UK, and the USA) from 1996 to 2014. The results show an overall female first author citation advantage, although in most broad fields it is reversed in all countries for some years. International differences include Medicine having a female first author citation advantage for all years in Australia, but a male citation advantage for most years in Canada. There was no general trend for the gender difference to increase or decrease over time. The average effect size is small, however, and unlikely to have a substantial influence on overall gender differences in researcher careers.

1. INTRODUCTION

Gender differences have disappeared or greatly shrunk in many areas of life, such as academic achievement at school and employment rates, reflecting broadly similar psychological capabilities (Hyde, 2005). They are nevertheless pervasive in some aspects of life in many countries, such as average income and job choices. In academia, the proportion of women is increasing overall, with women forming the majority in some fields, such as nursing and psychology (depending on the country). Nevertheless, some fields have been slow to recruit women and there is a widespread problem with a lack of women in senior positions (Solera & Musumeci, 2017). Initiatives to redress the balance, such as Athena SWAN (UK) and NSF-ADVANCE (USA) need to be informed about the reasons for the continuing problems, whether they include sexism, systemic bias, or other factors (Van Miegroet, Glass, et al., 2019). Citation impact may influence gender disparities when it is considered during funding, appointment and promotion decisions. Previous studies have found that women are more, equally, or less cited than men overall, depending on country and field (Elsevier, 2017; Larivière, Ni, et al., 2013; Thelwall, 2018a). For example, within medicine, a small male first author citation advantage has been attributed to greater male self-citations, journal prestige, and international collaboration (Andersen, Schneider, et al., 2019), but another medicine-based study found that

male greater self-citation was due to greater male output (Mishra, Fegley, et al., 2018), presumably due to male researchers being older, on average, and taking fewer career breaks and periods of part-time working for carer responsibilities. There is no clear overall pattern to the direction of citation gender disparities in terms of time, discipline, or country and most prior studies have not discussed the effect size of any difference. These shortfalls need to be addressed systematically on a large scale to draw conclusions about where gender differences in citation rates are important enough to need addressing in any country or field.

Several explanations have been proposed for field variations in gender differences in academia, typically focusing either on participation or promotion. Science, Technology, Engineering, and Mathematics (STEM) subjects have often been the focus, due to low female participation in many. Explanations have sometimes explored early life patterns and college degree choices. Explicit bias by senior male academics against women in some areas is a logical possibility, as are mainly male informal “old boys” circles in some that hire, promote, or reward insiders. More subtly, men may not appreciate the value of the work of women if they tend to work on different goals, with different methods or with different working practices. Some of these build into the idea that subjects can have “chilly climates” for female applicants and participants (Simon, Wagner, & Killion, 2017; Walton, Logel, et al., 2015), making them feel unwelcome, unappreciated or out of place. At least for STEM subjects in the USA, however, explicit bias does not seem to be an influential determinant of academic career outcomes (Ceci & Williams, 2011). This is a controversial issue and there are gender differences in academics’ and others’ assessment of the strength of evidence for and against the influence of gender bias in academia (Handley, Brown, et al., 2015; Moss-Racusin, Molenda, & Cramer, 2015). Moreover, it is not clear why bias might occur in fields that remain male-dominated, such as mathematics, but not in those that have shifted from male to female, such as veterinary science and psychology. Gender differences in abilities are also unlikely explanatory variables, because they seem to be minor and cover specialist tasks (Hines, 2011). There are stronger differences in field-related adult knowledge and expertise, but these may be accounted for by school-age social factors leading boys and girls toward different hobbies and school subjects (Ceci & Williams, 2011).

An explanation for field differences in participation with empirical evidence from vocational psychology (about largely nonacademic careers) is that women are more likely to have communal career goals, wishing to help society in their careers and wider lives, whereas men are more likely to prioritize self-advancement (Diekmann, Steinberg, et al., 2017). Within academia, this may translate into women being more likely to choose obviously socially helpful fields, such as education, nursing, medicine, social work, and immunology, rather than a more abstract subject, such as mathematics, or a more indirect field, such as engineering, politics, or computer science. Although there is no direct evidence for this hypothesis within academia, women in the USA and India have been shown to be more prevalent in people-related subjects (Thelwall, Bailey, et al., 2019a, 2019b), which are likely to be more directly socially helpful (e.g., nursing). If women are more likely to choose an academic subject for its affordance of communal benefits, then it is possible that women would also be more likely to target wider societal impact for their work. There is a little evidence in support of this in the form of apparently greater educational impact for research authored by women (Thelwall, 2018b).

Previous field-based studies of gender differences in citation impact have usually taken one of three approaches, with different interpretations that should not be conflated. Studies comparing career statistics (e.g., total citations, *h*-index, proportion of top-cited papers over a long period) have tended to find that men are cited more often. These reflect biases against women (for a comparison, see Reed, Enders, et al., 2011) because career statistics do not take into

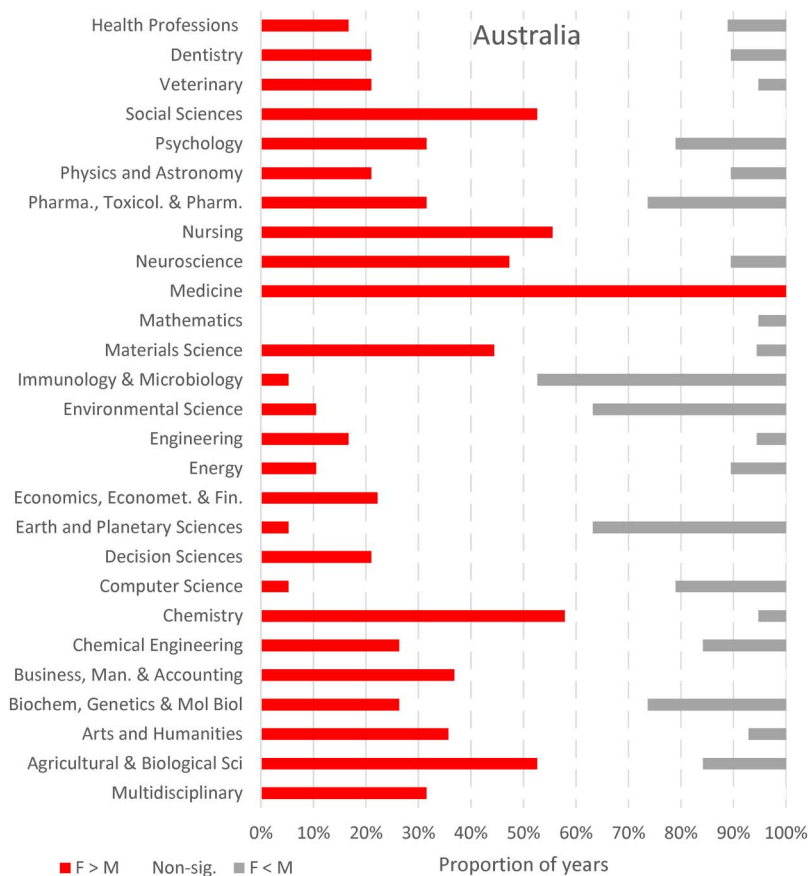


Figure 1. The number of years 1996–2014 for which the female NLCS is statistically significantly different from the male NLCS for Australia.

account that men care less (have shorter career breaks and part-time working for carer responsibilities for children and disabled and elderly dependants), have longer working lives (older retirement ages), and are demographically more prevalent in older age groups (because of the increasing proportion of female researchers over time).

Studies analyzing citations per paper do not have the same career problems, although they may be affected (probably differently by field) by gender differences in the proportions of junior and senior researchers. Citations per paper investigations have normally used statistical regression to assess whether gender helps to explain citation rates, taking a range of other variables into account (e.g., field, year, team size, team internationality, title length, abstract readability, journal impact factor). The simple comparison approach (used in the current article) instead just assesses whether there is a gender difference in average citation rates overall. This is relevant from the perspective of assessing whether any citation differentials could impact career progression (appointments, promotions). The simple approach does not explain why differences occur, which could be due to bias, different tendencies to work in teams, differences in the citation rates of subtopics studied by men and women, or, when a field is dominated by one gender, gender homophily in citation patterns (Dion, Sumner, & Mitchell, 2018; Potthoff & Zimmermann, 2017). A regression study might find that papers authored by women are more cited, but only because women work in larger teams and larger team papers tend to be more cited, irrespective of gender. A simple comparison in this case would just find an advantage for women.

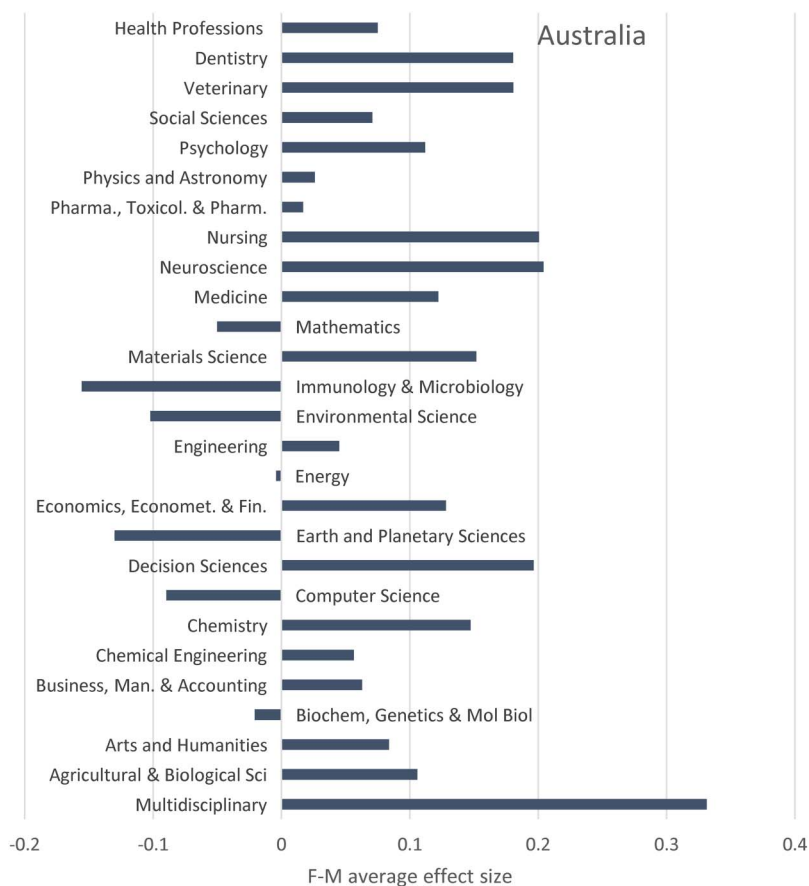


Figure 2. The female–male NLCS effect size (d) averaged over 1996–2014 for Australia.

Regressions use sets of independent variables encoding implicit assumptions about gender. For example, if a regression includes the number of title words, then it might conclude that there is no (residual) gender difference in citation rates because gender differences in title lengths “explain” overall differences. Nevertheless, this would not prove that women do not tend to create longer titles for salient reasons, such as a desire to express the impact of their paper. Similarly, if impact factors are included in a regression, then this ignores the fact that better papers may tend to be published in journals with higher impact factors due to author choices. The following list summarizes current mixed evidence about gender differences in citation rates for individual fields.

- Papers authored by men cited more: **astronomy** (astronomy papers in three major astronomy journals, *Science* and *Nature* 1950–2015; simple comparison, and regression type approach) (Caplar, Tacchella, & Birrer, 2017); **international relations** (papers in 12 journals 1980–2006; regression) (Maliniak, Powers, & Walter, 2013); **epidemiology**, mostly due to highly cited articles (papers in six journals 2018–12; simple comparison, and with h -index) (Schisterman, Swanson, et al., 2017).
- Statistically insignificant gender difference: **psychology** in Spain (Web of Science papers in 2007; average citation rate comparison) (Barrios, Villarroja, & Borrego, 2013); **ecology** (six journals 1997–2004; regression) (Borsuk, Budden, et al., 2009); **dendrochronology** (articles by 40 male or female researchers; average citation rate comparison)

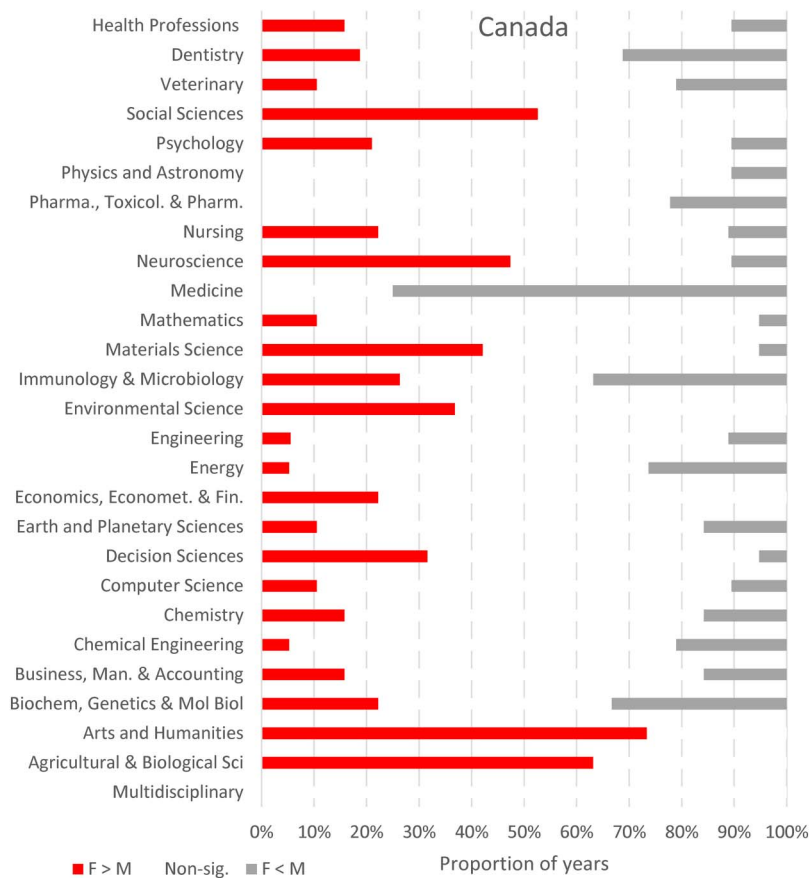


Figure 3. The number of years 1996–2014 for which the female NLCS is statistically significantly different from the male NLCS for Canada.

(Copenheaver, Goldbeck, & Cherubini, 2010); **economics, political science, and sociology** (nine journals 1983–2003; with and without regression) (Lynn, Noonan, et al., 2019); **peace research** (one journal 1983–2008; regression) (Østby, Strand, et al., 2013).

- Papers authored by women cited more: **management** in Australia, Canada, the Netherlands, the USA, and the UK, but nonsignificant for 14 other countries (46,549 Web of Science articles 2007–2013; regression) (Nielsen, 2017); **law** reviews (1980–1995 in HeinOnline; regression) (Ayres & Vars, 2000) and later confirmed (1990–2010 from the US-based HeinOnline; regression) (Cotropia & Petherbridge, 2017); Canadian **political science** (Web of Science articles 1985–2005; regression) (Montpetit, Blais, & Foucault, 2008).

To generate systematic evidence of field-based gender differences, this article assesses average gender differences in citation rates within 27 broad fields, using narrow fields for citation count normalization, for Scopus-indexed journal articles 1996–2014 from six large English-speaking countries. This uses most of the six million articles in a previous study of national gender differences (Thelwall, *in press*), excluding Jamaica and the years 2016–2018, to focus on disciplines rather than countries. This study is systematic in terms of covering all broad fields of academia and over two decades of research. New methods are used to investigate

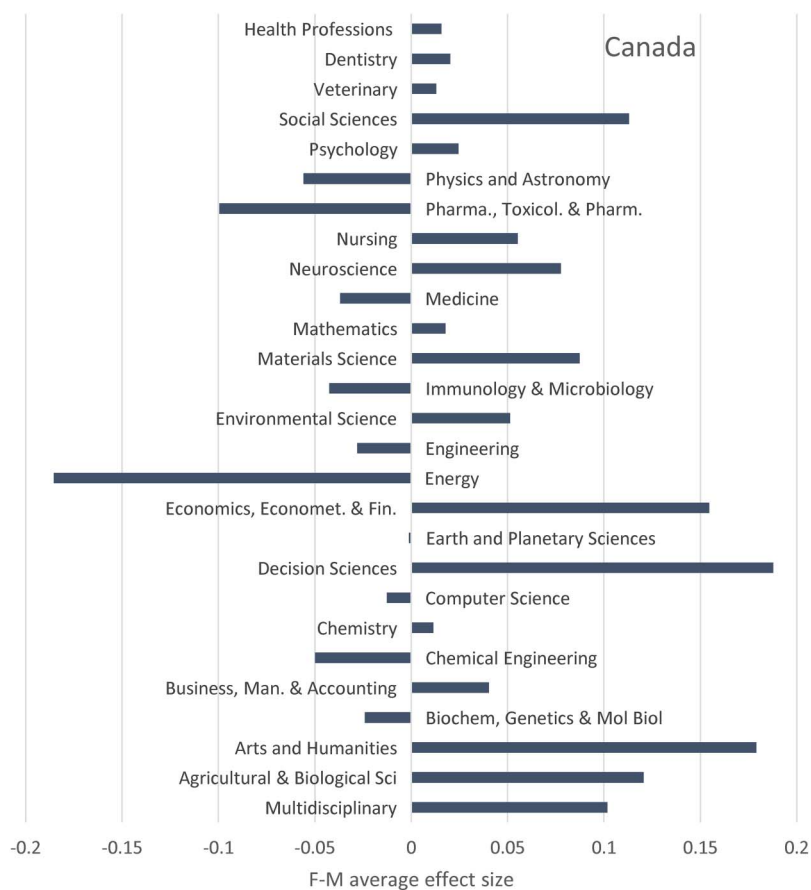


Figure 4. The female–male NLCS effect size (d) averaged over 1996–2014 for Canada.

gender differences, designed to identify the direction, statistical significance and practical significance of any differences for each field and country. The following research questions drive the study, using the simple comparison approach:

- RQ1: In which fields is there a gender difference in citation impact?
- RQ2: Does the answer to RQ1 vary between countries?
- RQ3: Does the answer to RQ1 vary over time?

2. METHODS

The research design was to recycle a large and reasonably comprehensive collection of refereed journal articles covering a long period, separate them by gender, and identify the magnitude and statistical significance of any gender differences in citation impact separately for multiple fields and countries. This is a follow-up of a previous study with the same data that did not analyze individual fields (Thelwall, *in press*).

2.1. Data

Scopus was chosen as the data source because it has a larger set of articles than the Web of Science (Mongeon & Paul-Hus, 2016). Only standard (nonreview) journal articles were

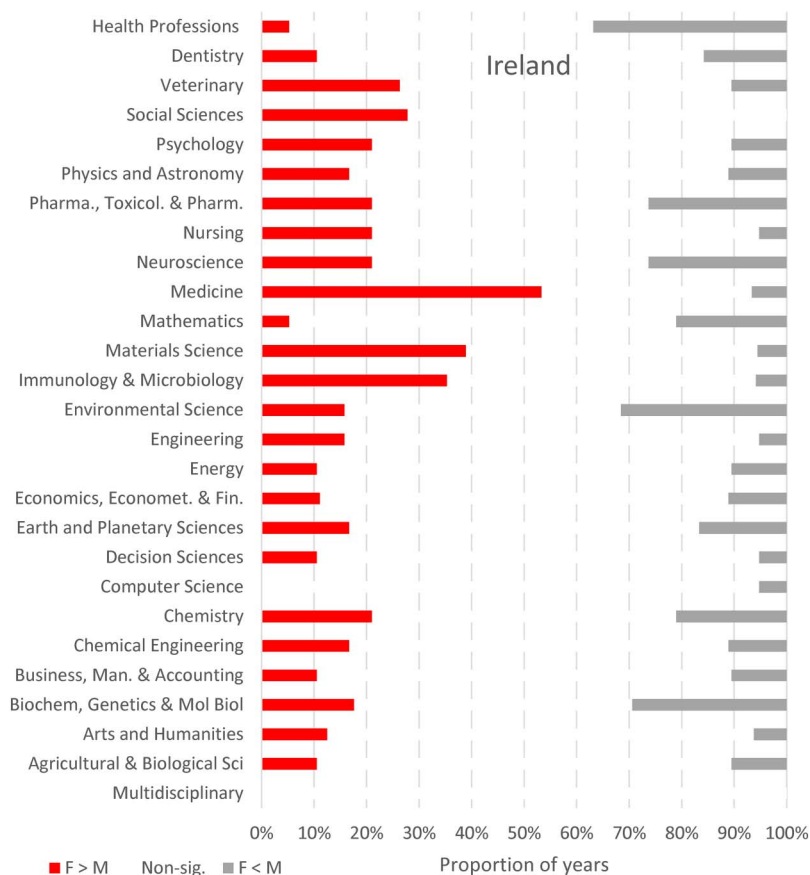


Figure 5. The number of years 1996–2014 for which the female NLCS is statistically significantly different from the male NLCS for Ireland.

included, because these are the primary vehicles to convey research findings in most fields. Conference papers, monographs, book chapters, and other outputs were ignored because there are not indexed systematically enough to effectively normalize their citation impact. These document types are important for the arts, humanities, some social sciences, computing, some engineering specialisms, and computational linguistics, so the article-based results for these fields here are not robust. This primarily applies to the following Scopus fields: Arts & Humanities; Computer Science; Energy; Engineering; and Social Sciences.

The period 1996–2014 was covered, as downloaded in November–December 2018 from Scopus. The start year 1996 is the logical choice, as the first after a Scopus coverage expansion. The end year 2014 gives every article at least 3 years to attract citations, which is sometimes considered the minimum window size for citation analysis (Abramo, Cicero, & D’Angelo, 2011; Wang, 2013). However, for each article, citations to 2018 were used rather than a fixed time window, as this would give more comprehensive data.

The six countries analyzed were Australia, Canada, Ireland, New Zealand, the UK, and the USA. These are all large predominantly English-speaking countries with a partially shared culture and similar levels of economic development. They represent a set for which there is no obvious reason why their results should differ and so form an interesting case in the sense that any differences must have more subtle causes. The largest excluded country, Jamaica, did not publish enough to allow broad field differences to be detected systematically. The focus on

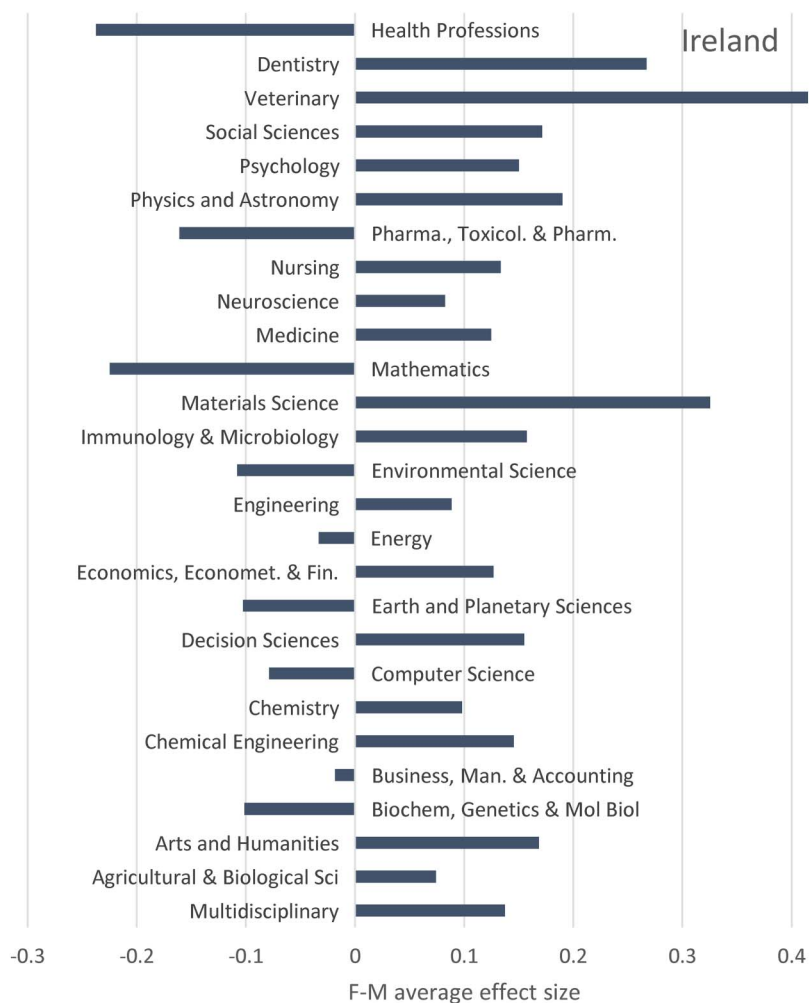


Figure 6. The female–male NLCS effect size (d) averaged over 1996–2014 for Ireland.

large countries is essential for the possibility of statistically significant results at the field level.

Articles were assigned a gender by using a first name heuristic on the first author. The first author tends to contribute the most in all broad fields (Larivière, Desrochers, et al., 2016) even though there is a degree of alphabetical authorship in some (Waltman, 2012) and the last author may be a senior author who determined the overall direction in others (Mongeon, Smith, et al., 2017), and for some PhD projects. Nevertheless, except in the minority of alphabetical (or partial alphabetical) ordering cases, the first author is the only one who can be reliably assumed to have made a major contribution to the published study. In the absence of systematic cross-science author contribution statements, this seems like a better choice than assuming that all authors contribute equally or weighting the first and last authors. The effect of this decision is to increase noise in the data (extra variability) for multiauthor articles when authors with a gender differing from the first author made a substantial contribution. It may also introduce systematic biases in fields where, for example, senior men tend to make substantial contributions, but the first author is a junior woman who conducted most of the research.

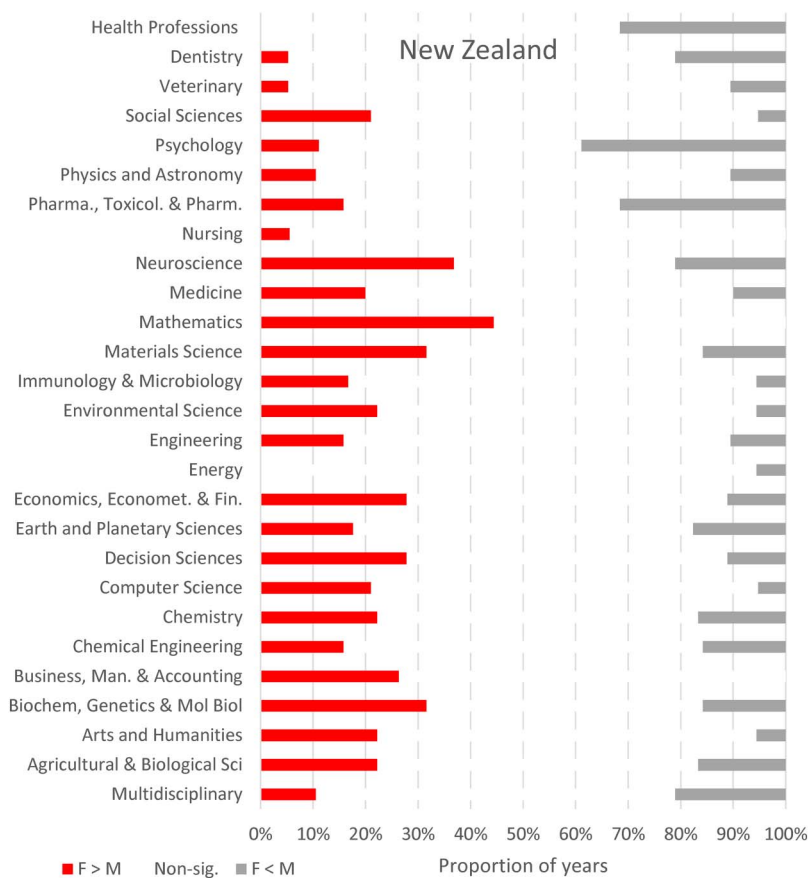


Figure 7. The number of years 1996–2014 for which the female NLCS is statistically significantly different from the male NLCS for New Zealand.

The first name list used was based on the US census 2010, including frequently occurring names that were used at least 90% by one gender. This was augmented by checking the most common other first names (as extracted from the Scopus data) with GenderAPI.com (Santamaría & Mihaljević, 2018), retaining first names as gendered when they were at least 90% of one gender and occurred reasonably frequently (e.g., 10 times if 100% for one gender, increasing to 500 times if 90% for one gender). This heuristic is imperfect because academics may have gender-neutral names, Chinese gendered names that are gender neutral in the Latin alphabet (e.g., Wei), gender-neutral short names (e.g., Pat), or names that are a different gender in another country (e.g., Nicola). A manual check of this method with the US census names only, using personal home page gender identifications for 1,010 US academics, found it to be 96.5% accurate on US first authors (Thelwall et al., 2019b), and a check on the same academics with the updated method used in the current paper found it to be 96.8% accurate. Example names with incorrectly guessed genders include Ronni, Yen, Shae, and Juan. A check of a similar method with 95% monogender names from the US census found that it agreed with Genni 2.0 (Smith, Singh, & Torvik, 2013; Torvik, 2018; Torvik & Agarwal, 2016) gender estimates (which include last name information to infer ethnicity) at a rate above 96% for various ethnicities for international PubMed authors (Mishra et al., 2018). Thus, the papers should be correctly identified for gender nearly all of the time for the USA, although the accuracy may be slightly lower for the other six countries despite

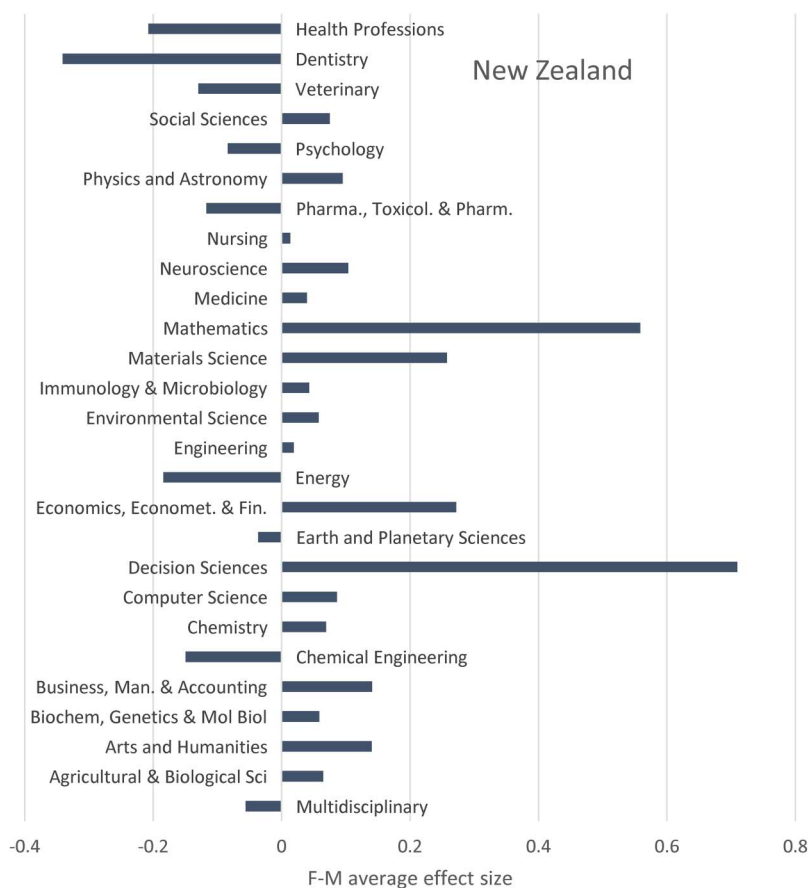


Figure 8. The female–male NLCS effect size (d) averaged over 1996–2014 for New Zealand.

the partly shared cultural heritage. Sample sizes for all field/year/country/gender combinations can be found within the spreadsheets holding the graphs in the online supplement (<https://doi.org/10.6084/m9.figshare.8081546>).

Citation counts were log transformed to reduce skewing and outliers, then normalized for the field and year of publication to allow cross-field, multiple-year comparisons. The log transformation was $\ln(1 + c)$, where 1 is added because many articles are uncited. The field normalization is to divide each article $\ln(1 + c)$ with the average of all the $\ln(1 + c)$ values for all articles published in the same Scopus narrow field (336 in total: Elsevier [2019]). For articles given multiple narrow field classifications, the average used in the denominator is the average of all relevant field averages. This is referred to as the *normalized log-transformed citation score* (NLCS). For any set of articles, the arithmetic mean of their NLCS is the *mean normalized log-transformed citation score* (MNLCS) indicator (Thelwall, 2017) variant of the *mean normalized citation score* (MNCS) (Waltman, van Eck, et al., 2011). The raw data is therefore a large set of NLCS, each associated with a country, gender, and Scopus broad field.

Scopus field classifications were used even though they are probably not as effective for citation analysis as the field definitions of the Web of Science or Science-Metrix (Klavans & Boyack, 2017) because classifications were needed for all articles in the set to maximize statistical power. The Science-Metrix scheme is relatively recent (Archambault, Beauchesne, & Caruso, 2011) and may not cover older journals.

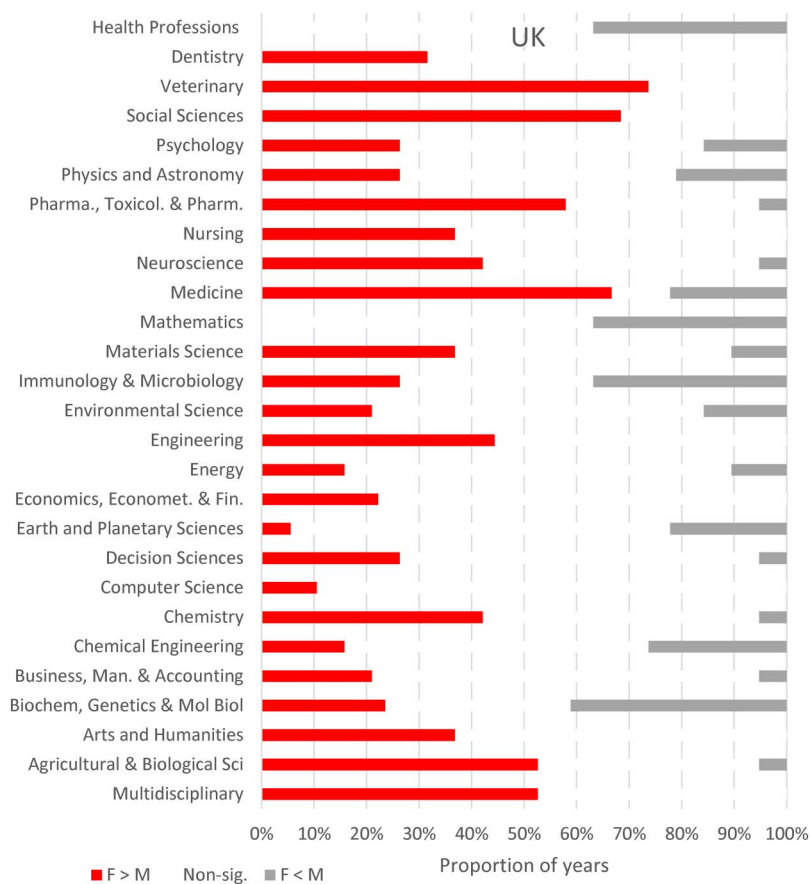


Figure 9. The number of years 1996–2014 for which the female NLCS is statistically significantly different from the male NLCS for the UK.

2.2. Analyses

Gender differences may vary by country, time, and discipline, generating a three-way analysis. The time effect could be nonlinear and the national MNLCS is impacted by changes in the journal coverage of Scopus, generating sudden national increases or decreases. A regression-type model including time is therefore not appropriate. Instead, two separate analyses are reported for each country: gender differences in disciplines, assuming no gender difference variability over time; and gender differences over time, assuming no gender difference variability between disciplines.

For each country/field/year combination, a *t*-test was conducted to compare male NLCS with female NLCS. The effect size was calculated by the difference in the means divided by the pooled standard deviation. All year/field/country combinations with fewer than two articles or excess kurtosis >3 or skew >3 were excluded from all calculations. After this exclusion, the *t*-test is a reasonable choice. Sample sizes, standard deviations, skewness, and individual *t*-test values are available in the online supplement at <https://doi.org/10.6084/m9.figshare.8081546>. Average effect sizes across all years and the number of positive tests at the *p* = 0.05 level are reported. Graphs were produced for each country/field combination for the average male and female NLCS to assess interactions between fields, years, and countries over time. The standard normal distribution formula with the *t*-distribution was used for confidence intervals on these graphs.

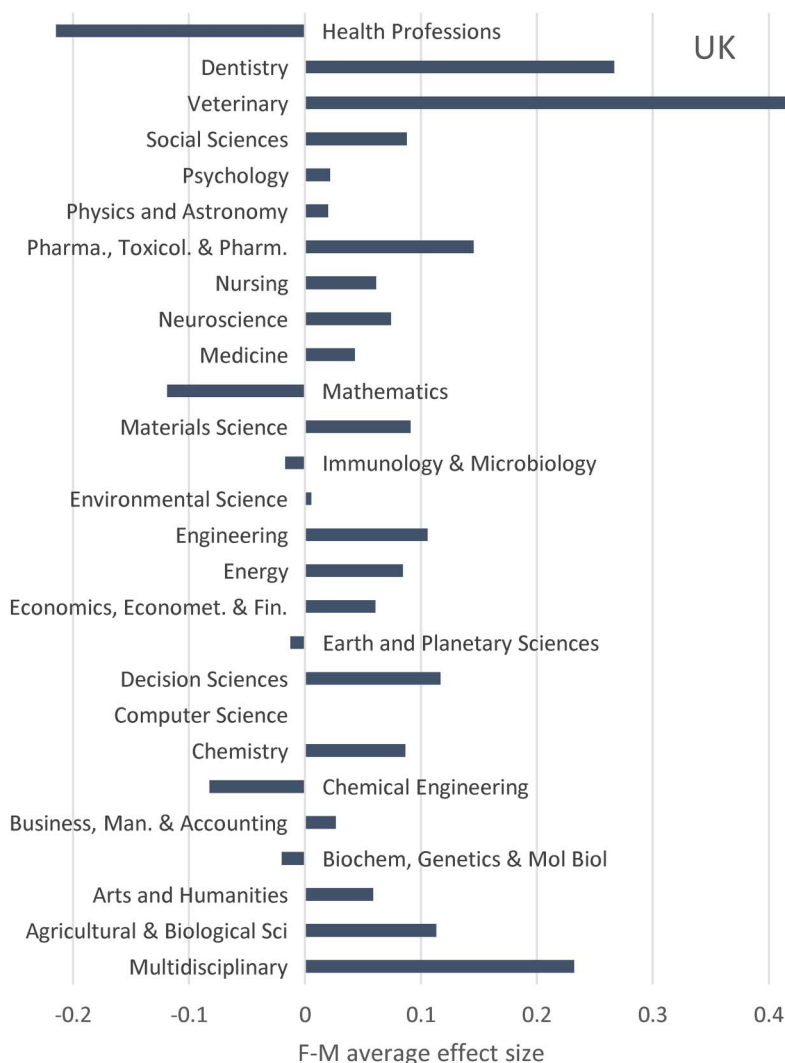


Figure 10. The female–male NLCS effect size (d) averaged over 1996–2014 for the UK.

3. RESULTS

3.1. RQ1 and RQ2

In all countries, there is a statistically significant gender citation advantage for women for multiple fields and years and a statistically significant gender citation advantage for men for multiple fields and years, with overall effect sizes being small (Figures 1–12). Overall, it is more common for there to be a statistically significant advantage for women than for men, and for the average effect size to be in favor of women (Figures 1–12).

It is rare for field citation advantages to be dominated by one gender in a country. The main exceptions are Medicine in Australia, Pharmacology, Toxicology, and Pharmaceuticals in the USA, and Arts and Humanities in the USA, all of which have a female citation advantage in at least 90% of the years studied. Two of these are not discipline-wide patterns because Medicine and Pharmacology, Toxicology, and Pharmaceuticals have mainly male citation advantages in Canada. In contrast, the Arts and Humanities have a female citation advantage

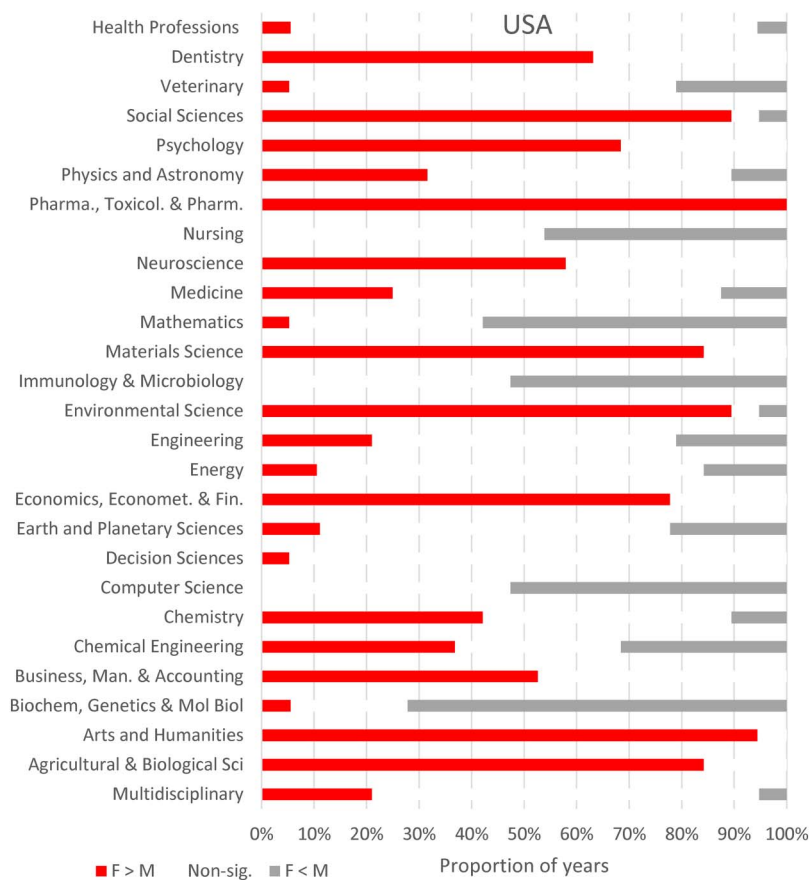


Figure 11. The number of years 1996–2014 for which the female NLCS is statistically significantly different from the male NLCS for the USA.

overall (average effect size) and a higher proportion of years with a statistically significant female citation advantage in all countries. Canadian medicine is the only case where men have a citation advantage for at least 50% of the years. The Social Sciences also have an international female citation advantage: All countries have more years with a female citation advantage than a male citation advantage, and some countries have a substantial difference. There does not seem to be a relationship between gender proportions in a field and citation advantages, however. For example, female-dominated nursing and male-dominated mathematics both have international variations in which gender tends to be cited statistically significantly most often.

3.2. RQ3

There is no trend for gender citation differentials to change over time. No country showed a pattern of gradually increasing or decreasing gendered shares of significant results or effect sizes. The possible slight exception is the USA, where there was a consistent decrease in the overall female advantage average effect size from 2010 to 2014 and a broadly decreasing trend from 2004 to 2014 (Figure 13; corresponding graphs for the other countries are available in the online supplement: <https://doi.org/10.6084/m9.figshare.8081546>).

Graphs for male and female average citation rates for all country and year combinations are available in the online supplement (6 × 27 = 162 graphs). From a large number of graphs, some are likely to show trends by accident, so these are difficult to draw strong conclusions

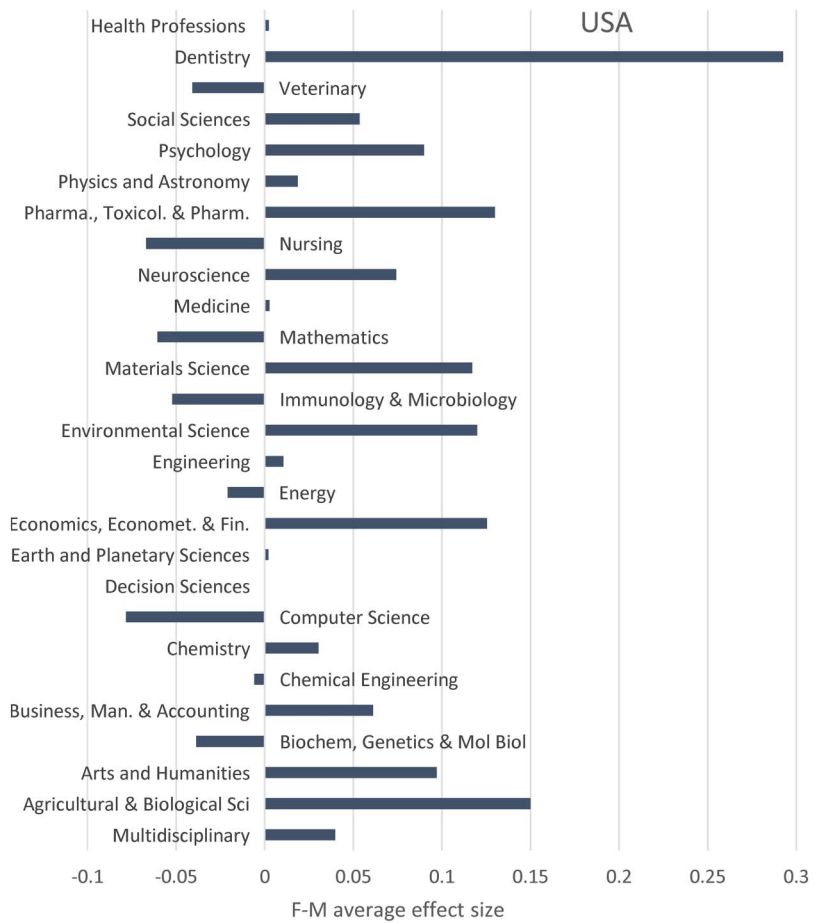


Figure 12. The female–male NLCS effect size (d) averaged over 1996–2014 for the USA.

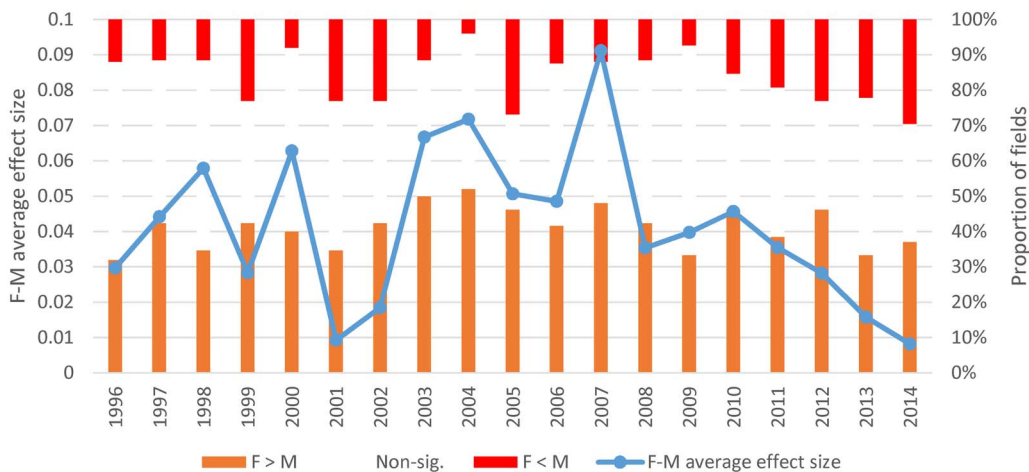


Figure 13. The female–male NLCS effect size (d) averaged over all 27 broad fields for the USA, and the number of broad fields (out of 27) for which the female NLCS is statistically significantly different from the male NLCS for the USA. Similar graphs for the other countries are available in the online supplement: <https://doi.org/10.6084/m9.figshare.8081546>.

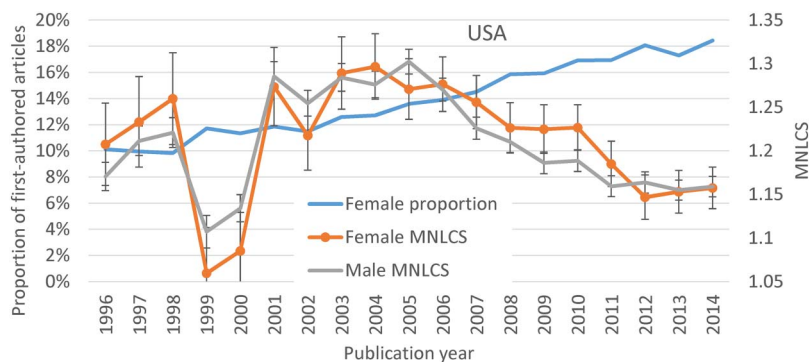


Figure 14. Female and male MNLCS 1996–2014 for Engineering in the USA. Error bars show 95% confidence intervals. Similar graphs for all other countries and disciplines are available online (<https://doi.org/10.6084/m9.figshare.8081546>).

from. Figure 14 (Engineering in the USA) shows periods with male (1999–2002) and female (1996–98; 2006–11) dominance, explaining the partly male and partly female significant results for a single subject and country. This graph also illustrates the impact of changes in the coverage of Scopus (the 1999–2000 dip) on the MNLCS magnitude.

4. DISCUSSION

Although this seems to be the largest systematic analysis of gender differences in citation rates yet, it has several limitations. The inclusion within Scopus categories of unusual or inappropriate journals with atypical authorship ratios for the category could influence each MNLCS. In addition, Scopus narrow categories (used for MNLCS) may combine specialisms with differing gender compositions and citation rates (e.g., librarianship and scientometrics). The statistical significance tests assume that the citation counts for each gender are independent of each other, which is false because one researcher may publish a set of high (or low) cited articles in the same year because they are working on a high (or low) citation topic. The use of multiple tests without a familywise correction procedure (e.g., Bonferroni) means that each individual test is unsafe; only the patterns are important. The gender identification heuristic may systematically exclude people with unusual gender characteristics for a country (e.g., with gender-neutral Sikh names in the UK). Each country's academics may also have research characteristics learned from another nation that educated them, and gender may influence this. For example, one gender may be more willing to travel to or from a country for a PhD position or research post. The analysis method assumes that the first author is the main author and that the other researchers had a minor contribution, which is not always true. In some fields, the last author is senior and may largely determine the research topic and methods. The results should not be extrapolated beyond the six English-speaking countries covered because of international differences in the relationship between gender and citation rates. This is evident in the results above, despite the relatively homogeneous set of countries. The results do not consider career stages, so it is possible that gender differences in citation rates differ between younger and older researchers. Finally, the results do not consider factors other than gender, field, and year that may influence citation rates, such as team size (Larivière, Gingras, et al., 2015) and researcher seniority (Slyder, Stein, et al., 2011). Thus, when there is a female citation advantage, it is not possible to infer that a paper authored by a woman is likely to be more cited than a paper by a comparable (e.g., PhD student, postdoc, junior or senior faculty) man.

The overall female citation advantage suggests that when citations are used in career progression decisions in a fair way (i.e., on a per-paper basis or considering career gaps and periods of part-time working), then women would have an advantage more often than men. Nevertheless, it is not clear that decision makers would often have detailed enough citation evidence (either field-normalized scores, or making comparisons only between papers from the same narrow field and year) to make an appropriately informed decision. Moreover, the differences in citation rates are not large. An effect size of 0.2 is “small” (Cohen, 1988) and average effect sizes are typically around 0.1. Given papers authored by a man and woman chosen at random, an effect size of 0.1 translates to the probability of the woman’s paper being more cited (ignoring ties) being 0.53. For an effect size of 0.2 the corresponding probability is 0.56 (Coe, 2002). Thus, it seems that the gender difference is typically not relevant. The largest average effect size of 0.7 gives a corresponding probability of 0.67. Thus, for two-thirds of gendered pairs of New Zealand Decision Sciences papers, the women’s papers would be more cited (using NLCS). This is a more substantial difference, but less than 30% of years have a statistically significant female advantage because the large effect size is based on low numbers of papers. Overall, then, it seems unlikely that the tendency for a female citation advantage has translated into a female career advantage often enough to make a difference to female academic career prospects overall.

Previous studies have found female (Thelwall, 2018a), male (Larivière et al., 2013), or non-significant (Thelwall, *in press*) citation advantages for the USA overall. The above results suggest an overall female citation advantage, varying by field, but with a mostly tiny average effect size: Only Dentistry has an effect size above 0.2. None of the field-based studies reviewed in the introduction are directly comparable in terms of field and country coverage, except that the regression-based female citation advantage for Management in Australia, Canada, the UK, and the USA (Nielsen, 2017) broadly agrees with the Business, Management, and Accounting category results above.

The two broad fields with a reasonably consistent gendered citation advantage, Social Sciences and Arts and Humanities, are also probably the most diverse in terms of the narrow fields subsumed within them (e.g., from Law to Cultural Studies and from History to Visual and Performing Arts). These may also be the broad fields in which citations are least valued, partly due to the lower importance of journal articles (e.g., REF2014, 2015). They may also be the fields in which the variety within narrow fields is greatest. Internal narrow field diversity would affect the MNLCS, so the results are not strong evidence for a generic female citation advantage in these two broad areas.

The results do not support prior claims of a citation bias against research by women because there is no broad field in which female first-authored research is not statistically significantly more cited for some years. Nevertheless, the frequent (small) female citation advantage is more impressive given the greater male self-citation tendency (King, Bergstrom, et al., 2017; even though it is a second-order effect: Mishra et al., 2018), gender citation homophily (given that most researchers are male), and a higher proportion of senior men (although seniority does not necessarily associate with higher citation impact). It is still possible that there is a degree of deliberate or unconscious gender bias in citing that has reduced the magnitude of the female citation advantage in some or all fields, years, and countries.

It is not possible to infer a cause-and-effect relationship in the data that would explain the overall tendency for female first-authored research to be statistically significantly more cited than male first-authored research in more fields or years than the other way around. This is not just because of the lack of a regression approach with an adequate selection of independent

variables. For example, if women tended to operate in larger teams and team size was the apparent “cause” of higher citation rates for women, an explanation would be needed for women being more likely to find themselves in larger teams, leaving open the possibility that gender is still important. Nevertheless, it is *possible* that there is a small underlying tendency for women to generate more impactful research as a side-effect of being more likely to have communal career goals (Diekman et al., 2017).

5. CONCLUSIONS

The results show a general, but small, tendency for female first-authored standard journal articles to be more cited than male first-authored standard journal articles in all 27 broad Scopus fields 1996–2014 in Australia, Canada, Ireland, New Zealand, the UK, and the USA. This tendency varies between fields and years in a way that does not appear to have a systematic pattern, except for a reasonably consistent tendency for a female citation advantage in Social Sciences and Arts and Humanities in all countries and multiple years (more than the other way around). However, the average effect size of the difference is almost certainly too small to have much influence on academic appointments, promotions, and tenure, in any field or country (from the six considered) over the past two decades. This is particularly true because promotion committees will not have access to the fine-grained field-normalized data here, which, due to their skew-resistant formulae, are more precise than available in the Web of Science and Scopus. The online graphs associated with this article (<https://doi.org/10.6084/m9.figshare.8081546>) give the most comprehensive data yet on gender differences and effect sizes for journal articles across academia, at least for the six large countries covered.

Based on these results, prior findings of a per-paper citation advantage for men and women for individual subjects and/or countries and/or small year ranges should not be extrapolated, because the results here show nonsystematic variations in these factors over time, between nations, and between fields.

COMPETING INTERESTS

The author has no competing interests to declare.

FUNDING INFORMATION

This research received no funding.

DATA AVAILABILITY

The data behind the results are on FigShare (<https://doi.org/10.6084/m9.figshare.8081546>).

REFERENCES

- Abramo, G., Cicero, T., & D’Angelo, C. A. (2011). Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics*, 5(4), 659–667.
- Andersen, J. P., Schneider, J. W., Jagsi, R., & Nielsen, M. W. (2019). Gender variations in citation distributions in medicine are very small and due to self-citation and journal prestige. *Elife*, 8, e45374. <https://doi.org/10.7554/eLife.45374>
- Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics* (pp. 66–77). South Africa: Durban.
- Ayres, I., & Vars, F. E. (2000). Determinants of citations to articles in elite law reviews. *The Journal of Legal Studies*, 29(S1), 427–450.
- Barrios, M., Villarroya, A., & Borrego, Á. (2013). Scientific production in psychology: A gender analysis. *Scientometrics*, 95(1), 15–23.
- Borsuk, R. M., Budden, A. E., Leimu, R., Aarssen, L. W., & Lortie, C. J. (2009). The influence of author gender, national language and number of authors on citation rate in ecology. *Open Ecology Journal*, 2(1), 25–28.
- Caplar, N., Tacchella, S., & Birrer, S. (2017). Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nature Astronomy*, 1(6), 0141.

- Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, 108(8), 3157–3162.
- Coe, R. (2002). It's the effect size, stupid. Paper presented at the British Educational Research Association Annual Conference. <https://www.leeds.ac.uk/educol/documents/00002182.htm>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Oxford: Routledge.
- Copenheaver, C. A., Goldbeck, K., & Cherubini, P. (2010). Lack of gender bias in citation rates of publications by dendrochronologists: What is unique about this discipline? *Tree-Ring Research*, 66(2), 127–134.
- Cotropia, C. A., & Petherbridge, L. (2017). Gender disparity in law review citation rates. *William & Mary Law Review*, 59, 771.
- Diekman, A. B., Steinberg, M., Brown, E. R., Belanger, A. L., & Clark, E. K. (2017). A goal congruity model of role entry, engagement, and exit: Understanding communal goal processes in STEM gender gaps. *Personality and Social Psychology Review*, 21(2), 142–175.
- Dion, M. L., Sumner, J. L., & Mitchell, S. M. (2018). Gendered citation patterns across political science and social science methodology fields. *Political Analysis*, 26(3), 312–327.
- Elsevier. (2017). *Gender in the Global Research Landscape*. <https://www.elsevier.com/research-intelligence/campaigns/gender-17>
- Elsevier. (2019). *Content—How Scopus Works* [Download source title list link; ASJC worksheet]. <https://www.elsevier.com/solutions/scopus/how-scopus-works/content>
- Handley, I. M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences*, 112(43), 13201–13206.
- Hines, M. (2011). Gender development and the human brain. *Annual Review of Neuroscience*, 34, 69–88.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581–592.
- King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2017). Men set their own cites high: Gender and self-citation across fields and over time. *Socius*, 3, 2378023117738903.
- Klavans, R., & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, 68(4), 984–998.
- Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, 504(7479), 211–213.
- Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, 46(3), 417–435.
- Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7), 1323–1332.
- Lynn, F. B., Noonan, M. C., Sauder, M., & Andersson, M. A. (2019). A rare case of gender parity in academia. *Social Forces*, 126, 126–142.
- Maliniak, D., Powers, R., & Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, 67(4), 889–922.
- Mishra, S., Fegley, B. D., Diesner, J., & Torvik, V. I. (2018). Self-citation is the hallmark of productive authors, of any gender. *PLOS ONE*, 13(9), e0195773.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, 106(1), 213–228.
- Mongeon, P., Smith, E., Joyal, B., & Larivière, V. (2017). The rise of the middle author: Investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. *PLOS ONE*, 12(9), e0184601.
- Montpetit, É., Blais, A., & Foucault, M. (2008). What does it take for a Canadian political scientist to be cited? *Social Science Quarterly*, 89(3), 802–816.
- Moss-Racusin, C. A., Molenda, A. K., & Cramer, C. R. (2015). Can evidence impact attitudes? Public reactions to evidence of gender bias in STEM fields. *Psychology of Women Quarterly*, 39(2), 194–209.
- Nielsen, M. W. (2017). Gender and citation impact in management research. *Journal of Informetrics*, 11(4), 1213–1228.
- Østby, G., Strand, H., Nordås, R., & Gleditsch, N. P. (2013). Gender gap or gender bias in peace research? Publication patterns and citation rates for *Journal of Peace Research*, 1983–2008. *International Studies Perspectives*, 14(4), 493–506.
- Potthoff, M., & Zimmermann, F. (2017). Is there a gender-based fragmentation of communication science? An investigation of the reasons for the apparent gender homophily in citations. *Scientometrics*, 112(2), 1047–1063.
- Reed, D. A., Enders, F., Lindor, R., McClees, M., & Lindor, K. D. (2011). Gender differences in academic productivity and leadership appointments of physicians throughout academic careers. *Academic Medicine*, 86(1), 43–47.
- REF2014. (2015). *Research Excellence Framework 2014: Panel Overview Reports*. <https://www.ref.ac.uk/2014/panels/paneloverviewreports/>
- Santamaría, L., & Mihaljević, H. (2018). Comparison and benchmark of name-to-gender inference services. *PeerJ Computer Science*, 4, e156.
- Schisterman, E. F., Swanson, C. W., Lu, Y. L., & Mumford, S. L. (2017). The changing face of epidemiology: Gender disparities in citations? *Epidemiology*, 28(2), 159–168.
- Simon, R. M., Wagner, A., & Killion, B. (2017). Gender and choosing a STEM major in college: Femininity, masculinity, chilly climate, and occupational values. *Journal of Research in Science Teaching*, 54(3), 299–323.
- Slyder, J. B., Stein, B. R., Sams, B. S., Walker, D. M., Jacob Beale, B., Feldhaus, J. J., & Copenheaver, C. A. (2011). Citation pattern and lifespan: A comparison of discipline, institution, and individual. *Scientometrics*, 89(3), 955–966.
- Smith, B., Singh, M., & Torvik, V. I. (2013). A search engine approach to estimating temporal changes in gender orientation of first names. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL 2013)*. New York, NY: ACM Press (pp. 199–208). <https://doi.org/10.1145/2467696.2467720>
- Solera, C., & Musumeci, R. (2017). The persisting glass ceiling in academia: A multidimensional cross-national perspective. *Polis*, 31(1), 17–44.
- Thelwall, M. (2017). Three practical field normalised alternative indicator formulae for research evaluation. *Journal of Informetrics*, 11(1), 128–151.
- Thelwall, M. (2018a). Do females create higher impact research? Scopus citations and Mendeley readers for articles from five countries. *Journal of Informetrics*, 12(4), 1031–1041.
- Thelwall, M. (2018b). Does female-authored research have more educational impact than male-authored research? *Journal of Altmetrics*, 1(1), p. 3.

- Thelwall, M. (in press). Female citation impact superiority 1996–2018 in six out of seven English-speaking nations. *Journal of the Association for Information Science and Technology*. <https://arxiv.org/abs/1904.12919>
- Thelwall, M., Bailey, C., Makita, M., Sud, P. & Madalli, D. (2019a). Gender and Research Publishing in India: Uniformly high inequality? *Journal of Informetrics*, *13*(1), 118–131.
- Thelwall, M., Bailey, C., Tobin, C. & Bradshaw, N. (2019b). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics*, *13*(1), 149–169.
- Torvik, V. I. (2018). Genni + Ethnea for the Author-ity 2009 dataset. University of Illinois at Urbana-Champaign. https://doi.org/10.13012/B2IDB-9087546_V1
- Torvik, V. I., & Agarwal, S. (2016). Ethnea: An instance-based ethnicity classifier based on geocoded author names in a large-scale bibliographic database. International Symposium on Science of Science, March 22–23, 2016. Library of Congress, Washington, DC, USA. <https://www.ideals.illinois.edu/bitstream/handle/2142/88927/ethnea.pdf>
- Van Miegroet, H., Glass, C., Callister, R. R., & Sullivan, K. (2019). Unclogging the pipeline: Advancement to full professor in academic STEM. *Equality, Diversity and Inclusion: An International Journal*, *38*(2), 246–264.
- Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics*, *6*(4), 700–711.
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. (2011). Towards a new crown indicator: An empirical analysis. *Scientometrics*, *87*(3), 467–481.
- Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, *107*(2), 468.
- Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872.