Authorship and citation gender trends in immunology and microbiology[1]

Mike Thelwall, University of Wolverhampton, UK

Immunology and microbiology research are essential for human and animal health. Unlike many other health fields, they do not usually centre around the curing or helping individual patients but focus on the microscopic scale instead. These fields are interesting from a gender perspective because two theories seeking to explain gender differences in career choices in the USA (people/things and communal/agentic goals) might produce conflicting expectations about their gender balances. This article assesses the gender shares of journal articles and gendered citation rates of five subfields of the Scopus Immunology and Microbiology broad category 1996-2014/18, for research with solely US author affiliations. Only Applied Microbiology and Biotechnology (38% female) had not reached gender parity in publishing by 2018. There was a female first author citation advantage in Parasitology but a disadvantage in Immunology. Immunology, Parasitology and Virology, had female last author citation disadvantages, but all gender effects were much smaller (<5%) than that of an extra author (10%-56%). Citation differences cannot therefore account for the current underrepresentation of women in senior roles.

Introduction

There is still a large gender gap in academia that may take over a century to eradicate in some disciplines (Holman, Stuart-Fox and Hauser 2018). For example, a female in the USA is 100 times more likely to publish in one area of academic nursing than in mathematical logic (Thelwall, Bailey, Tobin, and Bradshaw 2019). Biological sciences have attracted more females than other Science, Technology, Engineering and Mathematics (STEM) areas (e.g., NCES, 2018) but reasons for this relative success are not well understood (Sax, Lim, Lehman, and Lonje-Paulson 2018) and rates of females in senior positions have caused concern. For example, a 2016 survey of immunology departments and graduate programmes found women to form half (50.5%) of graduate students but only 22.3% of professors (Shapiro, Kovats, Parent, Gaffen, Hedrick, Jain, and Stephens 2016). Analyses of science, technology, engineering and maths subjects suggest that personal choice is more important than explicit gender bias for the gender composition of a field (Ceci & Williams, 2011).

Two competing psychological theories attempt to explain the gendering of careers, including within academia. According to one theory, as fields that clearly and directly benefit society, immunology and microbiology should have no problems attracting female researchers because females are more likely than males to have communal career goals that involve working to help society, whereas males are more likely to have self-focused agentic goals (Diekman, Steinberg, Brown, Belanger, and Clark 2017). In contrast, the people/thing theory (Su, Rounds, & Armstrong, 2009) suggests that males prefer careers dealing with inanimate objects, whereas females prefer careers working with people. Neither people nor things directly apply to immunology and microbiology. Both have some relevance, however, because the research must involve equipment (things) to operate at a microscopic level and much immunology and microbiology occurs, in principle, within people. Thus, the second theory does not make a strong prediction. This article analyses gender participation in immunology and microbiology research to shed light on this issue.

[1] Thelwall, M. (in press). Authorship and citation gender trends in immunology and microbiology. FEMS Microbiology Letters

Previous general bibliometric studies have given some information about gender in immunology and microbiology research. An investigation of 26 broad fields in the USA in 2017 found Immunology and Microbiology to be the seventh most female, but with 0.75 female first-authored articles for every male first-authored article. The most female specialty was Virology (0.86) and the least was Applied Microbiology & Biotechnology (0.54) (Thelwall, Bailey, Tobin, and Bradshaw 2019). A parallel study of India found Immunology and Microbiology to be the third most female broad field in the country, but still with a male majority of first authors (Thelwall, Bailey, Makita, Sud, and Madalli 2019). Only one article has analysed author gender within a purely immunology and microbiology contest. A bibliometric analysis of 5906 rotavirus articles 1900-2013 in the Web of Science, found that male authors formed a majority in the USA and all countries publishing at least 100 articles, except Brazil (Köster, Klingelhöfer, Groneberg, and Schwarzer 2016). Many other papers have used bibliometric methods to investigate the field, but typically to identify research topics, international productivity or highly cited papers (Moral-Munoz, Lucena-Antón, Perez-Cabezas, Carmona-Barrientos, González-Medina, and Ruiz-Molinero 2018; Sweileh, Al-Jabi, Sa'ed, Sawalha, and Abu-Taha 2018).

Despite the above review, no previous research seems to have focused on immunology and microbiology or to have investigated its gender evolution over time in terms of publishing and citation. This article assesses journal article authors 1996-2018 in the USA at the subfield level to give more detailed information about gender trends in the area. A citation analysis 1996-2014 is also included to investigate the main research question: *are citation rates gendered by first or last author in any or all subfields?* A positive result may help to explain the lack of senior women in the field. The focus is on the first and last authors, since these are typically the main contributors to an article. The number of authors is also investigated, since team size can influence citation rates, giving a secondary research question: *are citation rates gendered by first or last author in any or all subfields, after taking into account team size?* A restriction to a single country helps to ensure that no second order factors influence the results due to international differences in proportions of females and average citation impact.

Materials and Methods

Scopus journal articles from all seven Scopus Immunology and Microbiology categories from 1996 (the start date of greater Scopus indexing) to 2018 were downloaded between November 2018 and the end of January 2019, in chronological order. Only records for documents of type journal article were downloaded, excluding reviews, letters, editorials and other outputs. Author national affiliations were extracted from Scopus and first/last author genders were estimated from their first name, when recorded in Scopus, by matching it to a list of names associated with one gender at least 90% of the time in their country. These lists were obtained from the US 1990 census and gender-api.com. A test against a set of 1010 academics manually classified for gender from their web presences taken from a previous paper (Thelwall, et al., 2019a) suggest that this method is 98.5% accurate but, because it does not gender some names, overestimates the proportion of females by 1%. The number of authors for each article was also extracted from the Scopus record.

Scopus citation counts were extracted, log transformed and field/time normalised to allow comparisons between fields and years. A log transformation $ln(1+c)$ is necessary for analysing citations because sets of citation counts are highly skewed and averaging non-transformed citation counts gives results dominated by a few highly cited articles. Year normalisation is necessary for comparisons over time because older articles have had longer to be cited. Field normalisation is also necessary because fields generate citations at different rates due to journal reference list length policies, field norms in citing, and different needs to

cite journal articles within current papers. Articles after 2014 were excluded from the citation analysis, giving each article at least three years for citations to accumulate, which is adequate for comparison purposes (Campanario 2011).

The log-transformed citation counts were field/time normalised by dividing them by the average log-transformed citation count $ln(1+c)$ for all articles published in the same Scopus narrow field and year. This gives a score that is the ratio to the world average for the field and year. It is above 1 only if the article is more cited than average for its field and year. Most articles were recorded against multiple fields, and in these cases the average of the field average log-transformed citation counts was used as the denominator. This is the Normalised Log-transformed Citation Score (NLCS). Averaging these across multiple articles (e.g., all articles by males written in 1999 or overall) gives the Mean Normalised Log-transformed Citation Score (MNLCS) (Thelwall 2017). Because the normalisation transforms all citation counts to a common ratio scale where 0 is uncited, and 1 is the world average, it is fair to compare MNLCS values between different sets of articles even if they contain different balances of fields and years.

Analysis

MNLCS were plotted against time for first author gender in each Scopus subcategory for a visual inspection of any trends over time (e.g., gradually increasing male/female citation impact). No statistical tests were performed for this part because there is no prior expectation about what any trends could be.

Linear regression was used to test for an association between first author gender, last author gender and team size on citations. Team size was recorded as a categorical factor with five levels (solo, 2, 3, 4, and 5+ authors) rather than a formula (e.g., log(authors)) because there is enough data to allow this and prior research does not give a consensus about the relationship between team size and citations in the field. Publication year was added as a categorical factor with one level per year because the MNLCS varied between year due to relative changes in the relative impact of US research and changes in the coverage of journals by Scopus. Whilst the regression would be legitimate without publication year variables, adding them improves the accuracy of the fit. In this contest, multilevel modelling would also be a natural choice to deal with publication years (e.g., random intercepts models) but this requires extra distribution assumptions and is not necessary because there is sufficient data to accommodate the loss of degrees of freedom for the publication year variables.

The regression residuals suffer from heteroscedasticity due to differing ranges of citation counts between years. This influences the team size and gender variables (standard errors but not point estimates because the independent variables are all binary) because these change over time. A weighted least squares regression estimation technique that is robust to heteroscedasticity was used to overcome these issues (Yohai 1987), as implemented in lmrob, from the r package robustbase (Rousseeuw, Croux, Todorov, Ruckstuhl, Salibian-Barrera, Verbeke, and Maechler 2019). This technique does not require additional distributional assumptions, in contrast to multilevel modelling, and hence is safer.

The normalisation process can produce outliers through articles that are categorised in at least one subcategory that is small and with a low citation score. This issue was checked for by examining the largest residuals for common publication years and fields, but no problems were found.

Results

The results for the main five Immunology and Microbiology specialties are reported here, with the same data available for the other two Scopus subfields (General and Miscellaneous)

in the supplementary materials, together with full regression details, tests and additional graphs.

Four of the five fields had attained a small female majority by 2018 in terms of first-authored journal articles, but none had more than a third of female last authors by 2018 (Table 1). Since the last author is often senior, this suggests a continued lack of women in promoted positions. The field Applied Microbiology and Biotechnology had only 38% female first-authored articles by 2018, a fraction that had stabilised since 2004 (Figure 1).

Table 1. Gender identification and share statistics for Scopus Immunology and Microbiology narrow fields 1996-2018.

Narrow field	Articles with US only authors	Gendered first & last authors %	Female first in 1996 %	Female first in 2018 %	Change 1996-2018 %	Female last in 1996 %	Female last in 2018 %	Change 1996-2018 %
Applied Microbiology & Biotechnology	42646	50.0	27.4	38.2	10.8	15.3	25.4	10.1
Immunology	66882	55.8	34.5	50.8	16.3	20.1	32.3	12.2
Microbiology	78854	56.9	34.1	50.5	16.4	19.5	33.3	13.8
Parasitology	40222	58.2	27.6	52.0	24.4	18.5	33	14.5
Virology	15801	57.6	33.8	51.0	17.2	20.4	31.6	11.2

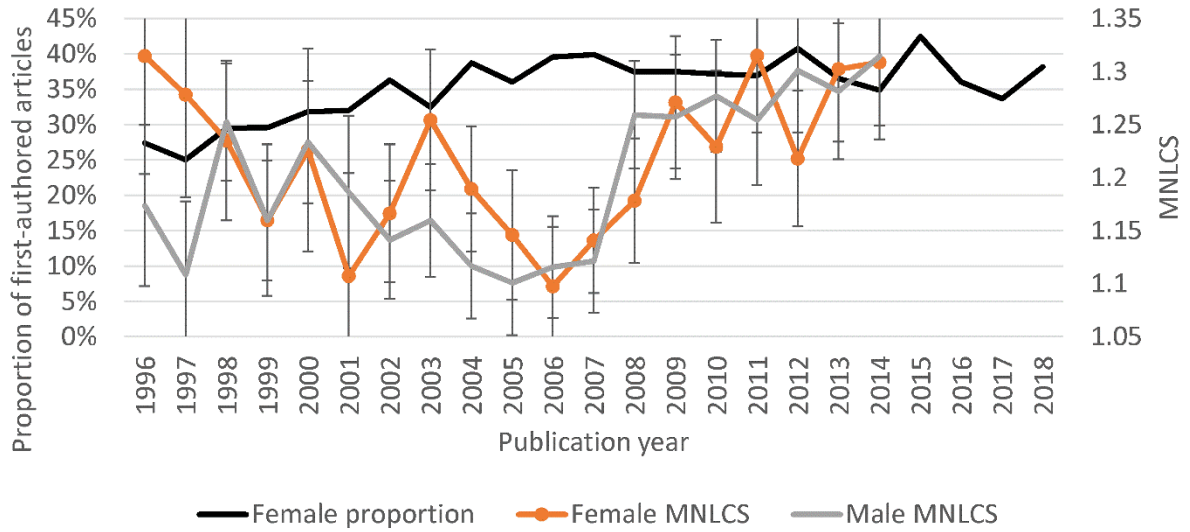

Figure 1. The proportion female first-authored and the average citation impact of female and male first-authored articles by year for *Applied Microbiology & Biotechnology*.

The regression model found first or last author gender effects in three of the five fields, but these were much smaller than the influence of extra authors (Table 2). Removing team size did not change the gender effects much (Table 3). There is a small female first and last author disadvantage in Immunology, a female first author advantage (3%) and last author disadvantage (-3%) in Parasitology, and a female last author disadvantage in Virology (-4%) (Table 2). The female first author citation advantage in Parasitology is consistent over time (Figure 2), despite the apparent conflict with the last author disadvantage.

Table 2. Regression coefficients for the full regression model for Scopus US Immunology and Microbiology narrow fields with gendered first and last authors. Publication years are not shown.

	Model 1					
Narrow field	Female first	Female last	2 auth	3 auth	4 auth	5+ auth
Applied Microbiology & Biotechnology	-0.008	-0.014	0.326***	0.454***	0.441***	0.560***
Immunology	-0.007*	-0.034***	0.121***	0.170***	0.183***	0.273***
Microbiology	-0.002	0.005	0.133***	0.183***	0.195***	0.274***
Parasitology	0.030***	-0.019*	0.149***	0.243***	0.264***	0.362***
Virology	-0.010	-0.043***	0.104***	0.146***	0.145***	0.215***

* p<0.05; ** p<0.01; *** p<0.001

Table 3. Regression coefficients for three reduced regression models for Scopus US Immunology and Microbiology narrow fields with gendered first and last authors. Publication years are not shown.

	Model 2			Model 3		Model 4
Narrow field	Female 1st	Female last		Female 1st		Female last
Applied Microbiology & Biotechnology	0.015	-0.016		0.013		-0.014
Immunology	-0.004	-0.040***		-0.007*		-0.040***
Microbiology	0.002	0.000		0.002		0.001
Parasitology	0.044***	-0.020*		0.043***		-0.015
Virology	-0.009	-0.048***		-0.011*		-0.048***

* p<0.05; ** p<0.01; *** p<0.001

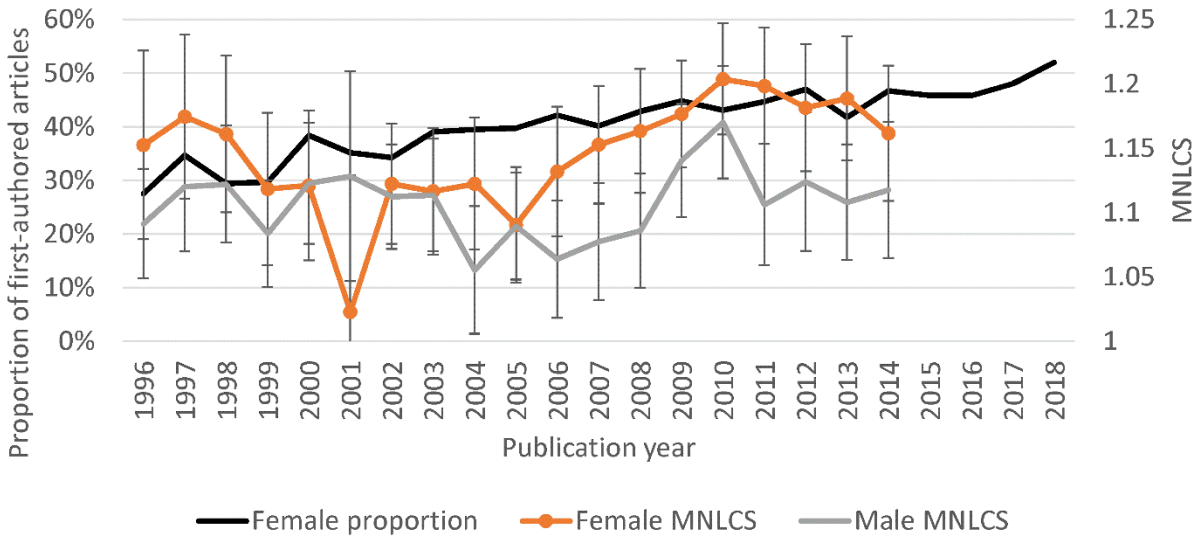

Figure 2. The proportion female first-authored and the average citation impact of female and male first-authored articles by year for *Parasitology*.

Discussion

The results show close to gender parity in the main (first) author of publishing in four of the five areas recognised by Scopus. The exception includes biotechnology, which is a male specialty in the USA. This accounts for the male dominance since biotechnology can involve

software development. The female minorities for last authors reflect the lack of senior females and shows that there are no exceptions within the five Scopus subfields analysed.

Whilst both research questions have positive answers in at least one field (i.e., there are gender differences in citation rates, irrespective of whether team size is taken into account), there does not seem to be a large enough gender effect on citation rates (<5%) for first authors to explain the lack of senior females. For example, if male first authors had tended to produce much higher citation impact research then this could have explained their greater chance for promotion. This effect is also balanced between fields for first authors.

There is a small last author citation advantage favouring males in most fields. If last authors tend to be senior then this suggests that senior males tend to generate slightly more citations for research that they supervise, but this difference is much smaller than the additional author effect and seems too small to account for the continuing domination of males in senior positions. It is also not universal in all fields. Since last authors are likely to be already senior and there is no corresponding male first author citation advantage, the small male last author citation advantage seems to occur after promotion. Second order effects may also influence this last author variation, however, such as a greater proportion of male last authors being senior (and hence perhaps better resourced), or minor gender differences in micro-specialties that influence citation rates.

References

Campanario JM. Empirical study of journal impact factors obtained using the classical two-year citation window versus a five-year citation window. *Scientomet* 2011;**87**:189-204.

Ceci SJ, Williams WM. Understanding current causes of women's underrepresentation in science. *PNAS* 2011;**108**:3157-3162.

Diekman AB, Steinberg M, Brown ER, Belanger AL, Clark EK. A goal congruity model of role entry, engagement, and exit: Understanding communal goal processes in STEM gender gaps. *Pers Soc Psych Rev* 2017;**21**:142- 175.

Holman L, Stuart-Fox D, & Hauser CE. The gender gap in science: How long until women are equally represented? *PLoS Biol* 2018;**16**:e2004956.

Köster C, Klingelhöfer D, Groneberg DA, Schwarzer M. Rotavirus - global research density equalizing mapping and gender analysis. *Vaccine*, 2016;**34**:90-100.

Moral-Munoz JA., Lucena-Antón D, Perez-Cabezas V, Carmona-Barrientos I, González-Medina G, Ruiz-Molinero C. Highly cited papers in Microbiology: identification and conceptual analysis. *FEMS Microb Lett* 2018;365:fny230.

NCES Table 325.20. *Degrees in the biological and biomedical sciences conferred by postsecondary institutions, by level of degree and sex of student: Selected years, 1951-52 through 2013-14*. https://nces.ed.gov/programs/digest/d15/tables/dt15_325.20.asp (23 May 2019, date last accessed).

Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, & Maechler M. robustbase: Basic Robust Statistics. R package version 0.92-8. (23 May 2019, date last accessed)

Sax LJ, Lim G, Lehman K, Lonje-Paulson L. Reversal of the Gender Gap: The biological sciences as a unique case within science, technology, engineering, and mathematics (STEM). *J Wom Min Sci Eng* 2018;**24**:291-324.

Shapiro VS, Kovats S, Parent, MA, Gaffen SL, Hedrick CC, Jain P, & Stephens R Update on gender equity in immunology, 2001 to 2016. *J Immunology* 2016;**197**:3751-3753.

Su R, Rounds J, Armstrong PI. Men and things, women and people: a meta-analysis of sex differences in interests. Psych Bull 2009;**135**:859.

Sweileh WM, Al-Jabi SW, Sa'ed HZ, Sawalha AF, & Abu-Taha AS. Global research output in antimicrobial resistance among uropathogens: A bibliometric analysis (2002–2016). *J Global Antimicrob Res* 2018;**13**:104-114.

Thelwall M, Bailey C, Makita M, Sud P, Madalli D. Gender and Research Publishing in India: Uniformly high inequality? *J Informetrics* 2019;**13**:118–131.

Thelwall M, Bailey C, Tobin C. Bradshaw N. Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *J Informetrics* 2019;**13**:149-169.

Thelwall M. Three practical field normalised alternative indicator formulae for research evaluation. *J Informetrics* 2017;**11**:128–151.

Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. *The Ann Statistics* 1987;**15**:642-656.