Journal of Informetrics 9 (2015) 1007–1017

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Gender differences in scientific performance: A bibliometric matching analysis of Danish health sciences Graduates☆

Tove Faber Frandsen[a,*], Rasmus Højbjerg Jacobsen[b], Johan A. Wallin[c], Kim Brixen[a], Jakob Ousager[c]

[a] Odense University Hospital, Sdr. Boulevard 29, DK-5000 Odense, Denmark
[b] KORA, Danish Institute for Local and Regional Government Research, Copenhagen, Denmark
[c] University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Article history:
Received 1 May 2015
Received in revised form
24 September 2015
Accepted 24 September 2015

Keywords:
Scientific productivity
Impact
Gender differences
Gender gap

ABSTRACT

The aim of this study is to compare PhD students' performance with respect to gender using a number of matching methods. The data consists of fine-grained information about PhD-students at the Institute of Clinical Research at the University of Southern Denmark. Men and women are matched controlling for sub-disciplinary affiliation, education, year of enrolment and age. Publications and citations are identified in Web of Science.

Our study shows that the average total number of publication is slightly higher for men than for women. Excluding the "other" group of publications from the analyses reveals that there is a negligible difference between men and women in terms of published articles. A substantial proportion of women is on maternity leave during the time period analysed and thus we would expect their productivity to be considerably lower. Similarly, we have found very little difference between the citation impact of men and women.

We find matching methods to be a promising set of methods for evaluating productivity and impact of individuals from various sub-fields, universities and time periods as we are able to discard some of the underlying factors determining the results of analyses of gender differences in productivity and citation impact.

1. Introduction

Students' decision to enter into research is influenced by many factors such as future career opportunities, chances of academic success and also considerations towards family issues. Graduate students' productivity may be influenced by family obligations such as children, partner, parents etc. According to Kelchtermans and Veugelers (2013) general factors determining the productivity of a researcher are talent, luck, effort and cumulative effect (a.k.a. the Matthew effect). For women, the choice of having children and to and take maternity leave will – all other things equal – make it more difficult to offer the extra effort that may be demanded in a competitive academic environment. Difference in the level of scientific productivity between men and women remains a research issue that has been explored bibliometrically in more than three decades (Mairesse & Pezzoni, 2013) and more studies are frequently called for (Fox & Stephan, 2001; Mairesse &

☆ The authors would like to thank Thomas Aagaard for assistance with data collection and are grateful for valuable comments and suggestions from the referees.

* Corresponding author.
E-mail address: t.faber@videncentret.sdu.dk (T.F. Frandsen).

http://dx.doi.org/10.1016/j.joi.2015.09.006

Pezzoni, 2013). Mairesse and Pezzoni have provided an overview of existing models for explaining the potentially lower productivity of women (family engagements, marital status and policies in favour of women, institutional specificities, discipline specificities, etc.). Hunter and Leahey (2010) analyse the effect of children on the entire careers of academics which is different for men and women. Fridner et al. (2015) investigate age, academic position, collaboration with former supervisor, control at work and exhaustion as predictors of productivity and finds different patterns for men compared to women. Cole and Zuckerman (1984) coined it the “productivity puzzle” while others prefer the term “productivity gap” (Abramo, Cicero, & D’Angelo, 2015).

Differentiating between academic fields on a relatively general level of e.g. social sciences, physical science etc. Duch et al. (2012) find differences in some but not all fields and find that the lower publication rates of female faculty members are correlated with the amount of research resources typically needed in the discipline. Only a few studies have taken the potential field or discipline differences in to account using a more fine-grained division between these. Some studies find no difference when controlling for academic rank (Paik et al., 2014; Tomei et al., 2014) whereas Tomei et al. do find a statistical significant difference in productivity (Paik et al., 2014). Findings by Abramo et al. (2015) suggest that even within the same discipline there are different tendencies and consequently, analyses may be more appropriately done on the level of sub-disciplines.

It is worth noting that the study by Khan et al. (2014) as it seems to indicate that there are significant differences across departments or programs within the same field. They found mean *h*-index values ranging from app. 5 to app. 30. Consequently, one should be careful when aggregating data from various universities, a point also noted by others (Mairesse & Pezzoni, 2013).

The existing studies of the gender productivity and impact puzzle typically include a much greater cohort of men than women (in a recent study the women only make up as little as 21 percent of the included faculty members, Duch et al., 2012). The over-representation of male academics is well known (see e.g. Sugimoto, Lariviere, Ni, Gingras, & Cronin, 2013), and data on the staff of national research systems indeed confirm that there is a significant deficit in the presence of women. The share of women ranges from 14 percent to 45 percent with a median value in 2011 of 33 percent (OECD, 2015). However, the share of women is not distributed equally across the data set. Compared with men, women tend to be younger and of lower academic rank. Moreover, some disciplines have very few female researchers whereas women dominate others.

The so-called leaking pipeline can be illustrated by depicting the obvious differences in the presence of women on bachelor, master, doctoral and faculty level (Duch et al., 2012). This will inevitably skew the dataset and possibly bias the analyses if not taken in to account. We therefore suggest the use of matching methods.

In this study, the aim is to analyse PhD students’ performance with respect to gender. Outcomes are publications and citations. We control for a number of factors considered important for productivity including the department (i.e. academic subfield) where the students were enrolled and use an exact matching method to construct comparable groups of men and women.

This paper uses a unique data set of health science PhD students within the clinical specialties from the University of Southern Denmark (formerly known as Odense University).

2. Overview of related literature

A number of studies have provided evidence of a gender gap documenting women researchers to be less productive than men. Most studies use samples restricted to one discipline and in some cases from one country e.g., Spanish psychologists (Barrios, Villarroya, & Borrego, 2013), Swedish physicians (Fridner et al., 2015), library and information scientist (Penas & Willett, 2006), German cardiologists (Bohm, Papoutsis, Gottwik, & Ukena, 2015), social psychologists (Cikara, Rudman, & Fiske, 2012), and German medical researchers (Kretschmer, Pudovkin, & Stegmann, 2012).

Other studies confirming the gender gap in productivity cover several disciplines. Baccini, Barabesi, Cioni, and Pisani (2014)study Italian researchers from various disciplines and find that women are less productive than men. Another Italian studies confirm their findings (Abramo, D’Angelo, & Caprasecca, 2009). Studies on Spanish data (Mauleon, Bordons, & Oppenheim, 2008) as well as Croatioan (Prpic, 2002) and American data (Xie & Shauman, 1998) are also available. Stack finds great differences in gender effect between different fields using data from the National Research Council (Stack, 2004). Canadian data supports their findings (Lariviere, Vignola-Gagne, Villeneuve, Gelinas, & Gingras, 2011).

A number of studies provide further insight into the existence of a gender gap in productivity and does not confirm the existence without reservations. Mairesse and Pezzoni (2013) study French physicists and find a substantial lower production by women, however, having controlled for a number of variables (e.g. non-equal chances of promotion and non-publishing spells) the differences disappear. Their findings are to a certain degree supported by a study of two disciplines (Bordons, Morillo, Fernandez, & Gomez, 2003) and nano science (Sotudeh & Khoshian, 2014). Eloy et al. (2013a) find that the productivity of women in Otolaryngology equals or surpasses that of men later in their careers. The differences between subfields are confirmed in a large study of various medical disciplines (Eloy et al., 2013b). A large-scale study confirms that the differences in publication rate and impact are discipline-specific (Duch et al., 2012). van Arensbergen, van der Weijden, and van den Besselaar (2012) stress the importance of keeping the skewed publication data in mind as few highly productive authors can skew the data set. Some studies find the gap to be decreasing over time (Abramo et al., 2009; Mauleon et al., 2008; van Arensbergen et al., 2012; Xie & Shauman, 1998). The large-scale analysis by West, Jacquet, King, Correll, and Bergstrom

(2013) stresses that although the differences are decreasing inequities remain due to e.g. author order and the number of authors.

Turning to the results of studies on scientific impact we also have a number of studies of which some use traditional citation based indicators whereas other use hybrid indicators such as the *h*-index. In the following an overview will be provided of both types of indicators regardless of indicator. Some of these studies cover a well-defined area, e.g. a study of the Martinez, Lopez, and Beebe (2015) Society, natural resources scientists and chemists (Bordons et al., 2003), social–personality psychology (Haslam et al., 2008), library and information science (Penas & Willett, 2006). Some of the studies covering one or two disciplines or even sub-disciplines have been unable to detect a gender gap (Bordons et al., 2003; Khan et al., 2014; Long & Fox, 1995; Penas & Willett, 2006; Slyder et al., 2011). Others show mixed results (Kretschmer et al., 2012; Lopez et al., 2014; Martinez et al., 2015; Paik et al., 2014). Finally, there are examples of a gender advantage for men in terms of impact (e.g. Andriessen, Krysinska, & Stack, 2015; Hunter & Leahey, 2010) and also of a gender advantage for women (Long, 1992).

Looking at studies covering more disciplines a study of 526 scientists (Cole & Zuckerman, 1984), a study of 6000+ Quebec university professors (Lariviere et al., 2011), and a study of 731 Spanish PhDs (Borrego, Barrios, Villarroya, & Olle, 2010) find an impact disadvantage for women. On the other hand a study of 4000+ faculty members at top US research universities finds that any gender difference is highly dependent on discipline, and in some disciplines women tend to have a greater fraction of higher impact publications than men (Duch et al., 2012).

A number of related measures also exist, and it may be worth mentioning that a study finds that men seem to apply for more patents (Ding, Murray, & Stuart, 2006), a study finds that men become principal investigator more often than women (van Dijk, Manor, & Carey, 2014), and a study finds that women have different collaboration patterns than men (Abramo, D'Angelo, & Murgia, 2013).

Summing up, the existing literature on gender differences in scientific productivity and impact are dominated by studies based on relatively small data samples. Larger studies tend to have lower data quality (typically in the process of attributing publications to individuals) but confirm that one should be careful when aggregating the data without considering a number of factors (e.g. discipline or even sub-discipline, development over time and career stage of the researchers). The existing literature also indicates that gender differences in productivity and impact may be changing over time.

3. Matching methods

A relationship between gender and bibliometric indicators may be caused by selection bias. Demographic characteristics of PhD students have changed over time, and a relationship between gender of PhD students and their publication activity as well as citation impact may be caused by these changing characteristics.

This study uses matching methods to compensate for the fact that the correct functional form that will eliminate selection bias in the relationship between gender and bibliometric indicators is unknown. In this approach, a treatment group is matched to controls. Matching methods seek to replicate or imitate a randomized experiment in which the matched treated and control observations do not differ systematically from each other. Matching methods are widely used in economics (see e.g. Jespersen, Munch, & Skipper, 2008; Lechner & Wiehler, 2011) and although the methods are not commonly used in bibliometrics examples do exist (Diekmann, Naf, & Schubiger, 2012; Eloy, Svider, Setzen, Baredes, & Folbe, 2014; Hopewell & Clarke, 2003). Cole and Zuckerman use matching for their study of gender differences in patterns of publication. They use department as matching criteria (Cole & Zuckerman, 1984). This study applies a matching method to form PhD student pairs for which treatment can be considered randomly assigned within each pair. This method is comparable to the case-control design in epidemiology.

Grouping PhD students with similar characteristics according to a few variables balances the students on these particular variables, but does not help to eliminate biases due to disparities in other variables. The challenge is to find the variables that encompass all the characteristics that are deemed to be of most importance both for the probability of high publication activity and high citation impact, such that any remaining differences may be attributed to gender alone.

The PhD students' academic discipline and age must be deemed important characteristics for the present study. There are great differences in citation practices across disciplines even within the medical area (Moed, 2006). Consequently, we need to take the research area into account. In this case educational background and research areas are used as proxies for discipline related differences.

Studies of publication activity and citation impact of specific researcher groups may include chronological age (Lissoni, Mairesse, Montobbio, & Pezzoni, 2011), career age (Xie & Shauman, 1998), academic experience (Frandsen & Nicolaisen, 2012) or professional categories (Costas & Bordons, 2011). In this case, however, their academic experience, career age or professional category is identical, and consequently we focus on chronological age as well as enrolment year. Lariviere finds mixed results in regard to the age of the PhD students (Lariviere et al., 2011). He finds that there is a tendency to older students publishing in higher impact factor journals than younger researchers However, looking at observed citation rates the opposite is found.

The beauty of the matching design is that after controlling for other key factors that may influence publications and citation patterns, we can answer the main question of whether there are gender differences in scientific productivity during and after graduate school by comparing the simple means from the two groups. Hence, if the average number of publications

Table 1
Reduction in the number of observations from master data to potential analysis data set.

		Removed	Remaining
A	Master data set	–	541
B	PhD-degree not completed	264	277
C	PhD-study started after 2008	30	247

Table 2
Reduction from potential analysis data set to actual analysis data set.

		Pairs	Females	Males
C	Potential analysis data set	–	118	129
D	Match found	78	78	63
E	After removal of outlier	77	77	62
F	After removal of non-unique names	73	73	61

is higher (in a statistically significant sense) in the male control group than in the female group, then we conclude that male graduate students have a higher productivity than female students.

4. Participants and methods

The data set of our study contains a total of 541 PhD-students at the Institute of Clinical Research at the University of Southern Denmark (until 1998 known as University of Odense) who enrolled the PhD-program between 1993 and 2013[1]. The participants in this study are selected at the time of enrolment. We could also have drawn a sample of faculty members, however, the career paths of men and women are not identical (Baccini et al., 2014; Kelchtermans & Veugelers, 2013), consequently, the characteristics of the group of women entering science are not equivalent to those of the female faculty members. This study focuses on potential gender differences in the early career phase whereas other studies have focused on later career stages (see e.g. van Arensbergen et al., 2012).

Table 1 shows the reduction in the data set from the master data set (A) to the potential analysis data set (C). Of the 541 individuals 264 had not completed their PhD-degree, resulting in 277 remaining participants (data set B). Most of the 264, who were excluded, had enrolled recently (i.e. less than 3 years previously), but 20 were dropouts. Due to the low number of dropouts we are not able to include them in the analyses as the matching criteria are the same as for the students completing their degree.

Since we need to follow the students five years after enrolment (see more on the specific choice of publication and citations windows below), the latest year for students in our analysis to be enrolled is 2008. Thus we excluded further 30 subjects resulting in a total of 247 (118 women and 129 men) participants for analysis (data set C). To control for sub-disciplinary affiliation, education, year of enrolment and age, we used the following matching criteria:

- *Sub-discipline:* This variable is specified as the hospital department that the individuals were affiliated with. In some cases individuals were affiliated with more than one department, and in this case we made use of both possible sub-disciplines to look-up a match. Using a set of matching methods that includes matching on the level of sub-discipline implies that the production of a given female researcher is compared with that of a male researcher within that same sub-discipline which means that field normalization is not necessary.
- *Education:* Most individuals in the data set were MDs, but about a third has another educational background, whereof most have a degree in nursing or science.
- *Age:* We allow up to five years difference in age. The male and female researchers can only be matched if there is no more than 5 years difference in chronological age. However, −5 to 5 years are computed on the basis of their specific enrolment year.
- *Enrolment year in the PhD-study program:* We also allow a maximum difference of five years. The data set covers a 20 years period; however a woman can only be matched to a man if there is no more than five years between their enrolment consequently, generation differences are left out of the analyses.

The matching was carried out in the following way: For each of the 118 women in the data set it was tested whether a man fulfilling the matching criteria could be found. If the was not the case, the woman was left out of the analysis. If multiple possible matches were found, the male match was found by drawing lots between the candidates.

Table 2 shows the reduction in the data set from the potential analysis data set (C) to the final data set used in analysis (F).

[1] For further description of the Danish PhD-program see the ministerial order (http://ufm.dk/en/legislation/prevailing-laws-and-regulations/education/files/engelsk-ph-d-bekendtgorelse.pdf) or an introduction by the Danish Ministry of Higher Education and Science: (http://studyindenmark.dk/study-options/find-your-international-study-programme/phd-research).

Table 3
Comparison of women and men matched and unmatched.

		Mean		*t*-Test	
		Women	Men	*t*	*p* > *t*
Birth year	Unmatched	1972.6	1970.9	2.32	0.021
	Matched	1968.5	1967.9	0.60	0.550
Enrolment year	Unmatched	2007.3	2005.0	4.56	0.000
	Matched	2001.3	2000.9	0.58	0.561
Not medical degree	Unmatched	0.3301	0.2085	3.15	0.002
	Matched	0.1096	0.1096	0.00	1.000

When using the above listed matching criteria, 78 pairs of woman-and-matched-man were found (data set D in Table 2). The 78 women were matched with 63 different men indicating that some men served as control for several women. When the treatment and comparison subjects are very different or the number of control units are very small, i.e. the overlap is relatively small, finding a satisfactory match by matching without replacement can be very problematic. Matching with replacement minimizes the differences between the matched comparison units and the treatment unit as matching with replacement allows a given untreated subject to be included in more than one matched set. Consequently, bias is reduced (Dehejia & Wahba, 2002). However, as matching with replacement allows the treatment effect estimate to on a relatively small number of controls, the number of times each control is matched should be monitored (Stuart, 2010). In this case 11 men each served as control for two different women, three men served as control for three women, and one man served as control for four women. Consequently, the replacement is at an acceptable level.

Of 78 matches found, one pair was excluded due to an extreme outlier[2] (data set E). Finally, 4 pairs had to be excluded, because it was not possible to identify the person(s) uniquely in the citation and publication databases. This leaves a total of 73 pairs to be used in the analysis (data set F). From the numbers in Table 2 we can see that some of the duplicate controls also were eliminated during this process. Thus in the final analysis data set 49 men each served as control for one woman, 9 men were control for two women, two men were control for three women, and one man was control for 4 women. Table 3 offers an overview of characteristics for matched and unmatched groups.

Publications and citations were identified for each individual using Web of Science. The following data were collected for each individual in the sample:

- The number of journal articles from year −5 to year +5 with enrolment year as year 0;
- The number of reviews (systematic as well as unsystematic) from year −5 to year +5 with enrolment year as year 0;
- The number of other publications (e.g. conference proceedings, meeting abstracts) from year −5 to year +5 with enrolment year as year 0; and
- Citations to all three publication categories from year −5 to year +5 with enrolment year as year 0

The number of publications has not been fractionalised. Gender may play a role in collaboration patterns in various ways that again affect the productivity as well as impact (Barrios et al., 2013; West et al., 2013). However, we have not included collaboration patterns in the analyses as we wish to detect gender differences not correct for them.

If the publication list of an individual could not be positively determined due to more authors with the same name online publication lists were consulted. If such a list was not available the individual was discarded from the analysis.

When using the matching method we find a subset of untreated individuals who are similar to those of the treated persons, or vice-versa and then we can estimate—in this case the effect of gender on productivity and impact by an ordinary *t*-test.

5. Results

Roughly two thirds of the PhD students are MDs with the remaining being divided among a variety of related health science degrees within nursing, public health, science, biomedicine and microbiology a.o. The data set reveals that the length of parental leave on average is 99 days for women and 9 days for men. Excluding any parental leave the men completes the PhD programme in 4.3 years on average and the women in 4.2 years.

Before looking at the development for the female PhD students and their matched male counterparts we first show a simple comparison of the productivity of the two groups before and after enrolment in the PhD-study. Table 4 shows the average scientific productivity for the two groups in the five years before enrolment as measured by the total number of publications, the total number of journal articles published and the total number of citations received.

First and foremost we can see from Table 4 that on the three measures shown, the average productivity of the female graduate students and their matched male counterparts does not differ in a statistically significant way. Also, the crude

[2] The pair, which was excluded do to an extreme observation, included an individual who was co-author of a paper with more than 50 authors, and the paper was cited more than 6000 times.

Table 4
Scientific productivity before enrolment.

Variable	Men		Women			
	Mean	Std.dev	Mean	Std.dev	*t*-Test	*p*-Value
Total publications	1.301	2.396	0.959	3.289	0.719	0.473
Journal articles	0.877	2.173	0.767	3.107	0.247	0.805
Citations received	8.589	26.420	18.055	81.730	0.942	0.348

Table 5
Scientific productivity after enrolment.

Variable	Men		Women			
	Mean	Std.dev	Mean	Std.dev	*t*-Test	*p*-Value
Total publications	7.370	7.338	5.562	5.907	1.640	0.103
Journal articles	4.534	3.902	3.877	4.079	0.995	0.321
Citations received	105.945	154.762	99.110	166.616	0.257	0.798

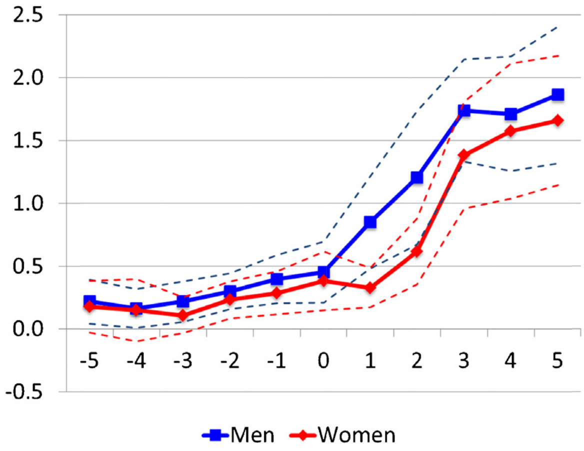

Fig. 1. Average number of publications according to gender. *Note:* The difference between the two groups is significantly different from zero in year 1 and year 2, when using a standard *t*-test and the 5 percent significance level. All other years are not different in a statistically significant way, based on a 5 percent significance level. The red and blue dashed lines show the 95 percent confidence interval for men and women, respectively. Data tables are available in Appendix A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

averages for total publications and journal articles are quite similar, but the women have received almost ten citations more one average. However, there is a large standard deviation for the number of citations, such that this difference is not statistically significant.

Table 5 shows the same numbers as Table 4, but for the five years after enrolment[3]. From the table we can see that scientific productivity is not statistically different between the groups of men and women during the five years after enrolment either. The number of publications and the number of citations received are much higher after enrolment than before, which is also what we would expect given that the students undertake a scientific education.

We now turn to the development in productivity over time. Fig. 1 shows the average number of publications for the male and the female group of graduate students over time. Fig. 1 and the following two figures should be read as follows: The numbers on the horizontal axis represents years from the enrolment year. Hence, the number zero is the enrolment year, the number 2 is two years after enrolment, and the number −3 is three years before enrolment etc. Appendix A presents the numbers from Figs. 1–3 and the corresponding statistical tests in three tables.

The tendencies depicted in Fig. 1 exhibit a pattern which could be expected given the population in the current analysis. The average number of published articles is very low – in the order of 0.1–0.2 – in the years until the enrolment in the PhD-study. Thereafter the number rises for both men and women up to a level of about 1.6–1.8 publications per year per individual.

[3] The numbers in Tables 4–5 show the averages for the five years before enrolment and the five years after enrolment, leaving out the year of enrolment from either table. This has been done, because our data does not allow us to observe the time of year of enrolment, which means that, on average, half of the year should be “before enrolment” and the other half should be “after enrolment”. Including the year of enrolment in either of Table 4 or 5 does not alter the conclusions.

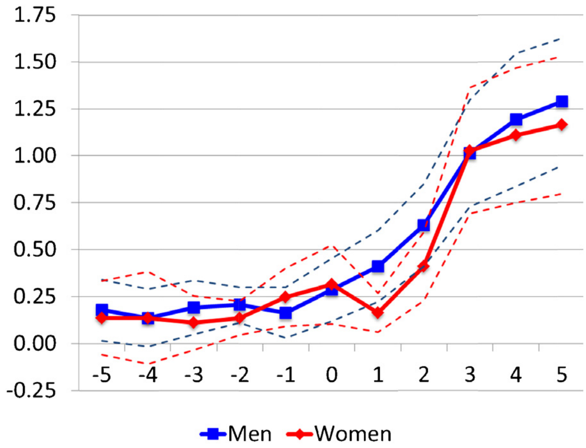

Fig. 2. Average total number of journal articles. *Note:* The difference between the two groups is significantly different from zero in year 1, when using a standard *t*-test and the 5 percent significance level. All other years are not different in a statistically significant way. The red and blue dashed lines show the 95 percent confidence interval for men and women, respectively. Data tables are available in Appendix A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

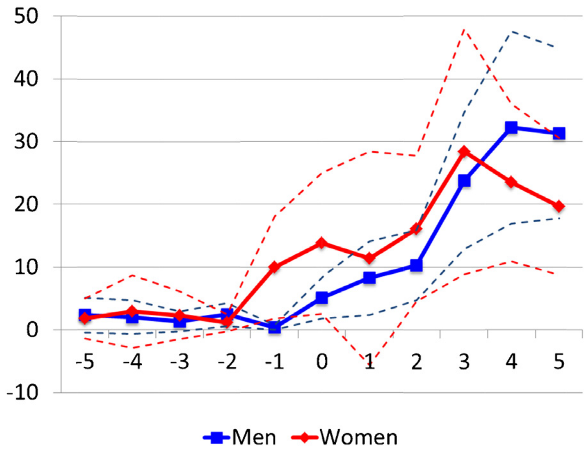

Fig. 3. Average number of citations to work published in year.
Note: The difference between men and women is statistically insignificant in all years except year −1, based on a 5 percent significance level. The red and blue dashed lines show the 95 percent confidence interval for men and women, respectively. Data tables are available in Appendix A. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 1 shows that the average total number of publication is slightly higher for men than for women, although this difference is only statistically significant in years 1 and 2. The main reason for men having a larger number of total publications is that men have more publications in the type ‘other’ which is reviews, book reviews, meeting abstracts and corrections. Including all other types of publications in the analysis reveals that the women produce less of these other types of publications, which is obvious when looking at Fig. 2.

Fig. 2 shows the number of articles and thus the “other” group of publications is excluded. Although articles are the main channel for publishing academic research, other types of publications, such as brief communications, meeting abstracts and reviews are also used. Some of them are probably of less scientific content (e.g. meeting abstracts), others are highly valued and cited (e.g. reviews). In the present study there are too few of these publications to perform more fine-grained analyses.

Excluding this group of publications from the analyses reveals that there is practically no difference (with exception of year one) between men and women in the productivity measured as the number of articles published. In this analysis the difference in productivity is only statistically significant in year 1 at the 5 percent significance level. However, even if a larger data sample should result in analyses showing that the differences are statistically significant in years 1 to 5 at the 5 percent significance level we find it remarkable that the difference in productivity is not greater. Based on their age we would expect a considerable share of these women to be on maternity leave at some time during the years −5 to 5. Hence we would expect their productivity to be substantially lower.

As the task of writing and publishing academic research is at the core of any graduate study, it could be the case that any inherent difference in the scientific productivity between male and female students would reveal itself in the number of publications, but rather in the impact of the publications. To examine whether this is the case we have also looked at the total number of citations to the published work from the two groups. This is shown in Fig. 3.

Looking at Fig. 3, we can see a pattern similar to what we observed in Figs. 1 and 2, namely that the graduate student receive very few citations to the material published in the years up to their enrolment in the PhD programme with only 2–3 citations per individual. This is obviously a consequence of the few published articles as seen above. As the amount of published material rises, so does the number of citations received, and in year 3–4 after enrolment the average student publishes material that receives a total of 20–30 citations. Turning to the gender difference, we observe a very little (and overall statistically insignificant) difference between men and women. At face value, it may appear that the difference in citations is bigger than the difference in published articles from Fig. 1 above, but as already stressed, this difference is not statistically significant, since the standard deviation of the number of citations is much larger than for published articles.

6. Discussion and conclusion

Our study shows that the average total number of publication is slightly higher for men than for women. Excluding the "other" group of publications from the analyses reveals that there is practically no difference (with exception of year one) between men and women in terms of published articles. We find it remarkable that the difference in productivity is not greater. A sizeable share of these women is on maternity leave at some time during the years −5 to 5, and consequently we would expect their productivity to be considerably lower. Likewise we find very little difference between the citation impact of men and women. At face value, it may appear that the difference in citations is bigger than the difference in published but this difference is not statistically significant, since the standard deviation of the number of citations is large.

Before turning to the discussion and conclusion we need to address the limitations of the study design. First of all the data set is limited to a relatively small number of health science PhD students within the clinical specialties from the University of Southern Denmark. Ideally the dataset should be considerably larger in order to be able to generalise our findings. However, the present study can shed some light on the productivity puzzle by using a set of methods that discards some of the underlying factors which could otherwise be determining the results of analyses of gender differences in productivity and citation impact. Second, the number of control variables is limited to educational background, age and sub-discipline. Although this list contains the most central control variables, other variables (e.g. whether the students actually have had a child) would be of interest. We have information on the length of maternity or paternity leave during the PhD study, but we do not know the number of children the students have had. This means that if a student just prior to starting the PhD study had a child he or she will appear as childless in our analysis. In order to avoid this bias of incomplete data, and because our focus is not on the differences between students with or without children we have not included the information in the analysis. Third, the time period analysed is limited to five years after enrolment in the PhD study, and it is possible that other effects would materialize if the analysis period was extended to, say, 10 years. Consequently, this study only allows for analyses of the production and citation of publications 5 years after enrolment.

Our study shows that there is practically no difference between men and women in the productivity measured as the number of articles published. The difference may be greater if we include other publications than just journal articles.

A large number of previous studies on the "productivity puzzle" or "productivity gap" confirm the gender gap in productivity (Abramo et al., 2009; Baccini et al., 2014; Barrios et al., 2013; Bohm et al., 2015; Cikara et al., 2012; Fridner et al., 2015; Kretschmer et al., 2012; Lariviere et al., 2011; Mauleon et al., 2008; Penas & Willett, 2006; Prpic, 2002; Xie & Shauman, 1998) A number of studies explore the causes and characteristics of the potential gender gap. These studies of a gender gap confirm that one needs to take a number of variables in to account to avoid bias (Bordons et al., 2003; Duch et al., 2012; Eloy et al., 2013b; Mairesse & Pezzoni, 2013; Sotudeh & Khoshian, 2014). Finally, the literature shows that the gender gap may be changing over time and that younger generations may not be characterised by gender differences to the same degree as the older (Abramo et al., 2009; Eloy et al., 2013a; Mauleon et al., 2008; van Arensbergen et al., 2012; Xie & Shauman, 1998).

Consequently, a fine-grained analysis is required to explore the gender differences in productivity taking a number of variables into account in order to avoid aggregating samples of men and women from various universities and fields or sub-fields. Furthermore, male academics should not be over-represented in the dataset. We base our analysis on matching methods. Using this fine-grained analysis we find a much smaller difference between genders with respect to scientific productivity than some of the previous studies.

The seemingly negligible productivity gap between male and female PhDs during and just after their PhD years may be somewhat unexpected given the fact that, all other things equal, women in this age are more on maternity leave than men, even in societies with a relatively high degree of gender equality. In our dataset the female PhDs spent roughly ten times as much time on maternity leave than their male counterparts (mean score 99 days for women, 9 days for men). But apparently the time used on maternity leave has no substantial effect on the academic productivity measured over an absolute span of time: In general, female PhDs seem to publish just as much as male PhDs over a period of in this instance 5 years, even though female researchers would spend a relatively larger portion of this time on maternity leave. One might speculate whether this phenomenon is best explained by women researchers being more effective than men, and therefore needing less time to produce the same. An alternative explanation could be that young researchers' academic ambitions and activities are in fact very seldom in a total stand still, regardless of the researcher being on maternity leave or not. Consequently, it could be hypothesised that the absence of a gender difference is in fact due to selection.

We find very little (and overall statistically insignificant) difference in the number of received citations between men and women. The existing literature is characterized by a number of studies covering one or two disciplines and some of them confirm the existence of a gender gap (e.g. Andriessen et al., 2015; Hunter & Leahey, 2010; Long, 1992) whereas other studies are unable to detect a gender gap (Bordons et al., 2003; Khan et al., 2014; Long & Fox, 1995; Penas & Willett, 2006; Slyder et al., 2011). Finally, some of the studies show mixed results (Kretschmer et al., 2012; Lopez et al., 2014; Martinez et al., 2015; Paik et al., 2014). Looking at studies covering more disciplines some detect a gender gap (Borrego et al., 2010; Cole & Zuckerman, 1984) whereas one study finds that any gender difference in impact is highly dependent on discipline, and in some disciplines women tend to have a greater fraction of higher impact publications than men (Duch et al., 2012).

Using matching methods including sub-disciplinary differences in impact patterns in the present study we find very little (and overall statistically insignificant) difference between men and women suggesting that if there is a difference it is remarkably small.

We find matching methods to be a promising set of methods for evaluating productivity and impact of individuals from various sub-fields, universities and time periods as we are able to discard some of the underlying factors that could be driving the results of analyses of gender differences in productivity and citation impact.

Previous studies have established that there seems to be a decreasing tendency in the productivity difference over time. The present study does not include analyses of the development over time. An analysis of the differences over time would require an increase in publication and citation window and we hope future research will be able to shed light on this hypothesis on a fine-grained level.

Appendix A. Data tables for Figs. 1–3

This appendix contains three tables, one for each of Figs. 1–3, showing the data points in the figures and the *t*-tests carried out to examine whether any gender differences were statistically significant.

Tables A1–A3.

Table A1
Data points and *t*-tests for Fig. 1 (total number of publications).

	Men		Women			
Year	Mean	Std.dev.	Mean	Std.dev.	*t*-Test	*p*-Value
−5	0.219	0.750	0.178	0.887	0.302	0.763
−4	0.164	0.667	0.151	1.063	0.093	0.926
−3	0.219	0.692	0.110	0.614	1.012	0.313
−2	0.301	0.617	0.233	0.635	0.661	0.510
−1	0.397	0.812	0.288	0.736	0.855	0.394
0	0.452	1.041	0.384	0.995	0.406	0.685
1	0.849	1.569	0.329	0.668	2.608	0.010*
2	1.205	2.273	0.616	1.126	1.984	0.049*
3	1.740	1.740	1.384	1.823	1.207	0.229
4	1.712	1.954	1.575	2.303	0.388	0.699
5	1.863	2.335	1.658	2.206	0.546	0.586

Note: *t*-Tests were carried out as standard two-tailed tests for equal sample size and equal variance as the expected variances of the two samples a priori were the same.
Asterisks (*) denote values that are statistically significant at a 5 percent significance level.

Table A2
Data points and *t*-tests for Fig. 2 (articles published).

	Men		Women			
Year	Mean	Std.dev.	Mean	Std.dev.	*t*-Test	*p*-Value
−5	0.178	0.694	0.137	0.839	0.323	0.747
−4	0.137	0.652	0.137	1.058	0.000	1.000
−3	0.192	0.616	0.110	0.614	0.808	0.421
−2	0.205	0.407	0.137	0.384	1.046	0.297
−1	0.164	0.578	0.247	0.662	0.799	0.425
0	0.288	0.716	0.315	0.896	0.204	0.839
1	0.411	0.814	0.164	0.441	2.276	0.024*
2	0.630	0.936	0.411	0.779	1.538	0.126
3	1.014	1.219	1.027	1.443	0.062	0.951
4	1.192	1.515	1.110	1.542	0.325	0.746
5	1.288	1.448	1.164	1.572	0.493	0.623

Note: *t*-Tests were carried out as standard two-tailed tests for equal sample size and equal variance as the expected variances of the two samples a priori were the same.
Asterisks (*) denote values that are statistically significant at a 5 percent significance level.

Table A3
Data points and *t*-tests for Fig. 3 (citations received).

Year	Men		Women		*t*-Test	*p*-Value
	Mean	Std.dev.	Mean	Std.dev.		
−5	2.342	11.764	1.781	13.745	0.265	0.791
−4	2.027	11.567	2.904	24.813	0.274	0.785
−3	1.356	6.871	2.301	16.359	0.455	0.650
−2	2.425	7.690	1.110	5.953	1.155	0.250
−1	0.438	1.563	9.959	34.741	2.339	0.021*
0	5.068	14.046	13.808	48.003	1.493	0.138
1	8.274	25.132	11.384	72.969	0.344	0.731
2	10.301	23.935	16.123	49.693	0.902	0.369
3	23.808	46.664	28.411	83.804	0.410	0.682
4	32.274	65.746	23.507	53.753	0.882	0.379
5	31.288	57.950	19.685	46.773	1.331	0.185

Note: t-Tests were carried out as standard two-tailed tests for equal sample size and equal variance as the expected variances of the two samples a priori were the same.
Asterisks (*) denote values that are statistically significant at a 5 percent significance level.

References

Abramo, G., Cicero, T., & D'Angelo, C. A. (2015). Should the research performance of scientists be distinguished by gender? *Journal of Informetrics*, *9*(1), 25–38, http://dx.doi.org/10.1016/j.joi.2014.11.002

Abramo, G., D'Angelo, C. A., & Caprasecca, A. (2009). Gender differences in research productivity: A bibliometric analysis of the Italian academic system. *Scientometrics*, *79*(3), 517–539. http://dx.doi.org/10.1007/s11192-007-2046-8

Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). Gender differences in research collaboration. *Journal of Informetrics*, *7*(4), 811–822. http://dx.doi.org/10.1016/j.joi.2013.07.002

Andriessen, K., Krysinska, K., & Stack, S. (2015). Predictors of article impact in suicidology: The bereavement literature, a research note. *Suicide and Life Threatening Behavior*, *45*(1), 18–24. http://dx.doi.org/10.1111/sltb.12106

Baccini, A., Barabesi, L., Cioni, M., & Pisani, C. (2014). Crossing the hurdle: The determinants of individual scientific performance. *Scientometrics*, *101*(3), 2035–2062. http://dx.doi.org/10.1007/s11192-014-1395-3

Barrios, M., Villarroya, A., & Borrego, A. (2013). Scientific production in psychology: A gender analysis. *Scientometrics*, *95*(1), 15–23. http://dx.doi.org/10.1007/s11192-012-0816-4

Bohm, M., Papoutsis, K., Gottwik, M., & Ukena, C. (2015). Publication performance of women compared to men in German cardiology. *International Journal of Cardiology*, *181*, 267–269. http://dx.doi.org/10.1016/j.ijcard.2014.12.018

Bordons, M., Morillo, F., Fernandez, M. T., & Gomez, I. (2003). One step further in the production of bibliometric indicators at the micro level: Differences by gender and professional category of scientists. *Scientometrics*, *57*(2), 159–173. http://dx.doi.org/10.1023/a:1024181400646

Borrego, A., Barrios, M., Villarroya, A., & Olle, C. (2010). Scientific output and impact of postdoctoral scientists: A gender perspective. *Scientometrics*, *83*(1), 93–101. http://dx.doi.org/10.1007/s11192-009-0025-y

Cikara, M., Rudman, L., & Fiske, S. (2012). Dearth by a thousand cuts?: Accounting for gender differences in top-ranked publication rates in social psychology. *Journal of Social Issues*, *68*(2), 263–285. http://dx.doi.org/10.1111/j.1540-4560.2012.01748.x

Cole, J. R., & Zuckerman, H. (1984). *The productivity puzzle: Persistence and change in patterns of publication of men and women scientists*. pp. 217–258. *Advances in motivation and achievement* (vol. 2).

Costas, R., & Bordons, M. (2011). Do age and professional rank influence the order of authorship in scientific publications? Some evidence from a micro-level perspective. *Scientometrics*, *88*(1), 145–161. http://dx.doi.org/10.1007/s11192-011-0368-z

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, *84*(1), 151–161.

Diekmann, A., Naf, M., & Schubiger, M. (2012). The impact of (Thyssen)-awarded articles in the Scientific Community. *Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie*, *64*(3), 563–581. http://dx.doi.org/10.1007/s11577-012-0175-4

Ding, W. W., Murray, F., & Stuart, T. E. (2006). Gender differences in patenting in the academic life sciences. *Science*, *313*(5787), 665–667. http://dx.doi.org/10.1126/science.1124832

Duch, J., Zeng, X. H. T., Sales-Pardo, M., Radicchi, F., Otis, S., Woodruff, T. K., et al. (2012). The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS ONE*, *7*(12), 11. http://dx.doi.org/10.1371/journal.pone.0051332

Eloy, J. A., Svider, P., Chandrasekhar, S. S., Husain, Q., Mauro, K. M., Setzen, M., et al. (2013). Gender disparities in scholarly productivity within academic otolaryngology departments. *Otolaryngology—Head and Neck Surgery*, *148*(2), 215–222. http://dx.doi.org/10.1177/0194599812466055

Eloy, J. A., Svider, P. F., Cherla, D. V., Diaz, L., Kovalerchik, O., Mauro, K. M., et al. (2013). Gender disparities in research productivity among 9952 academic physicians. *Laryngoscope*, *123*(8), 1865–1875. http://dx.doi.org/10.1002/lary.24039

Eloy, J. A., Svider, P. F., Setzen, M., Baredes, S., & Folbe, A. J. (2014). Does receiving an American Academy of Otolaryngology-Head and Neck Surgery Foundation centralized otolaryngology research efforts grant influence career path and scholarly impact among fellowship-trained rhinologists? *International Forum of Allergy & Rhinology*, *4*(1), 85–90. http://dx.doi.org/10.1002/alr.21224

Fox, M. F., & Stephan, P. E. (2001). Careers of young scientists: Preferences, prospects and realities by gender and field. *Social Studies of Science*, *31*(1), 109–122. http://dx.doi.org/10.1177/030631201031001006

Frandsen, T. F., & Nicolaisen, J. (2012). Effects of academic experience and prestige on researchers' citing behavior. *Journal of the American Society for Information Science and Technology*, *63*(1), 64–71. http://dx.doi.org/10.1002/asi.21666

Fridner, A., Norell, A., Akesson, G., Gustafsson Senden, M., Tevik Lovseth, L., & Schenck-Gustafsson, K. (2015). Possible reasons why female physicians publish fewer scientific articles than male physicians—A cross-sectional study. *BMC Medical Education*, *15*, 67. http://dx.doi.org/10.1186/s12909-015-0347-9

Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., et al. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, *76*(1), 169–185. http://dx.doi.org/10.1007/s11192-007-1892-8

Hopewell, S., & Clarke, M. (2003). How important is the size of a reprint order? *International Journal of Technology Assessment in Health Care*, *19*(4), 711–714 (Retrieved from <Go to ISI>: ⟨//WOS: 000220479700014⟩.

Hunter, L. A., & Leahey, E. (2010). Parenting and research productivity: New evidence and methods. *Social Studies of Science*, *40*(3), 433–451. http://dx.doi.org/10.1177/0306312709358472

Jespersen, S. T., Munch, J. R., & Skipper, L. (2008). Costs and benefits of Danish active labour market programmes. *Labour Economics*, *15*(5), 859–884. http://dx.doi.org/10.1016/j.labeco.2007.07.005

Kelchtermans, S., & Veugelers, R. (2013). Top research productivity and its persistence: Gender as a double-edged sword. *Review of Economics and Statistics*, *95*(1), 273–285. Retrieved from <Go to ISI>: ⟨/WOS: 000316153300019⟩.

Khan, N. R., Thompson, C. J., Taylor, D. R., Venable, G. T., Wham, R. M., Michael, L. M., et al. (2014). An analysis of publication productivity for 1225 academic neurosurgeons and 99 departments in the United States. *Journal of Neurosurgery*, *120*(3), 746–755. http://dx.doi.org/10.3171/2013.11.jns131708

Kretschmer, H., Pudovkin, A., & Stegmann, J. (2012). Research evaluation. Part II: Gender effects of evaluation: Are men more productive and more cited than women? *Scientometrics*, *93*(1), 17–30. http://dx.doi.org/10.1007/s11192-012-0658-0

Lariviere, V., Vignola-Gagne, E., Villeneuve, C., Gelinas, P., & Gingras, Y. (2011). Sex differences in research funding, productivity and impact: An analysis of Quebec university professors. *Scientometrics*, *87*(3), 483–498. http://dx.doi.org/10.1007/s11192-011-0369-y

Lechner, M., & Wiehler, S. (2011). Kids or courses? Gender differences in the effects of active labor market policies. *Journal of Population Economics*, *24*(3), 783–812. http://dx.doi.org/10.1007/s00148-009-0267-2

Lissoni, F., Mairesse, J., Montobbio, F., & Pezzoni, M. (2011). Scientific productivity and academic promotion: A study on French and Italian physicists. *Industrial and Corporate Change*, *20*(1), 253–294. http://dx.doi.org/10.1093/icc/dtq073

Long, J. S. (1992). Measures of sex-differences in scientific productivity. *Social Forces*, *71*(1), 159–178. http://dx.doi.org/10.2307/2579971

Long, J. S., & Fox, M. F. (1995). Scientific careers—Universalism and particularism. *Annual Review of Sociology*, *21*, 45–71. Retrieved from <Go to ISI>: ⟨//WOS:A1995RP07700004⟩.

Lopez, S. A., Svider, P. F., Misra, P., Bhagat, N., Langer, P. D., & Eloy, J. A. (2014). Gender differences in promotion and scholarly impact: An analysis of 1460 academic ophthalmologists. *Journal of Surgical Education*, *71*(6), 851–859. http://dx.doi.org/10.1016/j.jsurg.2014.03.015

Mairesse, J., & Pezzoni, M. (2013). Does gender affect scientific productivity. In *Paper presented at the: A critical review of the empirical evidence and a panel data econometric analysis for French physicists. Presented to AFSE Meeting* Aix en Provence.

Martinez, M., Lopez, S., & Beebe, K. (2015). Gender comparison of scholarly production in the Musculoskeletal Tumor Society using the hirsch index. *Journal of Surgical Education*, http://dx.doi.org/10.1016/j.jsurg.2015.06.020

Mauleon, E., Bordons, M., & Oppenheim, C. (2008). The effect of gender on research staff success in life sciences in the Spanish National Research Council. *Research Evaluation*, *17*(3), 213–225. http://dx.doi.org/10.3152/095820208×331676

Moed, H. F. (2006). . *Citation analysis in research evaluation* (vol. 9) Dordrecht: Springer Science & Business Media.

OECD. (2015). *Research and development statistics (RDS)*. Retrieved from ⟨http://www.oecd.org/science/inno/researchanddevelopmentstatisticsrds.htm⟩.

Paik, A. M., Mady, L. J., Villanueva, N. L., Goljo, E., Svider, P. F., Ciminello, F., et al. (2014). Research productivity and gender disparities: A look at academic plastic surgery. *Journal of Surgical Education*, *71*(4), 593–600. http://dx.doi.org/10.1016/j.jsurg.2014.01.010

Penas, C. S., & Willett, P. (2006). Gender differences in publication and citation counts in librarianship and information science research. *Journal of Information Science*, *32*(5), 480–485. http://dx.doi.org/10.1177/0165551506066058

Prpic, K. (2002). Gender and productivity differentials in science. *Scientometrics*, *55*(1), 27–58. http://dx.doi.org/10.1023/a:1016046819457

Slyder, J. B., Stein, B. R., Sams, B. S., Walker, D. M., Jacob Beale, B., Feldhaus, J. J., et al. (2011). Citation pattern and lifespan: A comparison of discipline, institution, and individual. *Scientometrics*, *89*(3), 955–966.

Sotudeh, H., & Khoshian, N. (2014). Gender differences in science: The case of scientific productivity in nano science & technology during 2005–2007. *Scientometrics*, *98*(1), 457–472. http://dx.doi.org/10.1007/s11192-013-1031-7

Stack, S. (2004). Gender, children and research productivity. *Research in Higher Education*, *45*(8), 891–920. http://dx.doi.org/10.1007/s11162-004-5953-z

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. http://dx.doi.org/10.1214/09-STS313

Sugimoto, C. R., Lariviere, V., Ni, C. Q., Gingras, Y., & Cronin, B. (2013). Global gender disparities in science. *Nature*, *504*(7479), 211–213. Retrieved from <Go to ISI>: ⟨//WOS:000328121500015⟩.

Tomei, K. L., Nahass, M. M., Husain, Q., Agarwal, N., Patel, S. K., Svider, P. F., et al. (2014). A gender-based comparison of academic rank and scholarly productivity in academic neurological surgery. *Journal of Clinical Neuroscience*, *21*(7), 1102–1105. http://dx.doi.org/10.1016/j.jocn.2013.11.006

van Arensbergen, P., van der Weijden, I., & van den Besselaar, P. (2012). Gender differences in scientific productivity: A persisting phenomenon? *Scientometrics*, *93*(3), 857–868. http://dx.doi.org/10.1007/s11192-012-0712-y

van Dijk, D., Manor, O., & Carey, L. B. (2014). Publication metrics and success on the academic job market. *Current Biology*, *24*(11), R516–R517. http://dx.doi.org/10.1016/j.cub.2014.04.039

West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The role of gender in scholarly authorship. *PLoS ONE*, *8*(7). Retrieved from <Go to ISI>: ⟨//WOS: 000322132100004⟩.

Xie, Y., & Shauman, K. A. (1998). Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review*, *63*(6), 847–870. http://dx.doi.org/10.2307/2657505