

# Identifying gender disparities in research performance: the importance of comparing apples with apples

Lynn P. Nygaard\*, Peace Research Institute Oslo (PRIO), Oslo, Norway, [lynn@prio.org](mailto:lynn@prio.org), ORCID: 0000-0003-3747-3074

Dag W. Aksnes, Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway, ORCID: 0000-0002-1519-195X

Fredrik Niclas Piro, Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo, Norway, ORCID: 0000-0002-8108-778X

\* corresponding author

## *Acknowledgements*

This research has been made possible through funding from the Research Council of Norway (grant number 295817).

## Conflict of Interest

The authors declare that they have no conflict of interest

## Introduction

Identifying factors associated with research performance and productivity has been the topic of considerable research in higher education (see, e.g., Fox & Nikivincze, 2020). Gender has often been singled out as a key factor, although the quest to uncover why men appear to produce more research output than women seems to have raised more questions than it answers. While men generally appear to produce more publications than women (see, e.g., Sugimoto et al., 2013), and score highest on most research performance indicators, the results can vary widely depending on the study in question (see., e.g., van Arensbergen et al., 2012).

One reason why results might vary relates to *what* exactly is being measured. Productivity is understood differently in different contexts, and often overlaps with broader concepts of performance. In its simplest sense, productivity refers to the number of academic outputs a

researcher produces. However, measuring productivity raises questions about what kind of academic outputs to include, how co-authorship should be counted, whether some outputs should count more than others, and whether citations should matter. The notion of what, exactly, constitutes a productive researcher and how research performance should be measured remains elusive, resulting in inconsistent modes of operationalization across the literature. Choices about operationalization matter because some publication practices appear to be gendered – for example, production of book chapters (Mayer & Rathmann, 2018) or co-authorship patterns (European Commission, 2019, p. 142).

A second reason relates to *who* is included in the study, i.e., the composition of the population. Men and women are distributed unevenly throughout the academy – both horizontally and vertically. For example, women comprise a distinct minority in most STEM (science, technology, engineering, and mathematics) fields in most countries, whereas they often constitute a majority in certain social sciences and humanities (SSH) fields. STEM fields have distinctly different publication patterns than the SSH fields, including more frequent co-authoring. Similarly, while women comprise the majority of doctoral students in many contexts, they constitute a distinct minority of full professors in almost all fields and countries (European Commission 2019, pp 118-123). Because doctoral students produce far fewer publications than professors, the decision to include doctoral candidates in productivity studies can shift productivity in favor of men simply because of the proportion of female doctoral students compared to professors.

Our inquiry examines whether gender differences in productivity (and inconsistent findings across studies) might simply reflect the uneven distribution of men and women throughout the academy, and differences in situated publication practices and the way these practices are measured. In other words, we are questioning whether previous studies are indeed observing essential differences between men and women, or whether they are comparing apples and oranges – while having an unclear notion of what fruit should taste like. Perhaps women do not publish more book chapters

than men (or co-author less) because they are women, but because they are more likely than men to be in fields that produce relatively more book chapters (or co-author less).

While this idea is not new (see, e.g., Nygaard & Bahgat, 2018), it has not been tested systematically across either different publication practices (the “what”) or different groups of researchers (the “who”). This study marks a significant advance on both fronts. With respect to the “what,” we step away from the problematic notion of productivity as a single indicator and instead look separately at a large set of individual practices related to academic publishing that are often considered in an evaluation context. With respect to the “who,” we examine the effect of slicing the aggregate population both vertically (by academic position and age) and horizontally (by scientific field and institutional affiliation).

Our aim is to discover whether we can see evidence of a composition effect, whereby non-gender-related attributes of a group can better explain gender differences in publishing practices than gender alone. Our specific research question is how horizontal and vertical disaggregation affects the gender gap compared to the aggregate level.

We answer our question through a bibliometric analysis of academic output from higher education institutions in Norway. Although Norway ranks highly on most indicators of gender equality, it still shows a lack of gender balance in institutions of higher education, particularly at the top of the academic hierarchy. The general demographic trend of unequal gender distribution across the academy, in combination with the publication patterns and practices associated with specific fields or academic positions, are comparable to what is seen in other countries (European Commission, 2019).

Perhaps the best reason for choosing Norway for our analysis, however, is its high-quality data on publication output. The Current Research Information System in Norway (Norwegian Science Index – NSI) database systematically collects bibliometric data on publications by researchers affiliated with public institutions in Norway. Data is collected on the type of publication (journal articles, book

chapters, and monographs); the level of publication (status of the journal or publisher); and authors and their affiliations. (See Sivertsen, 2018 for a full description of NSI). Although NSI imports data from Scopus, it also includes publications not captured by this database, including publications in non-English languages, book publication, and so on. The data is verified, quality assured and collected for *all* academic outputs published by *all* employees in the health sector, institute sector and higher education sector in Norway. The completeness of the dataset allows us to study an entire population rather than a sample, and its high quality makes it far more reliable than self-reported data and exceptionally well suited for bibliometric analysis.

In this study, we couple this detailed bibliometric data with data on individual characteristics of the persons from the Norwegian Research Personnel Register (NRPR), which covers all individuals in institutions with research and development activity in the public sector in Norway. Specifically, we combine detailed bibliometric data on all academic publications published by all Norwegian researchers in the higher education sector for the period 2015–2017 with data on gender, academic position, age, scientific field, and institutional affiliation. Before turning to details of how we analyze the data, we provide the conceptual framework that guides how we operationalize publication practices and producers of academic publishing in our study.

### The “what” and the “who”: Conceptualizing and operationalizing publication practices and producers of academic publishing in Norway

Our key theoretical assumptions stem from an Academic Literacies perspective that conceptualizes academic writing and publishing as a situated social practice, suggesting that the way academic writing is produced and the final form it takes depend on the context in which it is produced (see, e.g., Lillis & Scott 2007). Author identity – both in the sense of group membership (gender, race, class, as well as nationality, age, academic position, civil status, and so on) and beliefs about the self (assessment of skill, for example) – also plays a key role in the production of writing, not least in how

writing might be prioritized above other activities, confidence in one's own expertise, sense of ownership of the writing, and perception of agency (Ivanic, 1998). In essence, these assumptions mean that different corners of the academy will publish different kinds of things in different ways.

For example, while journal articles might be considered the gold standard of academic publication in most contexts, they are not produced the same way across all fields. Some fields rely on large laboratory-style forms of collaboration, with ambitious research projects bringing together hundreds of researchers from across the world and resulting in multiple publications. Other fields prefer a close examination of a single text by a single author. Moreover, some fields might value monographs and book chapters as much, if not more, than journal articles. And others might strive to make their research as relevant to the local setting as possible, which means a larger share of publications might be in a language other than English. The production of a journal article (or any other academic publication) also requires time for research and writing, as well as confidence in one's own ability to contribute to the academic discourse. Submitting to a high-prestige journal requires additional confidence – as does the choice to eschew the journal format altogether and aim for something that might matter deeply to the author or local context but not count in an evaluation. Time for research, the agency to make choices, and the expertise necessary to publish research vary considerably according to both academic position and the life course.

### The “what”: Publication practices

The way we have conceptualized the writing practices we examine in this study stem from a combination of which practices can be expected to differ across the academy, which practices are discussed in the productivity literature, and what kind of personal and bibliometric data we have access to. We focus on four different variable clusters, comprising twelve separate indicators:

***Publication volume:*** A fundamental way to conceptualize research productivity is through how many publications are generated by individual researchers over a certain period. In research environments that emphasize teams and collaborative writing (such as the STEM fields), researchers commonly

publish many papers but with a proportionately smaller number of author shares per publication (Lee & Bozeman, 2005). The converse is true for research environments that emphasize individual contributions (e.g., philosophy, anthropology, and history). Most studies measuring productivity solely in terms of publication volume, without taking into account co-authorship, show that men produce more than women (see, e.g., Beaudry & Larivière, 2016; Bendels et al., 2018; Sotudeh & Khoshian, 2014; Stack, 2002; Sugimoto et al., 2013). We examine output in numbers of both (1) *publications* (whole counts, regardless of type) and (2) *author shares*. The former treats all publication outputs equally (they all count as 1), while the latter gives an author their share of the publication according to their relative contribution (fractionalized publication counts). For example, if four researchers co-author a paper, they are each given 25% of the credit. Calculating author shares is commonly used to ensure greater equality between fields (see, e.g., Waltman & van Eck, 2015).

**Publication type:** In addition to journal articles, books and book chapters are also frequent outputs for most fields. However, they are also less frequently researched since most current bibliometric studies rely upon large databases like Scopus or Web of Science (WoS) where monographs and book chapters are less adequately covered (Aksnes & Sivertsen, 2019). Previous studies suggest that women lag behind men when it comes to article publishing and monographs (Puuska, 2010), but produce relatively more book chapters (Mayer & Rathmann, 2018). We look at the total volume of academic output in the 3-year period for each author and calculate the percentage of (3) *journal articles*, (4) *book chapters*, and (5) *monographs*. These three indicators add up to 100% of a researcher's total volume of peer-reviewed academic output.

**Authorship:** There is mixed evidence concerning gender differences in collaboration. While Uhly et al. (2017, p. 763) suggest that “women face greater challenges building informal and professional networks, which could impact their ability to ask or receive invitations to collaborate,” others find women more likely to collaborate than men (e.g., Fell & König, 2016). When it comes to

*international* collaboration in particular, most studies show men collaborating more extensively (Abramo et al., 2013; Uhly et al., 2017; Aksnes, Piro, & Rørstad, 2019; Sugimoto et al., 2013). While extensive collaboration may signal prestige in some fields, other fields view solo authorship as a hallmark of expertise. One recent Danish study found that women publish more single-authored articles (Nielsen, 2015) and argued that it could be disadvantageous (cf. similar findings in Zettler et al., 2017), but Kulczycki and Korytkowski (2020) note that solo-authored monographs are more likely to be produced by senior staff and highly productive researchers. We look at authorship in three different ways: (7) *authors per publication*, calculated as the mean average number of contributing authors for each publication in the researcher's total volume of publications; (8) *international collaboration*, calculated as the percentage of a researcher's publications with one or more co-authors affiliated with institutions in countries other than Norway; and (9) *solo authorship*, calculated as the percentage of publications where the researcher is the sole author.

**Impact or quality:** The final cluster concerns the assumed quality or impact of the research. Debates about research productivity and performance have often focused on how volume alone is insufficient; the quality or impact of the research – the extent to which the publication achieves high scores on indicators of excellence or uptake by the scientific community – should also matter. Quality and impact are harder to capture quantitatively than volume of output, but examining citations has been a common approach. The NSI database, however, relies on publication channel as a proxy for both quality and impact on the assumption that publication in prestigious journals is both an indicator of quality and will likely result in greater impact. Accredited journals and publishers with approved routines for peer review constitute “Level 1” channels; publication channels that represent the top 20% in their field are assigned “Level 2” status. The levels are determined, and revised annually, by national committees for different fields under the auspices of the higher education umbrella organization Universities Norway (UHR). We thus measure the percentage of (9) *Level 2 publications*.

We also add three additional indicators of impact or quality to this cluster to reflect the debates in the broader literature. First, we examine publication in English, on the assumption that English publication can be expected to lead to greater uptake in the wider scientific community than publication in Norwegian or other local languages, and that publication in English is often seen as a sign of prestige in non-Anglophone countries (Lillis and Curry, 2010). Some studies indicate that women might publish in local languages more frequently than men (e.g., Sugimoto 2013). We thus look at the percentage of (10) *English language publications*, distinguishing only between English and other languages rather than looking at Norwegian specifically. And finally, since examining citations has been a common approach in the literature, and many studies suggest gendered patterns of citation behavior (see, e.g., Nielsen, 2015; van den Besselaar & Sandström, 2016; Bendels et al., 2018), we include two complementary variables based on citations: (11) *Citation index (MNCS)* and (12) *Top10% cited papers* (Waltman et al., 2012). MNCS provides the average citations for all the researcher's output, from uncited works to top-cited works; TOP10% reflects the proportion of a researcher's publications with very high impact by measuring the proportion of a researcher's publications that, compared with other publications in the same field and in the same year, belong to the 10% frequently cited papers in the world.

Together, these twelve indicators, grouped into four clusters, provide different ways of measuring research productivity and performance and act as our dependent variables.

#### The "who": Disaggregating the population

Just as we understand academic publication to be a situated social practice, we also conceptualize gender as socially constructed, in the sense that gender carries social expectations for behavior, which vary from context to context (see, e.g., Witt 2011). To determine gender, we use the individual's legal gender recorded in the NRPR, which may or may not be the same as their biological sex or sex at birth.



We also understand that men and women play more than one social role, and that gender interacts with other aspects of identity. Here we focus on identity as an academic – discipline, area of expertise, and level of seniority – on the assumption that academic identity shapes an individual’s perceptions of what kinds of things should be produced and their agency in producing them. We focus on four different aspects of identity as our “compositional factors” that, in addition to gender, act as the independent variables in this study: *scientific field* and *academic position* act as our primary compositional variables (and the main focus of our analysis), while institutional affiliation and age act as secondary compositional variables and help us further isolate the effect of scientific field and academic position.

***Scientific field.*** Publication patterns vary significantly across fields (Piro, Aksnes, & Rørstad, 2013). In STEM fields, researchers are commonly listed as one of dozens of co-authors on multiple journal articles per year, whereas a more common expectation for a scholar in the SSH fields might be one or two articles per year – and any co-authoring might be limited to one or two other authors. And in STEM fields, PhD students frequently co-author with their supervisors, whereas in most SSH fields doctoral candidates carry out their work more independently, which affects the co-authoring patterns of both students and supervisors in these fields.

NSI categorizes all publications based on a classification system with 85 different fields. We merge fields that have strong similarities in publication patterns (publication volume, number of co-authors, and publication types), ending up with eight broad groups: Economics & Management; Engineering; Health Sciences (social medicine, nursing, psychology, etc.); Humanities; Mathematics & Informatics; Medicine (biomedicine and clinical medicine); Natural Sciences; and Social Sciences (political science, sociology, social anthropology, etc.). We classify each researcher by the field where they have the highest number of publications.

***Academic position.*** Just as not all academic fields have the same publication patterns, not all academic positions engender the same expectations for publication activity. The importance of

controlling for academic position is highlighted in several studies (Fox & Nikivincze, 2020; König et al., 2015; Rørstad & Aksnes, 2015; Nygaard & Bahgat, 2018). The NRPR operates with many different staff categories and academic positions, which we collapse into six main categories: Professor; Associate Professor; Postdoc/Researcher; PhD candidate; and Other. These categories represent a basic hierarchy with differing expectations for publication. Unlike many studies on productivity, we include doctoral candidates because in Norway most doctoral candidates write a PhD by publication and can be expected to have at least one article published during their doctoral candidacy. The category of “other” constitutes staff who are not hired to conduct research and comprises only 3.8% of the full study population.

***Institutional affiliation:*** The institutional context in which research is conducted and reported on also matters. Publicly funded research in Norway is carried out in three main sectors: the health sector (by practicing medical professionals); the research institute sector (specializing in applied research with a high degree of relevance to Norwegian society); and the higher education sector. Although NSI gathers data on all three sectors, we limit this study to only the higher education sector to ensure that our findings are as comparable as possible to other studies. However, even within this one sector we can expect differences related to time allotted to research (compared to teaching) and the academic profile of the institution (including composition of fields). We thus divide the institutions from the higher education sector into four groups: university colleges, new universities (that until recently were university colleges), specialized universities (e.g., the Norwegian School of Sport Sciences), and traditional universities (the four largest universities in Norway).

***Age.*** Like academic position, age can also say something about degree of expertise. However, biological age is not always correlated with “academic age”, as (particularly in some fields) some mature professionals return to the university as doctoral students for further education. In addition to expertise, age also reflects the life course, and we could expect different publishing behaviour for

someone with young children compared to someone close to retirement. While many studies have shown a curvilinear relationship (peaking and then declining), Rørstad and Aksnes (2015) suggest that for women productivity increases with age. We operationalize age by using the individual's age in 2015. Although age could be expected to strongly correlate with academic position, the correlation between these variables in this study is only .413 (Pearson's  $r$ , two-tailed, significant at the 0.01 level), and thus does not cause any issues of multicollinearity.

## Methods

We first establish a baseline by identifying the gender differences in each of the publication practices we observe (our dependent variables) at the aggregate level – where all women (regardless of where they are situated in the academy) are compared to all men. As an intermediate step, we conducted a bivariate analysis using scientific field and academic position as the two compositional variables; the results of this analysis are presented online as supplementary information. Since the purpose of our study is to explore whether gender variations across a wide spectrum of variables subject to an ecological fallacy, we use a straightforward ordinary least square (OLS) regression analysis to assess the extent to which each of the compositional variables can explain the overall gender differences on the dependent variables. This statistical approach allows us to observe the overall impact of horizontal and vertical disaggregation (i.e., whether disaggregation increases or decreases the gender gap relative to the aggregate level) across the large number of dependent variables. Although OLS does not allow us to estimate the exact sizes or strengths of these relationships, our purpose is merely to examine whether measurements of gender differences could be sensitive to how productivity is operationalized and how the sample population is disaggregated. The unit of analysis is the individual researcher. The sample we analyze comprises all individuals registered in NRPR in the higher education sector for the years 2015–2017. Limiting the population to only those in the higher education sector leaves us with 17,878 individuals. The institutions that

these researchers represent comprise 8 general universities, 7 public university colleges, 3 private university colleges and 3 public specialized universities.

All publication variables refer to the individual's total publication output registered in the NSI database in the period 2015–2017. This interval is chosen to meaningfully analyze citations during the same period, as citations are counted from the time of publication up to and including 2019, normalized by field and publication year. The citation analysis is limited to WoS indexed publications only, as NSI does not collect data on citations. We use the academic positions and institutional affiliations held by the individual in 2015 for the entire study interval, regardless of whether they have changed affiliation or position in later years. Our assumption is that the time-lag between research and publishing makes it reasonable to expect that, for most researchers, their position and affiliation in 2015 will correspond with publication output in subsequent years. A three-year publication interval was chosen to account for random annual fluctuations in the publication numbers (Abramo et al., 2012).

Table 1 shows the distribution of researchers in our sample by the primary compositional factors and illustrates the unequal distribution of gender across fields and academic positions. Except for the Humanities (where the gender distribution is almost equal), all fields are notably either male- or female-dominated. Women have a higher representation in Health Sciences, Medicine and Social Sciences. Men have a higher representation in Economics & Management, Engineering, Mathematics & Informatics, and Natural Sciences. We also observe that women are strongly underrepresented at the full professor level (both in absolute numbers and in percentages), whilst being far better represented at the lower levels of academia, foremost among PhD candidates and in the diverse group "Other".

[Insert Table 1 about here]

## Results

The first step of our analysis establishes the baseline figures: gender differences in publication practices without taking into account any of the compositional variables. Table 2 provides an overview of the average values for all individuals included in the analysis, averages for men and women separately, and gender parity ratios where women's mean scores are compared to men's mean scores. We use ratios to enable comparison between variables that are otherwise not comparable because they use different units of measurement. A ratio of 1.00 represents gender parity: A ratio below 1.00 means that women score lower than men, and a ratio above 1.00 means that women score higher. The more the ratio differs from 1.00 (either higher or lower), the bigger the gender gap. Note that we focus on observing how this gap changes after introducing compositional variables rather than attempting to estimate the exact size of each relationship. The aggregate figures reflect the findings in much of the literature: women publish less, produce relatively more book chapters, have a fewer number of co-authors, publish less in English, and publish less in high-prestige channels. Surprisingly, however, there seems to be only a negligible difference in citations.

[Insert Table 2 about here]

Our next step, the multivariate regression analysis looks at each of the 12 publication practices separately – first starting with the aggregate level (“gender-alone” model), then by examining each of the four compositional variables one at a time, and finally by adding all compositional variables at once in what we call the “full model”. Table 3 thus displays gender regression estimates from a total

of 72 regression models. For readability purposes, we include the regression estimates for each individual scientific field, academic position, or institution only for the full model (Table 4).

[Insert Table 3 about here]

[Insert Table 4 about here]

**Publication volume:** For number of publications, academic position alone reduces the gender coefficient by 41% (Std. Beta values from .092 ( $p < .000$ ) to .054 ( $p < .000$ ); all remaining Std. Beta values referred to in the following text have  $p$ -values  $< .000$ , unless reported otherwise). In comparison, scientific field only explains 10% (Std. Beta values from .092 to .083). Age and institutional affiliation account for 1% and 4%, respectively. In total, compositional variables explain 58% of the gender coefficient, as the gender Beta value is reduced from .092 to .038. This means that there is still a gender gap in productivity, but it is much smaller compared to the aggregate figure. Academic position explains most of the gender differences: productivity increases moving up the hierarchy and there are relatively few female full professors (cf. Table 1). For author shares, academic position alone reduces the gender coefficient by 57%, and age by 4%. Although scientific field and institution amplify the gender differences, the compositional variables in sum reduce the gender coefficient by 62%.

**Publication type:** Journal articles constitute the bulk of academic output for both genders, with no statistically significant gender differences in either the gender-alone model or in the full model. For monographs, we observe a change from a non-significant difference at the aggregate level to significant after entering scientific field as control variable. This is perhaps because most fields produce very few monographs, whereas in the humanities, enough monographs are produced to create a significant set of observations, and within this field, men are more likely to produce this

publication type. For book chapters, an aggregate gender imbalance (women producing more chapters, Std. Beta coefficient  $-.017$ ,  $p < .05$ ) becomes non-significant in the full model because women are more strongly represented in fields where book chapters are more common. Entering the compositional variables one by one to the model does not substantially reduce the gender gap, but in the full model we see how most compositional variables are statistically significant, thus *in combination* they remove all the gender difference.

**Authorship:** For *number of authors*, the small but positive significant gender coefficient is marginally affected by compositional factors except when controlling for institutional affiliation, which explains all significant gender difference. Table 4 shows that it is foremost affiliation with one of the old, traditional universities that influences the number of authors. For *international collaboration*, each of the compositional variables – except for age – reduce the gender coefficient: academic position by 29% (Std. Beta coefficient from  $.077$  to  $.055$ ), scientific field by 47%, and institutional affiliation by 18%. Age slightly increases the gender coefficient. In sum, the compositional variables explain 73% of the gender differences in international collaboration. The relative importance of academic field probably results from international collaboration being much more frequent in STEM fields than in non-STEM fields. Solo authorship is more puzzling: All compositional factors (except institutional affiliation) increase the size of the gender coefficient. At the aggregate level, women appear to produce more solo works than men (Std. Beta coefficient  $-.019$ ,  $p < .05$ ), and adding academic position (Std. Beta coefficient  $-.029$ ) and age (Std. Beta coefficient  $-.023$ ) strengthens this relationship. However, accounting for scientific field has the opposite effect and suggests that men are far more solo-oriented than women. Consequently, in the full model we end with a positive regression estimate of  $.025$  (in men's favor) compared to the baseline number  $-.019$  (i.e., in favor of women). Interestingly, adding institution alone eliminates all evidence of a statistically significant gender coefficient for solo authorship. We speculate on the reason for this puzzling finding in the discussion below.

**Impact or quality:** For *publication in Level 2 journals*, the small (but statistically significant) gender coefficient (Std. Beta coefficient .030) at the aggregate level is no longer statistically significant when academic position is accounted for and disappears almost entirely in the full model (Table 3). This is particularly interesting in light of Table 4, which shows that each individual cell (except for Economics) is significantly associated with Level 2 publication. For *English language publication*, academic position and institutional affiliation reduce the gender coefficient somewhat, while scientific field has a major influence on the gender coefficient, reducing it by 68% (Std. Beta coefficient from .112 to .036). In sum the compositional variables explain 79% of the gender gap (Std. Beta coefficient from .112 to .026). For the *citation variables*, the lack of significant gender differences that appear at the aggregate level remain unchanged when composition variables are added – apart from scientific field. When *only* scientific field is taken into account, a slight gender difference appears that favors men. This could be a result of the medical/health fields, which show a moderate gender difference in citations and are both relatively large fields with strong presence of women. Notably, this difference disappears when all other compositional variables are accounted for.

## Discussion

Table 5 summarizes the results from the regression analysis. ~~Four~~ Three of the variables (percentage of articles, authors per publication, and the two citations variables) showed no gender differences to begin with and remain unchanged after controlling for compositional variables. In four of the variables – namely, publications, author shares, international collaboration, and publication in English – a gender gap remained, although it was strongly reduced after controlling for the compositional variables. Additional compositional variables could perhaps explain even more of the remaining gender difference.

[Insert Table 5 about here]



Three such avenues for further analysis present themselves: First, a more fine-grained disaggregation of scientific field could be undertaken. Political science and education, for example, are both social sciences with different gender profiles (with a larger relative share of women in education) and different publication practices (with, for example, less international collaboration and less publication in English in the field of education). Second, if we want to focus on the “typical” researcher, we could filter out the most prolific researchers by taking into account “productivity tier”. Individual productivity is highly skewed, with very few academics producing the bulk of publications. Dividing researchers into three groups – prolific, regular, and sporadic, for example – might provide a more meaningful comparison of “typical” men and women because the most highly prolific researchers tend to be men, and their exceptionally high scores have a significant impact on the average for all men (Abramo et al., 2021). And third, we could consider leaves of absence and part-time positions. In bibliometric studies, productivity over three years presupposes that the individual was employed full time, without leave, during that period. Since women are more likely to work part time and take longer parental leave, their publication output could be expected to be lower than men for that reason alone. Calculating “total months worked” rather than calendar years for the chosen period might reduce the remaining gender gap in publication output significantly.

Table 5 also indicates that two variables did not show a clear pattern: monographs and solo authorship – both of which revealed greater gender disparities than were visible at the aggregate level. For solo authorship, women’s relatively higher share at the aggregate level turns into a relatively lower share within each scientific field. This might be explained by Simpson’s paradox (Pearl, 2014), which in our case relates to the different gender composition of researchers across fields. Since men and women are differently distributed in each subgroup – with women better represented in those with high values – they end up with a higher overall score than men. When it comes to monographs, we see an example of how a non-significant difference at the aggregate level

becomes significant through disaggregation because monographs are disproportionately produced in only one field – the humanities. Consequently, in that field, the gender regression coefficient changes from non-significant to significant.

Both these anomalies are in line with our premise that the production of academic publications is highly situated, but further suggest that gender might interact with these different environments in different ways (Acker, 2006). The way that gendered social norms interact with the specific demands of a scientific field or academic position – and perhaps the institutional history of a particular environment – may create “hot spots” that need greater attention. For example, both solo authorship and the production of monographs are not only considered more prestigious in some fields than others (e.g., the humanities), but also require a considerable amount of “alone time,” which may be more difficult for academic positions that come with high demands for teaching or supervision (especially in a field where doctoral candidates do not work with their supervisors on a common project). These challenges may be exacerbated in a context where gendered social norms encourage women to take on a larger share of the “academic housekeeping” or make it difficult for women to spend time writing after work hours (Aiston & Jung, 2015; Seierstad & Healy, 2012). These factors might all come together in a specific institutional environment that may also have a weak history of implementing measures to improve gender balance. And across different countries, gender-role expectations might operate differently with respect to self-selection into fields and possibilities for promotion (Correll, 2004).

While the strength of our simple statistical approach (comparing means and OLS regression analyses) is that we can examine multiple independent and dependent variables simultaneously, a clear limitation is that we cannot capture more fine-tuned interactions between gender, the compositional factors, and publication indicators. Structural equation modeling (SEM) would better isolate each independent variable’s influence (and how they work together) on publishing indicators, and Blinder-Oaxaca decomposition (Elder, Goddeeris & Haider, 2010) would more accurately

partition the influence of gender and other compositional variables, respectively. Such approaches would be useful for more in-depth studies of how gender interacts with the dependent variables, but preferably for *one* such indicator at the time.

A second limitation of our study is that along both the “who” and the “what” dimensions we were limited to data that is systematically gathered by NSI and the NRPR databases. We could not, for example, take a more intersectional approach to defining author identity by also considering race, class, nationality, disability, or any other number of important identity markers because this data was not available to us. Similarly, we could not consider the production of non-scientific output, which is becoming increasingly important as academics are required to demonstrate greater social accountability and relevance.

## Conclusion

We set out to examine how the “what” (which publication practices are measured) and the “who” (how the sample is disaggregated) matter in assessing apparent gender differences in research productivity and performance. We looked closely at four clusters of publication practices related to *publication volume, publication type, authorship, and impact or quality* (12 indicators in total), and examined how controlling for scientific field, institutional affiliation, academic position, and age would change the apparent differences between men and women.

Our findings suggest that for most publication practices, *most* gender variations at the aggregate level disappear or are strongly reduced after disaggregation – and including more than one compositional variable reduces the gender gap more than one on its own. In other words, the more we could compare apples with other apples, the smaller the gender differences became. Political scientists will write and publish like other political scientists – and unlike organic chemists; professors will write and publish like other professors – and unlike doctoral candidates. Of course, gender still matters – after all, it is as a woman or a man in a gendered society that one chooses a

discipline, is promoted through the ranks, applies for a position in an institute, and moves through the life course (Witt, 2011). Thus, when it comes to quantitative measures of research performance and productivity, our study shows that gender seems to matter less than the compositional factors – but it is not entirely negligible.

While we do not set out to examine whether or why women might *actually* be less productive than men, only how gender gaps might change depending on the analysis performed, we are aware that there are very real obstacles to women's productivity and possibilities for promotion both in Norway and other countries. In that respect, our findings strongly suggest that it would be prudent to avoid using aggregate numbers as a basis for designing interventions. Because gender inequalities are not the same across different publication practices or different groups of academics, our findings suggest that gender equality measures should be tailored to the specific context, not one-size-fits-all. A second implication of our findings is that rather than targeting women's productivity per se, gender inequality measures could focus on reducing demographic differences by recruiting more women into STEM fields, addressing gendered obstacles to professorship, and so on.

In conclusion, we argue that aggregate figures about gender differences in research productivity or performance can prove misleading because they compare apples and oranges. In our study, not only do the aggregate figures make the gender gap appear larger than it is for most publication practices, but in some contexts, they also obscure gender inequalities that are bigger, or more complex, than the aggregate figures suggest. While the ratios and coefficients in our study might be specific to Norway, we expect that the general problematic relationship between aggregate and disaggregate figures will be generalizable. Acknowledging the situated nature of academic publishing can help shift the discourse about gender and research productivity away from essentialist understandings of gender to more nuanced discourses about how other aspects of identity – such as disciplinary belonging and degree of expertise – interact with gender to shape everyday publication practices in the academy.

## References

- Abramo, G., D'Angelo, C. A., & Cicero, T. (2012). What is the appropriate length of publication period over which to assess research performance? *Scientometrics* 93 (3), 1005–1017.
- Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). Gender differences in research collaboration. *Journal of Informetrics*, 7(4), 811–822.
- Abramo, G., Aksnes, D. W., & d'Angelo, A. C. (2021). Gender differences in research performance within and between countries: Italy vs Norway. *Journal of Informetrics*, 15(2), 101–144.
- Acker, J. (2006). Inequality regimes. Gender, class, and race in organizations. *Gender and Society*, 20(4), 441–464.
- Aiston, S. J. & Jung, J. (2015). Women academics and research productivity: An international comparison. *Gender and Education*. 27(3), 205-220. <https://doi:10.1080/09540253.2015.1024617>
- Aksnes, D. W., Piro, F. N. & Rørstad, K. (2019). Gender gaps in international research collaboration: a bibliometric approach. *Scientometrics*, 120, 747–774.
- Aksnes, D. W. & Sivertsen, G. (2019). A criteria-based assessment of the coverage of Scopus and Web of Science. *Journal of Data and Information Science*, 4(1), 1–21.
- Beaudry, C. & Larivière, V. (2016). Which gender gap? Factors affecting researchers' scientific impact in science and medicine. *Research Policy*, 45(9), 1790–1817.
- Bendels, M. H. K., Müller, R., Brueggmann, D., et al. (2018). Gender disparities in high-quality research revealed by Nature Index journals. *PLOS ONE*, 13(1), e0189136.
- Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review*. 69(1), 93–113.
- Elder, T.J., Goddeeris, J.H., & Haider, S.J. (2010). Unexplained gaps and Oaxaca-Blinder decompositions. *Labour Economics*, 17(1), 284-290.

European Commission (2019). *She Figures 2018*: Publications Office, Luxembourg, available at:  
[https://ec.europa.eu/info/publications/she-figures-2018\\_en](https://ec.europa.eu/info/publications/she-figures-2018_en).

Fell, C. B. & König, C. J. (2016). Is there a gender difference in scientific collaboration? A scientometric examination of co-authorships among industrial-organizational psychologists, *Scientometrics*, 108, 113–141.

Fox, M. F. & Nikivincze, I. (2020). Being highly prolific in academic science: Characteristics of individuals and their departments. *Higher Education*. Advance online publication.  
<https://doi.org/10.1007/s10734-020-00609-z>

Ivanic, R. (1998). *Writing and Identity: The Discoursal Construction of Identity in Academic Writing*. Amsterdam: John Benjamins Publishing Company.

König, C. J., Fell, C. B., Kellnhöfer, L., et al. (2015). Are there gender differences among researchers from industrial/organizational psychology? *Scientometrics*, 105(3), 1931–1952.

Kulczycki, E. & Korytkowski, P. (2020). Researchers publishing monographs are more productive and local-oriented. *Scientometrics*, 125, 1371–1387.

Lee, S. & Bozeman, B. (2005). The impact of research collaboration on scientific productivity, *Social Studies of Science*, 35(5), 673–702.

Lillis, T., & Curry, M. J. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. London: Routledge.

Lillis, T., & Scott, M. (2007). Defining academic literacies research: Issues of epistemology, ideology and strategy. *Journal of Applied Linguistics*, 4(1), 5-32.

Mayer, S. J. & Rathmann, J. M. K. (2018). How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. *Scientometrics*, 117, 1663–1693.

Nielsen, M. W. (2015). Gender inequality and research performance: Moving beyond individual-meritocratic explanations of academic advancement. *Studies in Higher Education*, 4(11), 2044–2060.

Nygaard, L. P. (2017). Publishing and perishing: An academic literacies framework for investigating research productivity. *Studies in Higher Education*, 42(3), 519–532.

Nygaard, L. P. & Bahgat, K. (2018). What's in a number? How (and why) measuring research productivity in different ways changes the gender gap. *Journal of English for Academic Purposes*, 32, 67–79.

Pearl, J. (2014). Comment: Understanding Simpson's paradox. *American Statistician*, 68(1), 8–13.

Petr, M., Engels, T. C. E., Kulczycki, E., Duskova, M., Guns, R., Sieberova, M., and Sivertsen, G. (2021). Journal article publishing in the social sciences and humanities: A comparison of Web of Science coverage for five European countries. *Plos One*, 16(4).

Piro, F. N., Aksnes, D. W. & Rørstad, K. (2013). A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *Journal of the American Society for Information Science and Technology (JASIST)*, 64(2), 307–20.

Puuska, H. M. (2010). Effects of scholar's gender and professional position on publishing productivity in different publication types. Analysis of a Finnish university. *Scientometrics*, 82(2), 419–437.

<https://doi.org/10.1007/s11192-009-0037-7>

Sivertsen, G. (2018). The Norwegian Model in Norway. *Journal of Data and Information Science*, 3(4), 3–19.

Rørstad, K. & Aksnes, D. W. (2015). Publication rate expressed by age, gender and academic position – A large scale analysis of Norwegian academic staff. *Journal of Informetrics*, 9(2), 317–333.

Seierstad, C., & Healy, G. (2012). Women's equality in the Scandinavian academy: A distant dream? *Work, Employment, & Society*, 26(2), 296–313.

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.

Stack, S. (2002). Gender and scholarly productivity: 1970–2000. *Sociological Focus*, 35(3), 285–296.

Sugimoto, C. R., Lariviere, V., Ni, C. Q., Gingras, Y., & Cronin, B. (2013). Global gender disparities in science. *Nature*, 504(7479), 211–213.

Uhly, K. M., Visser, L. M., & Zippel, K. S. (2017). Gendered patterns in international research collaborations in academia. *Studies in Higher Education*, 42(4), 760–782.

Van Arensbergen, P., Weijden, I. v. d., & Besselaar, P. v. d. (2012). Gender differences in scientific productivity: A persisting phenomenon? *Scientometrics*, 93(3), 857–868.

Van den Besselaar, P. & Sandström, U. (2016). Gender differences in research performance and its impact on careers: A longitudinal case study. *Scientometrics*, 106, 143–162.

Waltman, L., Calero -Medina, C., Kosten, J., et al. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419–2432.

Waltman, L. and van Eck, E.J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9(4), 872–894.

Zettler, H. R., Cardwell, S. M. & Craig, J. M. (2017). The gendering effects of co-authorship in criminology & criminal justice research. *Criminal Justice Studies*, 30(1), 30–44.



**Table 1: Distribution of researchers by the primary compositional variables**

<b>Scientific field</b>	<b>Women (N)</b>	<b>% of Total</b>	<b>Men (N)</b>	<b>% of Total</b>	<b>Sum (N)</b>	<b>% Women</b>
Economics & Management	271	3.39	617	6.24	888	30.52
Engineering	580	7.25	2038	20.63	2618	22.15
Health Sciences	1132	14.15	636	6.44	1768	64.03
Humanities	1010	12.63	1131	11.45	2141	47.17
Mathematics & Informatics	136	1.70	496	5.02	632	21.52
Medicine	1713	21.42	1356	13.72	3069	55.82
Natural Sciences	1252	15.65	2144	21.70	3396	36.87
Social Sciences	1904	23.81	1462	14.80	3366	56.57
<b>Total</b>	<b>7998</b>	<b>100.00</b>	<b>9880</b>	<b>100.00</b>	<b>17878</b>	<b>44.74</b>

  

<b>Academic position</b>	<b>Women (N)</b>	<b>% of Total</b>	<b>Men (N)</b>	<b>% of Total</b>	<b>Sum (N)</b>	<b>% Women</b>
Professor	969	12.12	2597	26.29	3566	27.17
Associate professor	1444	18.05	1616	16.36	3060	47.19
Postdoc/Researcher	1386	17.33	1871	18.94	3257	42.55
PhD candidate	2161	27.02	2307	23.35	4468	48.37
Other	2038	25.49	1489	15.07	3527	57.78
<b>Total</b>	<b>7998</b>	<b>100.00</b>	<b>9880</b>	<b>100</b>	<b>17878</b>	<b>44.74</b>

**Table 2: Aggregate-level publication characteristics**

	Publication volume		Publication type			Authorship			Impact or quality			
	Publications	Author shares	Articles (%)	Chapters (%)	Monographs (%)	Authors per publication	International collab. (%)	Solo authorship	Level 2 (%)	English (%)	MNCS	Top 10%
Mean for full sample	5.69	1.74	81.76	16.90	1.34	5.17	34.43	18.29	23.08	83.65	1.25	10.83
Women, mean	4.29	1.43	81.26	17.49	1.26	4.97	31.12	19.01	22.01	79.60	1.24	10.71
Men, mean	6.81	1.99	82.16	16.42	1.41	5.34	37.11	17.71	23.94	86.92	1.25	10.92
Ratio w:m	0.63	0.72	0.99	1.06	0.89	0.93	0.84	1.07	0.92	0.92	0.98	0.98

**Table 3: Regression analysis: Gender and 12 publication indicators when controlling for the different compositional variables. (Standardized Beta coefficients)**

	Publication vol.		Publication type			Authorship			Impact or quality			
	Publications	Author shares	Articles	Chapters	Monographs	Authors per publication	International collaboration	Solo authorship	Level 2	English	MNCs	Top 10%
<b>Gender alone</b>	.092***	.125***	.014	-.017*	.009	.019*	.077***	-.019*	.030***	.112***	.004	.005
<b>Gender + academic position</b>	.054***	.053***	.015*	-.015*	.000	.016*	.055***	-.029***	.012	.095***	-.002	.000
<b>Gender + scientific field</b>	.083***	.128***	-.009	.001	.026**	.017*	.041***	.032***	.021*	.036***	.022*	.020*
<b>Gender + age</b>	.091***	.120***	.018*	-.021**	.007	.020**	.079**	-.023**	.033***	.118***	.005	.006
<b>Gender + institution</b>	.088***	.126***	.000	-.003	.010	.063	-.014	.010	.014*	.094***	.001	.000
<b>Full model: Gender + all compositional variables<sup>1</sup></b>	.038***	.048***	-.006	.001	.016*	.021**	.025***	.012	.004	.024***	.016	.014
<b>Adjusted R square (full model)</b>	.071	.233	.276	.242	.045	.102	.231	.430	.061	.330	.011	.017

\*\*\* Sig. <0.001 level, \*\* Sig. <0.01 level, \* Sig. <0.05. <sup>1</sup> Estimates for compositional variables are shown in Table 6.

**Table 4: Regression analysis: All independent and 12 dependent variables (full model), individually. (Standardized Beta coefficients)**

	Publication vol.		Publication type			Authorship			Impact or quality			
	Publications	Author shares	Articles	Chapters	Monographs	Authors per publication	International collaboration	Solo authorship	Level 2	English	MNCS	Top 10%
<b>Scientific field</b>												
Econ & Management	-.092***	.033***	-.101***	.099***	.018*	-.144***	-.138***	.096***	-.010	-.088***	-.019	-.022*
Engineering	-.047***	.151***	-.160***	.166***	.000	-.190***	-.245***	.021**	-.083***	.001	-.054***	-.074***
Health Sciences	-.061***	.084***	-.036***	.037***	.001	-.140***	-.189***	.053***	-.030**	-.105***	.013	.004
Humanities	-.144***	.189***	-.402***	.363***	.183***	-.252***	-.384***	.610***	-.019*	-.354***	.012	-.020*
Math & Informatics	-.059***	.047***	-.070***	.072***	.003	-.119***	-.072***	.058***	-.020*	-.011	-.018*	-.032***
Medicine	-.064***	-.015	.019*	-.020*	.001	-.022*	-.106***	-.007	-.033***	-.008	.041***	.030**
Social Sciences	-.137***	.187***	-.410***	.393***	.112***	-.271***	-.415***	.412***	-.045***	-.396***	.017	-.018
<b>Academic position</b>												
Professor	.271***	.467***	-.012	.003	.030*	.053***	.154***	-.027**	.107***	.073***	.050**	.073***
Associate Professor	.118***	.202***	-.017	.007	.033**	.026**	.076***	.004	.038***	.024*	.007	.027*
Postdoc/ Researcher	.076***	.113***	.015	-.020*	.014	.040***	.148***	-.012	.082***	.025**	.051***	.083***
Other	.040***	.034***	-.031***	.036***	-.012	.037***	.022*	-.052***	-.022*	-.088***	-.006	.014
<b>Age</b>												
	-.033**	-.052***	-.069***	.065***	.023*	-.036***	-.094***	.060***	-.132***	-.143***	-.050***	-.075***
<b>Institution</b>												
University college	.013	-.029***	-.012	.008	.015	.028**	-.010	-.002	-.049***	-.073***	-.021	-.014
Specialized university	.000	-.006	.021**	-.030***	.025**	.012	.015*	.020**	.063***	-.005	.019	.021*
Traditional university	.028**	.002	.046***	-.050***	.009	.053***	.043***	.028***	.104***	.041***	.012	.024*

\*\*\* Sig. <0.001 level, \*\* Sig. <0.01 level, \* Sig. <0.05. Reference value for scientific field is Natural sciences. Reference value for academic position is PhD candidates. Reference value for institution type is new universities.

**Table 5: Summary of regression results**

	Publication volume		Publication type			Authorship			Impact or quality			
<i>After controlling for compositional variables</i>	Publications	Author shares	Articles	Chapters	Monographs	Authors per publication	International collaboration	Solo authorship	Level 2	English	MNCS	Top 10%
Little or no statistical gender difference to begin with, unchanged												
Gender difference strongly reduced												
Gender difference eliminated												
Unclear pattern												