Professor Marek Kwiek
Center for Public Policy Studies, Director
UNESCO Chair in Institutional Research and Higher Education Policy
University of Poznan, Poland
kwiekm@amu.edu.pl

Dr. Wojciech Roszka
Poznan University of Business and Economics, Poznan, Poland
wojciech.roszka@ue.poznan.pl

Gender Disparities in International Research Collaboration: A Large-scale Bibliometric Study of 25,000 University Professors

Abstract

In this research, we examine the hypothesis that gender disparities in international research collaboration differ by collaboration intensity, academic position, age, and academic discipline. The following are the major findings: (1) while female scientists exhibit a higher rate of general, national, and institutional collaboration, male scientists exhibit a higher rate of international collaboration, a finding critically important in explaining gender disparities in impact, productivity, and access to large grants. (2) An aggregated picture of gender disparities hides a more nuanced cross-disciplinary picture of them. (3) An analysis of international research collaboration at three separate intensity levels (low, medium, and high) reveals that male scientists dominate in international collaboration at each level. However, at each level, there are specific disciplines in which females collaborate more than males. Further (4), gender disparities are clearly linked with age. Until about the age of 40, they are marginal and then they begin to grow. Finally, we estimate the odds of being involved in international research collaboration using an analytical linear logistic model. The examined sample includes 25,463 internationally productive Polish university professors from 85 universities, grouped into 27 disciplines, who authored 159,943 Scopus-indexed articles.

1. Introduction

Gender is of significance for research performance and citation impact in science, both for national populations of scientists and national research top performers (Abramo, D'Angelo, and DiCosta, 2019; Larivière, Vignola-Gagné, Villeneuve, Gelinas, and Gingras, 2011). The relative rate of collaborative research has increased over time and it currently dominates solo authorships in all fields, except for the humanities (Wuchty, Jones, and Uzzi, 2007). International research collaboration is a hot topic in policy studies on increasingly globalized and networked (Wagner, 2018) science. Further, females in science (and tackling their discrimination through various equality strategies)—with significant policy implications—is another hot policy topic (Zippel, 2017) that makes gender disparities in international research collaboration an ultra-hot theme, particularly if large-scale data encompassing entire national systems are utilized. However, as Abramo, D'Angelo and Murgia (2013) note, the debate on gender aspects in research systems has focused primarily on the overrepresentation of male academics, the productivity gap, and gender discrimination, and only rarely on collaboration patterns. In this research, we examine the hypothesis that gender disparities in collaboration patterns in science differ by collaboration intensity, academic position, age, and discipline, thereby reflecting wider gender disparities in all collaboration types (general, international, institutional, and national). In addition, we estimate the odds of being involved in international research collaboration using an analytical linear logistic model.

An integrated dataset of all Polish scientists with their administrative, biographical, publication, and citation data is used in this study ("The Polish Science Observatory" data set maintained by the authors includes 99,535 scientists and 377,886 Scopus-indexed articles published in the decade 2009–2018). The sample examined in this paper comprises 25,463 internationally productive Polish university professors from 85 universities, grouped into 27 disciplines who authored 159,943 Scopus-indexed articles.

The paper is structured in the following manner. The next section provides a literature review, followed by data and methods. The results section presents discussions on gender disparities in international collaboration from the perspectives of collaboration intensity, academic disciplines, age, and academic positions and on the results of the linear logistic model. The last section ends the paper with a summary of the findings, discussion, and conclusions.

2. International Research Collaboration and Gender: Literature Review

2.1. The Homophily Principle: Gender Similarity Breeding Connection Among Scientists

The general picture in which there are "two scientific worlds, one male, the other female" is highly relevant in any discussion of research collaboration (Etzkowitz, Kemelgor, and Uzzi, 2000, p. 137). The male world is characterized by stronger social and professional ties than the female world: while men form close social ties with other male colleagues, which facilitate access to collegial resources and information, a lone scientist conception may characterize females to a much higher degree than males. Male and female scientists often pursue or are pushed onto somewhat different career tracks (Baker, 2012) and are located in different academic structures, with "differential access to valuable resources" (Xie and Shauman, 2003, p. 193)—one of them being extended international networks. The social capital, or networks of contacts (see Villanueva-Felez, Molas-Gallart, and Escribá-Esteve, 2013), of male and female scientists differs substantially as do faculty network relationships (Etzkowitz et al., 2000, p.157-177). The lower social capital of female scientists is linked to gender-based homophily in research collaborations: the tendency for scientists to collaborate (and co-author) with individuals of the same gender.

Female scientists exhibit stronger gender homophily (Jadidid, Karimi, Lietz, and Wagner, 2018). Evidence from co-authorship patterns in economics indicates that team formation in academic publishing is not gender-neutral: there is powerful gender sorting in team formation (Boschini and Sjögren, 2007). Women are less likely than men to co-author and when they do, they are less likely to occupy the first or last author positions; moreover, they also single-author significantly more than men (Wang, Lee, West, Bergstrom, and Erosheva, 2019). However, the practices of collaboration between males and females differ across disciplines (e.g., between management sciences and economics, as shown on an empirical basis by Maddi, Larivière, and Gingras, 2019). The gender gap in the propensity to co-author with a woman increases in the presence of females in the field (Boschini and Sjögren, 2007), with certain fields being more female-orientated and others continuing to be dominated by men (Halevi, 2019). The author's gender affects the citations received: as the proportion of females as authors per article increases, the citations tend to decrease (in economics, Maddi et al., 2019), which is consistent with previous studies showing that female scientists have fewer international collaborations than male scientists and that the level of citations is higher for articles written as part of international collaborations (Larivière et al., 2011). Female scientists are reported (at universities in Québec) to, on average, receive less funding for research than men, to be generally less productive in terms of publications, and to be at a slight disadvantage in terms of the scientific impact of their publications as measured by citations (Larivière et al., 2011; as in Norway, see Aksnes, Rørstad, Piro, and Sivertsen, 2011).

Male-female collaboration practices are governed by the homophily principle: similarity breeds connection between individuals (McPherson, Smith-Lovin, and Cook, 2001) and structures network ties of every type. Homophily limits people's social worlds in a manner that has powerful implications for the information they receive, the attitudes they form, and the interactions they experience (McPherson et al., 2001). Gender-based homophily has substantial implications for access to networks in science and for forming international research collaborative teams. While "old boy networks" in science impact the decision to publish and co-author, gender differences in network access are reported to change over time as females become more represented across disciplines (e.g., economics) (McDowell, Singell, and Stater, 2006). Female researchers may particularly require good mentoring (Long, 1990) because science is traditionally dominated by men (Fell and König, 2016). "Female" networks in science (e.g., computing) (Jadidi et al., 2018, pp. 20-21) are significantly smaller, much more closed, contain fewer brokerage opportunities (links), and are more short-lived (according to median values of collaboration duration); moreover, females also tend to collect knowledge in tightly knit communities.

2.2. Gender Differences in Collaboration Networks and Interdisciplinarity

Bibliometric studies usually refer to international research collaboration defined as the production of internationally co-authored publications; in contrast, survey-based studies usually define international research collaboration as research conducted with international collaborators. Both survey and bibliometric approaches are closely linked and examine related phenomena from different angles; both approaches clearly show that male and female scientists collaborate differently (except for top performers) (Abramo et al., 2019; Yemini 2019). Female scientists tend to be more focused on national and institutional collaborations, and male scientists are more focused on international collaborations. In addition, female scientists are often less specialized and more interdisciplinary in their collaborations (Rhoten and Pfirman, 2007), which has negative implications for the acceptance of their papers in top academic journals and, more generally, for research-based academic reputation in which specialization promotes excellence more than interdisciplinarity (or diversification). Female scientists are reported to specialize less than male scientists and thereby lose out on an important means of increasing their productivity, having a specialized research program (that is, repeatedly writing papers in the same specialty area), and promoting productivity (Leahey, 2006). Men specialize because they think “a diversified research program indicates a failure to excel in any one area, whereas females diversify because they think it indicates scholarly breadth” (Leahey, 2006, p. 774). Research collaboration—particularly international research collaboration—requires reliable collaboration networks, which are smaller, more interdisciplinary, and more egalitarian (Araújo, Araújo, Moreira, Herrmann, and Andrade, 2017) in the case of female scientists.

The increasing global focus on research and its internationalization—supported by the vertically stratifying mechanisms of global university rankings (and their increasing reliance on metrics favoring publishing in top academic journals and publishing in international collaboration)—appear to have powerful gendered consequences (see the following three monographs in particular: Xie and Shauman, 2003; Baker, 2012; Zippel, 2017). Further, the four aspects of academic work and lives persistently studied within the academic profession literature—research productivity, teaching and research time distribution, teaching versus research role orientation, and research internationalization—are clearly gendered. As we have shown elsewhere based on large-scale survey data (Kwiek, 2019a; Kwiek, 2016), female scientists across 11 European systems are less productive on average, spend a higher proportion of time on teaching and a lower proportion of time on research, exhibit lower levels of research orientation and higher levels of teaching orientation, and are less internationalized in research than their male peers. The combination of these factors is important, as they tend to define advancement in academic careers in research-intensive universities.

In addition, female scientists are less prone to leading large-scale projects with multiple collaborators favored by granting agencies (Baker, 2012), which increases academic visibility (Maddi et al., 2019) through prestigious multi-authored international publications and their citations, thereby leading male scientists to an increased amount of collaboration, more visibility, and more grants in the future. Bruno Latour's "credibility cycle" (Latour and Woolgar, 1986) in academic careers may be repeated faster in the current output-oriented research-focused environment for males than for females (each cycle leading from prestigious publications to collaborative grants to new collaborative publications to more prestige and reputation and again to new collaborative grants, partially explaining the comparative advantage of male scientists and their faster academic promotions). Female scientists are also punished in terms of the research-defined academic success in terms of a natural coincidence of the rapid research output development phase in science with the final decade of childbearing—"many women simply do not get past the critical threshold, and are unable to participate"—thereby often clustering in part-time teaching-focused academic work (O'Brien and Hapgood, 2012).

In the long run, the globalization of science as is currently developing (including sustained global focus on international collaborative research, large-scale research grants, the overwhelming role of top journals in academic knowledge production, and the increasing global role of productivity metrics in career progression) presents greater disadvantages for female scientists than for male scientists. Specifically, the growing importance of international research collaboration in academic promotions entails comparative disadvantages for females (Zippel, 2017) who are, on average, less internationalized.

2.3. "Internationalization Accumulative Advantage" and "Localization Accumulative Disadvantage"

The combination of over-reliance on interdisciplinary collaborations and under-reliance on international collaborations among female researchers, both with powerful underlying structural factors, may slow down their academic careers as compared to those of males. While men are found to produce more publications during their PhD and postdoctoral years, females in these years are often diverted by marriage, starting a family, and child bearing (Halevi, 2019). For females, there is always the contradiction between "the tenure clock and the biological clock," in which the biological clock does not fit traditional assumptions of (male) youthful achievements in science according to which there is "a forced march in the early years, allowing a slower pace later. This is exactly the opposite of the structure that would be preferable for most women" (Etzkowitz et al., 2000, p. 141; Long, 1990, p. 1302). In addition, the collaboration patterns and professional networks of female scientists reflect their greater focus on teaching and service compared to males. The gap in productivity between males and females is explained by several reasons, of which one is different collaboration patterns and professional networks (Halevi, 2019; Mayer and Rathmann, 2018). Productivity in science is tied to the environment of work: the signals, resources, and reward schemes, as well as the networks of communication and exchange in the larger community of science (Fox, 1991).

Family choices are more significant for female than for male scientists: while marriage, starting a family, and child care create a gap in productivity that can be difficult to bridge (Halevi, 2019), the decision to remain single also diminishes young female scientists' opportunities of collaboration with male scientists, including with their male PhD mentors (Long, 1990, p. 1307), particularly in the early phases of their academic career. Collaboration with the mentor was found to be the most important factor affecting productivity at the beginning of academic career, with a concentration of small disadvantages explaining gender differences in productivity (Long, 1990). As small advantages accumulate for males and small disadvantages accumulate for females, in the dual processes of "accumulative advantage" and "accumulative disadvantage" (Cole and Cole, 1973, p. 146), the gender disparity in productivity may grow over time. Therefore, differences in productivity at the beginning of the career are essential for understanding the remainder of the career (Long 1990). As noted by Cole and Cole (1973, p. 151), "if women fail to be as productive in the years immediately following their degree, the social process of accumulative disadvantage may take over and contribute to their falling further behind in the race to produce new scientific discoveries." At the same time, the danger is that the objective "meritocracy-driven reliance on quantitative measures of scientific output" may, in fact, prevent female researchers from "proving their worth"—that is, from getting employed or moving up the academic ladder (Nielsen, 2016, p. 2057).

Further, productivity is strongly correlated with international collaboration. In the Polish context, certain scientists are clearly more internationalized than others. Polish research is characterized by two parallel processes that are termed "internationalization accumulative advantage" and "localization accumulative disadvantage." As more international collaboration tends to imply higher publishing rates (and higher citation rates), research internationalization plays an increasingly stratifying role within the Polish academic profession, thereby leading to internationalization accumulative advantage. Increasingly, those who do not collaborate internationally are likely to suffer localization accumulative disadvantage in terms of resources and prestige. The male/female distinction is particularly relevant in this context, as male scientists are more internationalized in research than female scientists. These processes divide the academic community—both across institutions (vertical differentiation) and across faculties within institutions (horizontal segmentation)—into highly internationalized institutions, faculties, research groups, and individual scientists and their less internationalized or more localized counterparts. These processes also divide the academic community by gender.

2.4. The Combined Effect of Gender, Age, and Academic Rank

In terms of survey-based studies, beyond the numerous studies on general research collaboration and gender, several studies have focused specifically on the role of gender in international research collaboration. The findings indicate that being female is a negative predictor of international research collaboration (Rostan, Ceravolo, and Metcalfe, 2014; Vabø, Padilla-Gonzales, Waagene, and Naess, 2014; Kwiek, 2015a), as the prototypical academic figure in international research collaboration is a male full professor in his mid-50s (Rostan et al., 2014). Vabø et al. (2014) found that female scientists report a lower amount of international research collaboration than males, regardless of the intensity of international collaboration in the world regions studied. However, when the data are disaggregated by academic rank, the significance of the gender gap among junior faculty disappears in certain countries (USA, Canada, South Africa, and Australia). Moreover, while male scientists are reported to be generally more involved in international research collaboration, female academics tend to be more involved in internationalization at home—for example, teaching in a foreign language (Vabø et al., 2014, p. 202).

Survey-based studies also reveal that being male significantly increases the odds of involvement in international research collaboration (by 69%) in 11 European countries (see Kwiek, 2018a). In Fox, Realff, Rueda, and Morn (2017, p. 1304), female engineers identified funding and finding collaborators as external barriers to internationalization, while personal or family concerns were perceived as significantly less important barriers for themselves than for others. Further, the “glass fences” concept of unequal access to international research collaboration persists as a form of gendered inequality: “while women do sometimes climb over these fences, they require extra efforts” compared to men (Uhly et al., 2015, p. 3). For an account of how science globalization perpetuates gender inequalities and creates disadvantages for female scientists, see Zippel (2017). For an account of internationalization (particularly international mobility) as “indirect discrimination” against female scientists, see Ackers (2008).

Bibliometric research on gender disparity in international collaboration at a national level has been conducted in Norway and Italy. The general conclusion for Norway was that the propensity to collaborate in international research was similar for both male and female scientists; however, for Italy, this propensity is higher for male scientists across the entire population but similar for male and female top performers (Abramo et al., 2019). Successive studies have addressed the gap in research on gender differences in research collaboration in general and international research collaboration in particular by taking the individual scientist as the unit of analysis for both entire populations and for the sample of top performers. In the case of all Italian scientists, Abramo et al. (2013) showed that female scientists are more likely to collaborate domestically both intramurally and extramurally but are less likely to engage in international collaboration. The methodology used in their study avoids distortion by outliers—that is, by cases of highly productive and highly internationalized scientists whose extensive publications distort aggregate index values (Abramo et al., 2013, p. 820; on impact, see Piro, Rørstad, and Aksnes, 2016). The same approach is adopted in this paper.

In Norway, Aksnes, Piro, and Rørstad (2019) used the Cristin bibliographic database (Norwegian Science Index of all peer-reviewed publications) to study gender differences in international collaboration. Their unit of analysis was also the individual scientist; counting all individuals equally as single units, regardless of productivity (Aksnes et al. 2019: 8), limited the effect of the outliers. Analyzed by field, academic position, and publication productivity, scientific discipline emerged as the most important determinant of international research collaboration, while gender differences were not statistically significant.

A few studies combine age, academic rank, and international research collaboration (see Zeng, Duch, Sales-Pardo, Moreira, Radicchi, Ribeiro et al., 2016) —as is the case of our paper—because only a few data sets combine administrative and biographical data at the individual level on the one hand and publication and citation data on the other hand. These combinations can be studied at the level of individual institutions; however, large-scale studies at the national level depend on data set mergers (for Italy, see Abramo, D'Angelo, and Solazzi, 2011; Abramo, D'Angelo, and Murgia, 2016) or on comprehensive national databases such as Norway's Cristin. Given the policy challenge posed by the progressive aging of European scientists, data-driven studies of national populations of scientists, as well as their propensity to collaborate, are particularly useful. Our data set enables us to examine the propensity to undertake major collaboration types not only by gender and academic position but also by age. For example, in a study of Italian full professors, Abramo et al. (2016, p. 318) concluded that productivity declines significantly with age. However, professors appointed at a young age were more likely to maintain and increase their productivity as compared to colleagues promoted at a later age. In a study of age and productivity among Italian National Research Council scientists, Bonaccorsi and Daraio (2003, p. 75) concluded that productivity declines with age and that the average age of researchers is increasing, with severe policy implications for national science systems.

International research collaboration can be studied by age or by academic generation. Belonging to a specific historical generation can impact both individual productivity (Kwiek, 2019a) and individual opportunities to engage in international collaboration (Rostan et al., 2014, p. 125). Here, "generation" may refer to "biographical generation" (expressed as biological age) or "status generation" (expressed as career stage) (Jung, Kooij, and Teichler, 2014). Seniority by age and by career stage tend to overlap in most countries, including Poland. There is a simple explanation for senior and older academics' higher propensity to collaborate internationally. A study of 19 countries found that scientists collaborating internationally have "more power, better networks, and longer experience" (Jung et al., 2014, p. 214) and that senior positions entail more resources in terms of "power, prestige, visibility, and scientific standing" (Rostan, 2015, p. 257). Younger academics may also have less success in international collaborations because they are more expensive than national or intra-institutional collaboration. However, for the same 19 countries, Rostan et al. (2014, p. 129) reported that the oldest generation of scientists are an exception to this rule, but gender was not examined in the context of age cohorts.

International research collaboration is becoming increasingly common among younger generations. As one recent study showed, the international collaboration rate in Norwegian research universities increased from 58% in 1992 to 66% in 2001 and to 71% in 2013. Not only are younger generations more internationalized, but almost all generations become increasingly involved in international research collaboration as they age (Kyvik and Aksnes, 2015, p. 1448-1449). As they clearly demonstrated for the youngest age cohort in 1989–1991 (Kyvik and Aksnes, 2015, p. 1448), certain generations excel in international collaboration over time and as they age—younger and older Polish academics are a textbook example of this. Career opportunities and academic norms differed significantly for those entering the academic labor force prior to 1989 and those who came after; the same was applicable to those entering the Polish profession before and after the reforms of the 2010s.

2.5. The Polish Context

Gender disparity in Polish science has rarely been studied. Kosmulski (2015) examined the productivity and impact of male and female scientists in the period 1975–2014 based on a limited set of authors bearing one of 26 most popular "–ski" or "–cki" names, showing that male scientists have generally higher productivity and impact, except for biochemistry where they are almost equal. Siemienska (2007) based her research on two small-scale surveys of full professors and young academics and revealed that cultural capital (measured as the level of parents' education) was particularly important for research productivity of females. As measured by a proxy of internationally co-authored publications, Poland had the lowest level of research internationalization in the European Union in 2018 (based on Scopus data). There are numerous underlying reasons for this; however, in general terms, this relates to the systematic "deinstitutionalization" of the research mission of Polish universities in 1990–2010, followed by its slow "reinstitutionalization" (Kwiek, 2012) powered by two waves of higher education reforms in the previous decade (for overviews of the Polish higher education and science systems, see Antonowicz, 2016; Kwiek, 2015b; Urbanek, 2018; Bieliński and Tomczyńska, 2018; Ostrowicka and Stankiewicz, 2018; Wolszczak-Derlacz and Parteka, 2010).

In addition, Poland is gradually implementing a performance-based research funding system (Kulczycki, Korzeń, and Korytkowski, 2017). Funding is linked either directly to prior research outputs (through subsidies allocated to individual academic units) or indirectly in the form of grant-based competitive funding for academics. However, the Polish science system remains heavily underfunded. According to *Main Science and Technology Indicators* (OECD, 2019), Poland's gross domestic spending on R&D (GERD) in 2017 as a percentage of gross domestic product (GDP) was the fourth-lowest in the European Union (at 1.03 as compared to 1.97 for EU-28 countries and 2.37 for OECD countries). Poland's Higher Education Expenditure on R&D (HERD) as a percentage of GDP also remains among the lowest in the European Union. The low levels of investment in R&D are reflected in publication, citation, and international collaboration data for the period 2009–2018 (Scopus 2020). In 2018, the total Polish publication output was approximately 51,000, with 34,200 articles (5.59% of the total output of 28 European Union member states). Moreover, Poland's share of internationally co-authored articles is the lowest in the EU-28. Although this increased from 29.1% in 2009 to 35.8% in 2018, the EU-28 average was 45.7% in 2018. Polish science exhibits high levels of national (25.4%) and intra-institutional collaboration (28.1%). In terms of Field-Weighted Citation Impact (FWCI), Poland has struggled to achieve the world average of 1.0, which it attained only in the last three years.

However, Poland's average international collaboration impact for 2009–2018 roughly matches the average for EU-13, EU-15, and EU-28 countries. The structure of publications indexed in the Scopus database has remained almost unchanged for a decade, and although research internationalization is a key element of the recent reforms, growth is slow.

3. Data and Methods

3.1. The Dataset

Two large databases were merged in the present research: Database I was an official national administrative and biographical database of all Polish scientists (the Polish Science—*Nauka Polska*) and Database II was the Scopus database as an official publication and citation source used for individual- and institutional-level evaluation in Poland. Database I (created by the OPI National Research Institute) comprised 99,535 scientists employed in the Polish science sector as of November 21, 2017. Only scientists with at least a doctoral degree (70,272) and only scientists employed in the higher education sector were selected for further analysis (54,448 or 54.70% of all scientists with at least a doctoral degree working at 85 universities of various types). The data used were both demographic (gender and date of birth) and professional (highest degree awarded; date of granting PhD, habilitation, and full professorship; and institutional affiliation), with each scientist identified by a unique ID. Database II, the original Scopus publication and citation database, included 169,775 names from 85 institutions whose publications were included in the database for the decade analyzed (2009–2018). Authors in Database II were defined by their institutional affiliations, documents, and individual Scopus IDs.

These two data sources, although overlapping in terms individual IDs, were produced by data managers with different methods of data storage, processing, presentation, and unit identification. The key procedure was to appropriately identify authors with their individual IDs in the two databases and to provide them with a new ID in the integrated database. The integration of the two databases in a deterministic manner (e.g., by first name, last name, and institutional affiliation) would raise issues related to various name-spelling errors routinely found in the Scopus-derived Database II. In deterministic merging, our awareness of the errors made and of the potential impact of these errors on the quality of the new integrated database would be limited (the integration uncertainty is undefined; see Herzog, Scheuren, and Winkler, 2007, p. 82-83). However, these limitations were reduced using probabilistic methods of database integration (as defined in Fellegi and Sunter, 1969; Herzog et al., 2007; Harron, Dibben, Boyd, Hjern, Azimaee, Barreto, and Goldstein, 2017; and Enamorado, Fifield, and Imai, 2019).

In the probabilistic linkage method used, one estimates the probability that a given record in the database A and a given record in the database B belong to the same statistical unit based on the so-called matching variables present in both data sets (e.g., first name, last name, address, date of birth, etc.). Fellegi and Sunter (1969) proposed a method for verifying the accuracy of a combination of records in two data sets using the similarity function of the two combined statistical units. The idea of the method is to classify pairs in space A × B created from joined sets A and B into the set *M* of exact matches and the set *U* of non-matches. Classification is done using the formula of the ratio of probabilities:

$$R = \frac{P(\gamma \in \Gamma | M)}{P(\gamma \in \Gamma | U)}, \quad (1)$$

where γ denotes the arbitrary pattern of compliance in the comparative space Γ. If the probability that the compared pair of records is a match is high, then R is high, otherwise it is small. These patterns are estimated by comparing the compliance of entries in the matching variables by calculating string distance.

Separately within each of the 85 universities, the first name and the last name records of each record in Database I were compared with each of the records in Database II using the Jaro-Winkler string distance (with values from 0 to 1; see Jaro, 1989; Winkler, 1990; Winkler, 2006). Pairs of strings with a distance greater than 0.94 were considered identical (2) (see Table 1), pairs with a distance greater than 0.88 but less than 0.94 were considered similar (1), while those with a distance less than 0.88 were considered disparate (0). Next, using expectation maximization (Enamorado et al., 2019), the probability that a given pair of records belongs to the same unit was estimated. If the probability was greater than 0.85, the pair was considered to be part of the same unit (Harron, Dibben, Boyd, Hjern, Azimaee, Barreto, and Goldstein, 2017). The computation was made using the fastLink R package.

Table 1. An example of probabilistic integration output (identical, similar and disparate pairs of strings).

Last name, Database II	**First name, Database II**	**Last name, Database I**	**First name, Database I**	**Last name compliance**	**First name compliance**	**Posterior probability**
Kwiek	Marek	Kwiek	Marek	2	2	0.9975556
Mrowiec	Bozena	Mrowiec	Bożena	2	1	0.9946168
Sobkow	Agata	Sobków	Agata	1	2	0.9991700
Wltek	Bozena	Witek	Bożena	1	1	0.9073788
Mudry	Z.	Mudryk	Zbigniew	2	0	0.8846165

By employing a probabilistic approach to the merging of the data sets, it was possible to estimate the uncertainty of the process and, thus, assess the quality of the new integrated database by calculating the percentage of records incorrectly classified as matches (false discovery rate, FDR) and the percentage of records incorrectly classified as non-matches (false negative rate, FNR). An integrated database obtained in accordance with the above procedures and used in our research finally included 37,081 records.[1] Next, Database I was deterministically merged with Database II.

[1] There were 38,750 records referring to 32,937 unique authors (more than one occurrence in Database II was found for 4,452 people or 13.51% of unique authors. With regard to quality, FDR was 0.21% and FNR was 39.91%. The high value of FNR is the result of duplicate instances in the database. There were 9,931 records that referred to more than one person,

Database I contained information on all documents published by authors affiliated with the 85 largest Polish universities in the 2009–2018 reference period. Database II contained metadata on 377,886 papers, with titles and documents' individual Scopus IDs, author names and their individual Scopus IDs, year of publication, journal title, journal ratings, authors affiliations, and the All Science Journal Classification (ASJC) disciplines to which papers were ascribed. ASJC is a subject classification scheme employed by Scopus to index source titles in a structured hierarchy of disciplines (27) and sub-disciplines (334). For all cross-disciplinary analyses, the ASJC subject categories were employed. From among the 377,886 papers in the original Database II, 230,007 were written by the authors included in Database I. Subsequently, only articles written in journals were selected for further analysis, with the number of papers in the database reducing to 159,943 articles. Approximately half of the Polish scientists from the higher education sector did not publish a paper indexed in the Scopus database in the reference period— which is in line with previous findings about the distribution of Polish publications—with the overwhelming majority of publications belonging to national publication outlets.

3.2. Methods

Finally, following the idea of Abramo, Aksnes, and D'Angelo (2019, pp. 7-8) who defined the dominant Web of Science subject category for each Italian and Norwegian professor, every Polish scientist present in the integrated database was ascribed to one of 334 ASJC disciplines at the four-digit level and one of 27 ASJC disciplines at the two-digit level. In the ASJC system used, a given paper can have one or multiple disciplinary classifications. The ASJC discipline ascribed to each scientist was the one that occurred the most frequently among all articles authored by each scientist in the decade studied; in the case of an equal number of dominant disciplines, they were selected at random from among the dominant ones.[2] Further, three disciplines were omitted from analysis as they did not meet the minimum threshold of 50 scientists per

---

where 3,679 (82.63%) occurred twice, 609 (13.68%) occurred three times, and 169 (3.68%) occurred four or more times. Therefore, among duplicated records, a clerical review was performed (as suggested in Herzog et al., 2007). Manual verification of duplicate records revealed that 1,207 (12.15% in terms of duplicated records and 3.11% of all integrated records) records were incorrectly assigned to the same person. These records were deleted from the integrated database, yielding N = 37,081 records.

[2] The ASJC discipline codes were described in the paper in the following manner: AGRI Agricultural and Biological Sciences; HUM Arts and Humanities; BIO Biochemistry, Genetics and Molecular Biology; BUS Business, Management and Accounting; CHEMENG Chemical Engineering; CHEM Chemistry; COMP Computer Science; DEC Decision Science; EARTH Earth and Planetary Sciences; ECON Economics, Econometrics and Finance; ENER Energy; ENG Engineering; ENVIR Environmental Science; IMMU Immunology and Microbiology; MATER Materials Science; MATH Mathematics; MED Medicine; NEURO Neuroscience; NURS Nursing; PHARM Pharmacology, Toxicology and Pharmaceutics; PHYS Physics and Astronomy; PSYCH Psychology; SOC Social Sciences; VET Veterinary; DENT Dentistry; and HEALTH Health Professions.

discipline (GEN, NEURO, and NURS). Therefore, the sample studied in this paper comprised 25,463 scientists.

Gender determination of names—for example, by manual identification of gender by author names—was not necessary: the administrative and biographical Database I contained gender information and date of birth information for all the observations. The integrated dataset of Polish scientists with their administrative, biographical, publication and citation data used in this research (termed "The Polish Science Observatory") is maintained and periodically updated by the two authors as part of ongoing research programs in the Center for Public Policy at the University of Poznan.

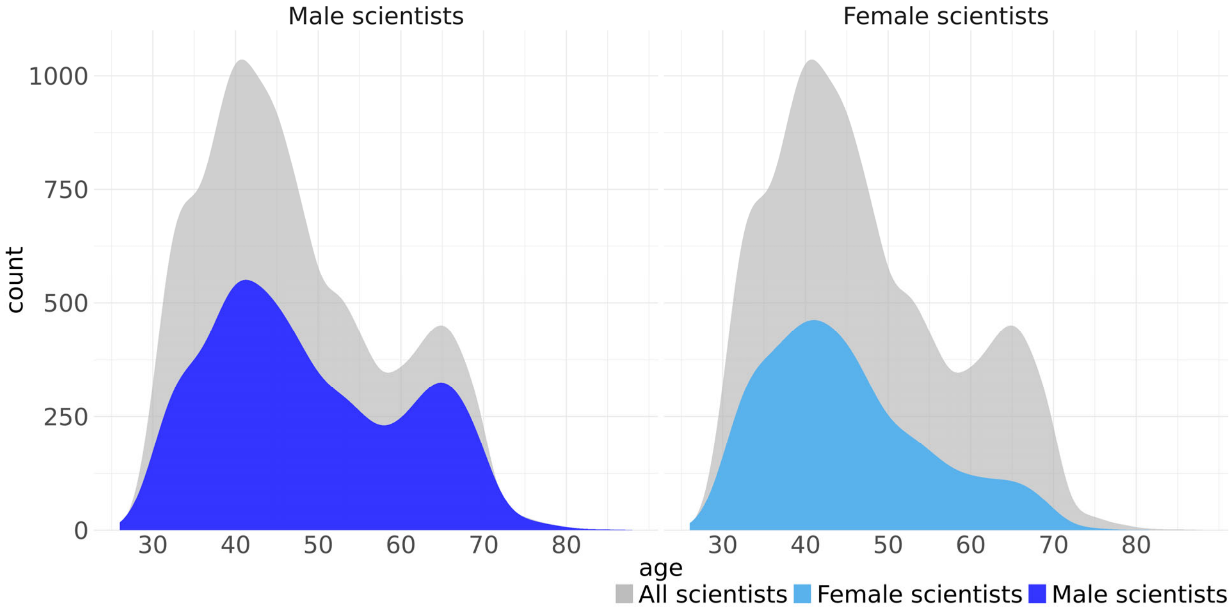
3.3. The Sample

The structure of the sample (N = 25,463) is presented in Tables 2 and 3: in terms of age groups, approximately half of the scientists are in the 36–50 age bracket (51.6%) and in terms of academic positions, over half of them are assistant professors (56.0%). Column percentages enable the analysis of gender distribution of the Polish academic profession by age groups, academic positions, and disciplines, while row percentages enable analysis of how male and female scientists are distributed by a given age group, academic position, and discipline (Table 2) (In a two-dimensional cross-tabulation, row percentages are computed by dividing the count for a cell by the total sample size for that row. A row percentage shows the proportion of scientists in a column category from among those in the row). The table also enables a study of age distribution for each academic position from a gender perspective (Table 3). The three largest disciplines represented in the sample are agricultural and biological sciences, engineering, and medicine (AGR, ENG, and MED), representing over one-third of the scientists (37.8%). Female participation in the academic profession was found to decrease with age: while female scientists represent approximately half of all scientists aged 31–35, they represent only about a quarter of all scientists aged 61–65 years (49.8% and 26.7%, respectively). The theme of female scientists as newcomers to the Polish academic sector, and as a group clustered in lower academic positions, is strengthened by the finding that while females constitute approximately half of all assistant professors, they represent only approximately a quarter of full professors (48% and 24%, respectively). Polish assistant professors under 45 (with doctorates only) have an almost equal gender distribution. The older professors (aged 41–55 years) with a habilitation degree (a second, postdoctoral degree) are already dominated by male scientists (who represent approximately 60% of associate professors). In the case of full professors, the number of males is at least three times higher than the number of females (see Table 3) for every age group, for both young professors (aged 41–45) and the oldest ones (aged 61–65).

Further, female scientists are severely underrepresented in computer science (COMP 16.5%), engineering (ENG 14.9%), physics and astronomy (PHYS 16.6%), and mathematics (MATHS 25.2%). The only larger discipline in which female scientists are overrepresented is biochemistry, genetics, and molecular biology (BIO 60.0%), with all other disciplines in which they may be overrepresented being small. In arts and humanities (HUM) and social sciences (SOC), the distribution of scientists by gender is practically equal (49.8%).

The age structure of the sample by gender is presented in Figure 1: the striking feature is the ever-decreasing number of female scientists after the peak in their numbers at approximately 40 years (as in the case of male scientists) contrasted with an increasing number of older male scientists, with a second peak at approximately 65 years. It must be remembered that our sample contains only scientists who had at least a single publication in the Scopus database in the period 2009–2018 and, therefore, it includes internationally productive Polish academic scientists.

Figure 1. Age structure of the sample, all Polish internationally productive university professors (N = 25,463), by gender. All university professors in grey.



In this paper, international research collaboration (defined as the occurrence of an article with at least two authors, of which at least one has a non-Polish institutional affiliation) is examined in the context of three other collaboration types: collaboration in general (defined as the occurrence of an article with at least two authors); national collaboration (article with at least two authors with two different Polish affiliations), and institutional collaboration (article with at least two authors with the same Polish affiliation). An article published in an international collaboration can also be counted as an article published in national collaboration (if it has authors with at least two different Polish affiliations) and institutional collaboration (if it has authors with at least two of the same Polish affiliations). An article published in national collaboration can also be counted as an article published in institutional collaboration (if at least two authors have the same Polish institutional collaboration), following traditional distinctions between collaboration types (see Abramo et al., 2013).

4. Results

4.1. Four Collaboration Types and Three Collaboration Intensity Levels

Three approaches to measuring international research collaboration were tested in this paper: a threshold approach (one article only, as in Abramo et al., 2013), a minimum 50% approach, and a minimum 75% approach, representing a low, medium, and high collaboration intensity, respectively. Distributions by gender and dominating discipline were examined for each approach: for a minimum 50% (and 75%) approaches, over 50% (and 75%) of collaborative papers in an individual portfolio of articles from 2009 to 2018 were ascribed to a given scientist.

Table 2. Structure of the sample, all Polish internationally productive university professors, by gender, age group, academic position, and discipline, presented with column and row percentages.

		Female			Male			Total		
		N	% col	% row	N	% col	% row	N	% col	% row
Age groups	26 - 30	245	2.3	46.8	279	1.9	53.2	524	2.1	100.0
	31 - 35	1 653	15.6	49.8	1 664	11.2	50.2	3 317	13.0	100.0
	36 - 40	2 148	20.3	47.1	2 411	16.2	52.9	4 559	17.9	100.0
	41 - 45	2 272	21.5	45.7	2 696	18.1	54.3	4 968	19.5	100.0
	46 - 50	1 569	14.8	43.8	2 013	13.5	56.2	3 582	14.1	100.0
	51 - 55	993	9.4	40.3	1 471	9.9	59.7	2 464	9.7	100.0
	56 - 60	671	6.3	37.7	1 108	7.4	62.3	1 779	7.0	100.0
	61 - 65	563	5.3	26.7	1 548	10.4	73.3	2 111	8.3	100.0
	66 - 70	417	3.9	23.0	1 396	9.4	77.0	1 813	7.1	100.0
	71+	46	0.4	13.3	300	2.0	86.7	346	1.4	100.0
	Total	**10 577**	**100.0**	**41.5**	**14 886**	**100.0**	**58.5**	**25 463**	**100.0**	**100.0**
Academic positions	Assistant Pr.	6 851	64.8	48.0	7 420	49.8	52.0	14 271	56.0	100.0
	Asssoc. Pr.	2 822	26.7	38.0	4 596	30.9	62.0	7 418	29.1	100.0
	Full Pr.	904	8.5	24.0	2 870	19.3	76.0	3 774	14.8	100.0
	Total	**10 577**	**100.0**	**41.5**	**14 886**	**100.0**	**58.5**	**25 463**	**100.0**	**100.0**
Disciplines (ASJC)	AGRI	1 444	13.7	53.4	1 258	8.5	46.6	2 702	10.6	100.0
	BIO	1 068	10.1	60.0	712	4.8	40.0	1 780	7.0	100.0
	BUS	372	3.5	52.1	342	2.3	47.9	714	2.8	100.0
	CHEM	756	7.1	51.3	719	4.8	48.7	1 475	5.8	100.0
	CHEMENG	185	1.7	38.5	296	2.0	61.5	481	1.9	100.0
	COMP	170	1.6	16.5	860	5.8	83.5	1 030	4.0	100.0
	DEC	24	0.2	44.4	30	0.2	55.6	54	0.2	100.0
	DENT	57	0.5	76.0	18	0.1	24.0	75	0.3	100.0
	EARTH	385	3.6	33.4	769	5.2	66.6	1 154	4.5	100.0
	ECON	186	1.8	49.1	193	1.3	50.9	379	1.5	100.0
	ENER	82	0.8	27.8	213	1.4	72.2	295	1.2	100.0
	ENG	501	4.7	14.9	2 857	19.2	85.1	3 358	13.2	100.0
	ENVIR	848	8.0	50.5	832	5.6	49.5	1 680	6.6	100.0
	HEALTH	23	0.2	34.3	44	0.3	65.7	67	0.3	100.0
	HUM	527	5.0	49.8	531	3.6	50.2	1 058	4.2	100.0
	IMMU	90	0.9	75.6	29	0.2	24.4	119	0.5	100.0
	MATER	495	4.7	33.9	967	6.5	66.1	1 462	5.7	100.0
	MATH	259	2.4	25.2	767	5.2	74.8	1 026	4.0	100.0
	MED	1 920	18.2	53.7	1 654	11.1	46.3	3 574	14.0	100.0
	PHARM	169	1.6	66.5	85	0.6	33.5	254	1.0	100.0
	PHYS	182	1.7	16.6	916	6.2	83.4	1 098	4.3	100.0
	PSYCH	194	1.8	63.8	110	0.7	36.2	304	1.2	100.0
	SOC	494	4.7	49.8	498	3.3	50.2	992	3.9	100.0
	VET	146	1.4	44.0	186	1.2	56.0	332	1.3	100.0
	Total	**10 577**	**100.0**	**41.5**	**14 886**	**100.0**	**58.5**	**25 463**	**100.0**	**100.0**

Table 3. Structure of the sample, all Polish internationally productive university professors by gender, age group, and academic position, presented with column and row percentages.

	Assistant Professor						**Asssociate Professor**						**Full Profesor**						**Total**					
	Female			**Male**			**Female**			**Male**			**Female**			**Male**			**Female**			**Male**		
	N	**% col**	**% row**	**N**	**% col**	**% row**	**N**	**% col**	**% row**	**N**	**% col**	**% row**	**N**	**% col**	**% row**	**N**	**% col**	**% row**	**N**	**% col**	**% row**	**N**	**% col**	**% row**
26 - 30	246	3.6	46.8	280	3.8	53.2	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	0	0.0	0.0	246	2.3	46.8	280	1.9	53.2
31 - 35	1636	23.8	50.2	1621	21.8	49.8	23	0.8	32.4	48	1.0	67.6	0	0.0	0.0	1	0.0	100.0	1659	15.6	49.8	1670	11.2	50.2
36 - 40	1937	28.2	49.8	1956	26.3	50.2	219	7.7	32.6	453	9.8	67.4	2	0.2	22.2	7	0.2	77.8	2158	20.3	47.2	2416	16.2	52.8
41 - 45	1589	23.1	49.0	1652	22.2	51.0	664	23.5	40.4	981	21.3	59.6	25	2.8	24.8	76	2.6	75.2	2278	21.5	45.7	2709	18.1	54.3
46 - 50	820	11.9	46.4	946	12.7	53.6	695	24.6	43.2	914	19.8	56.8	57	6.3	26.6	157	5.5	73.4	1572	14.8	43.8	2017	13.5	56.2
51 - 55	361	5.2	45.9	425	5.7	54.1	523	18.5	40.9	757	16.4	59.1	112	12.3	27.7	292	10.1	72.3	996	9.4	40.3	1474	9.9	59.7
56 - 60	195	2.8	46.7	223	3.0	53.3	336	11.9	39.6	513	11.1	60.4	144	15.9	27.8	374	13.0	72.2	675	6.4	37.8	1110	7.4	62.2
61 - 65	80	1.2	25.0	240	3.2	75.0	257	9.1	31.5	558	12.1	68.5	227	25.0	23.2	753	26.2	76.8	564	5.3	26.7	1551	10.4	73.3
66 - 70	13	0.2	12.9	88	1.2	87.1	105	3.7	23.7	338	7.3	76.3	302	33.3	23.7	974	33.9	76.3	420	4.0	23.1	1400	9.4	76.9
71+	0	0.0	0.0	13	0.2	100.0	8	0.3	15.1	45	1.0	84.9	38	4.2	13.5	243	8.4	86.5	46	0.4	13.3	301	2.0	86.7
Total	**6877**	**100.0**	**48.0**	**7444**	**100.0**	**52.0**	**2830**	**100.0**	**38.1**	**4607**	**100.0**	**61.9**	**907**	**100.0**	**24.0**	**2877**	**100.0**	**76.0**	**10614**	**100.0**	**41.6**	**14928**	**100.0**	**58.4**

Further, we examined the gender disparity for each of the four collaborative types (general, international, national, and institutional) according to the three levels of collaboration intensity. In the case of general collaboration (any collaboration type, a superset of all other collaboration types), the gender disparity emerges as differentiated by discipline and collaboration intensity. While at low levels of intensity in general collaboration, gender disparities are negligible, they increase with intensity. The same pattern is observed for national, institutional, and international collaboration (for example, in the case of institutional collaboration—a definitely dominating collaboration type in Poland—gender differences increase with collaboration intensity in the following manner: from 83.1% vs. 82.6% at a low intensity level to 67.1% vs. 62.6% at a medium intensity level to 53.6% vs. 47.2% at a high intensity level).

In this section, we focus on gender disparities in high-intensity collaborations. In the case of high-intensity collaborations for all disciplines combined, collaboration rates by gender are higher for females in general collaboration, institutional collaboration, and national collaboration (83.2% vs. 80.6%, 53.6% vs. 47.2%, and 4.5% vs. 3.9%, respectively (see Total in Table 4). Male scientists exhibit higher collaboration rate only for the most demanding and most expensive collaboration type: international collaboration (4.1% and 5.2%).

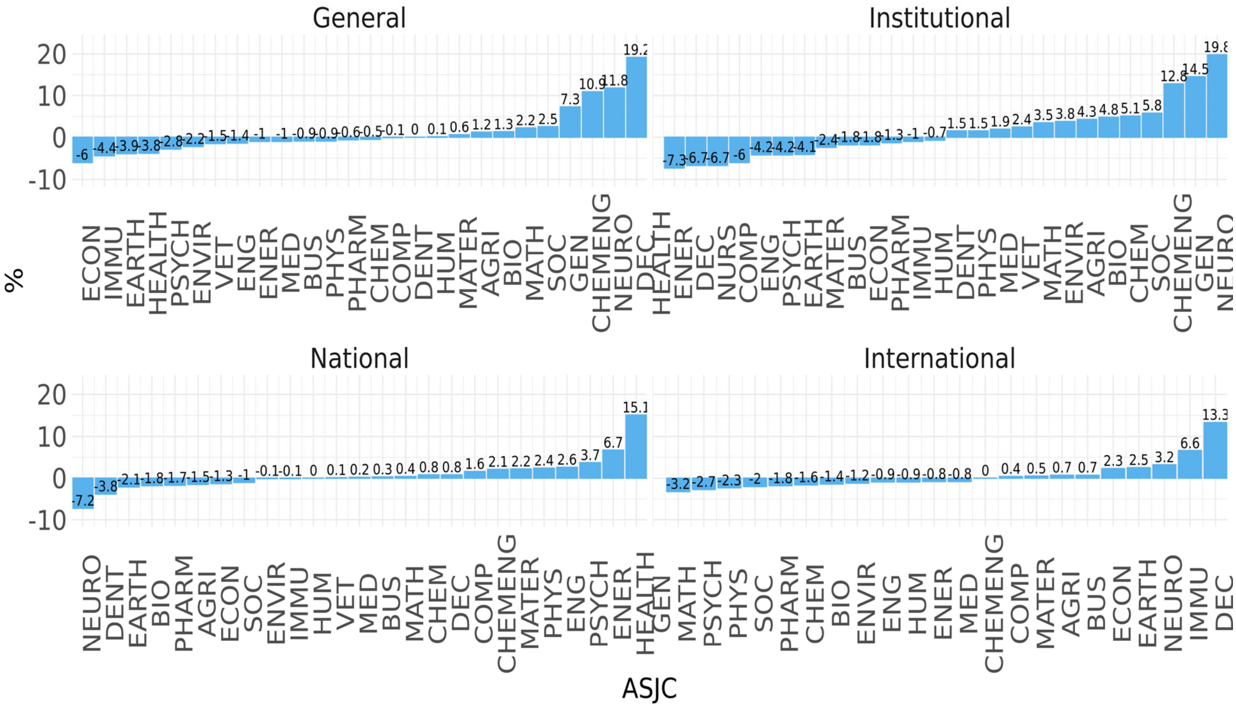
However, the data analysis for all disciplines combined do not tell the entire story of gender disparity. There is a fascinating cross-disciplinary gender disparity in all four collaboration types. In general, in national and institutional collaboration, there are specific disciplines in which male scientists exhibit higher collaboration propensity; in contrast, in international collaboration, there are specific disciplines in which female scientists exhibit higher collaboration propensity. Figure 2 presents gender differences by collaboration type (four panels) and discipline in greater detail: results above zero indicate a female advantage in a given discipline and results below zero indicate a male advantage. The differences by discipline are presented for high collaboration intensity only.

In general, the propensity of male scientists to collaborate is higher in 13 out of 24 disciplines (see panel 1 in Figure 2). Among the three largest disciplines (AGRI, ENG, and MED), female scientists show higher propensity in only one—AGRI. In contrast to Abramo et al. (2013), who found higher propensity to general collaboration among Italian females for almost all disciplines, the propensity to collaborate in general is higher for Polish male scientists for over half of all disciplines, with the highest percentage difference for ECON—reaching 6.0 percentage points For national collaboration, the male advantage was found for 10 disciplines, with HUM exhibiting no gender difference. The female advantage reached higher levels for HEALTH and ENERGY, two small disciplines. For institutional collaboration, the male advantage was found for 13 disciplines.

Table 4. Percentage differences in high-intensity collaborations (>75% articles published in international collaboration in the scientist's individual publication portfolio for 2009–2018), all Polish internationally productive university professors, by collaboration type, gender, and ASJC discipline (in %).

		Scientists: total		Scientists: general collaboration			Scientists: international collaboration			Scientists: national collaboration			Scientists: institutional collaboration		
		N	% col	N	% col	% row	N	% col	% row	N	% col	% row	N	% col	% row
AGRI	F	1444	53.4	1361	53.8	94.3	44	59.5	3	52	44.8	3.6	896	55.2	62
	M	1258	46.6	1170	46.2	93	30	40.5	2.4	64	55.2	5.1	727	44.8	57.8
BIO	F	1068	60	1053	60.3	98.6	48	53.3	4.5	59	53.2	5.5	782	61.6	73.2
	M	712	40	693	39.7	97.3	42	46.7	5.9	52	46.8	7.3	487	38.4	68.4
BUS	F	372	52.1	237	51.7	63.7	43	53.8	11.6	23	53.5	6.2	125	50.8	33.6
	M	342	47.9	221	48.3	64.6	37	46.3	10.8	20	46.5	5.8	121	49.2	35.4
CHEM	F	756	51.3	723	51.1	95.6	34	43.6	4.5	29	56.9	3.8	449	53.5	59.4
	M	719	48.7	691	48.9	96.1	44	56.4	6.1	22	43.1	3.1	390	46.5	54.2
CHEM-ENG	F	185	38.5	174	41.4	94.1	5	38.5	2.7	12	48	6.5	120	43.8	64.9
	M	296	61.5	246	58.6	83.1	8	61.5	2.7	13	52	4.4	154	56.2	52
COMP	F	170	16.5	132	16.5	77.6	13	17.3	7.6	8	22.9	4.7	55	14.3	32.4
	M	860	83.5	669	83.5	77.8	62	82.7	7.2	27	77.1	3.1	330	85.7	38.4
DEC	F	24	44.4	15	53.6	62.5	4	80	16.7	1	50	4.2	4	36.4	16.7
	M	30	55.6	13	46.4	43.3	1	20	3.3	1	50	3.3	7	63.6	23.3
DENT	F	57	76	57	76	100	0	0	0	1	50	1.8	42	76.4	73.7
	M	18	24	18	24	100	0	0	0	1	50	5.6	13	23.6	72.2
EARTH	F	385	33.4	272	32.2	70.6	28	43.1	7.3	10	21.7	2.6	125	30.8	32.5
	M	769	66.6	573	67.8	74.5	37	56.9	4.8	36	78.3	4.7	281	69.2	36.5
ECON	F	186	49.1	92	46.2	49.5	10	62.5	5.4	12	44.4	6.5	42	47.2	22.6
	M	193	50.9	107	53.8	55.4	6	37.5	3.1	15	55.6	7.8	47	52.8	24.4
ENER	F	82	27.8	65	27.5	79.3	4	25	4.9	14	38.9	17.1	36	24.8	43.9
	M	213	72.2	171	72.5	80.3	12	75	5.6	22	61.1	10.3	109	75.2	51.2
ENG	F	501	14.9	377	14.7	75.2	10	10.9	2	25	26.6	5	197	13.7	39.3
	M	2857	85.1	2189	85.3	76.6	82	89.1	2.9	69	73.4	2.4	1242	86.3	43.5
ENVIR	F	848	50.5	749	49.9	88.3	16	38.1	1.9	33	50	3.9	501	52.1	59.1
	M	832	49.5	753	50.1	90.5	26	61.9	3.1	33	50	4	460	47.9	55.3
HEALTH	F	23	34.3	19	33.3	82.6	0	0	0	4	80	17.4	7	24.1	30.4
	M	44	65.7	38	66.7	86.4	0	0	0	1	20	2.3	22	75.9	50
HUM	F	527	49.8	99	50	18.8	17	43.6	3.2	16	50	3	52	48.1	9.9
	M	531	50.2	99	50	18.6	22	56.4	4.1	16	50	3	56	51.9	10.5
IMMU	F	90	75.6	86	74.8	95.6	9	90	10	3	75	3.3	55	75.3	61.1
	M	29	24.4	29	25.2	100	1	10	3.4	1	25	3.4	18	24.7	62.1
MATER	F	495	33.9	464	34	93.7	29	35.8	5.9	23	48.9	4.6	270	32.9	54.5
	M	967	66.1	901	66	93.2	52	64.2	5.4	24	51.1	2.5	551	67.1	57
MATH	F	259	25.2	154	26	59.5	13	17.1	5	8	27.6	3.1	53	29	20.5
	M	767	74.8	439	74	57.2	63	82.9	8.2	21	72.4	2.7	130	71	16.9
MED	F	1920	53.7	1862	53.5	97	44	46.3	2.3	89	54.9	4.6	1351	54.4	70.4
	M	1654	46.3	1621	46.5	98	51	53.7	3.1	73	45.1	4.4	1132	45.6	68.4
PHARM	F	169	66.5	166	66.4	98.2	3	50	1.8	9	60	5.3	129	66.2	76.3
	M	85	33.5	84	33.6	98.8	3	50	3.5	6	40	7.1	66	33.8	77.6
PHYS	F	182	16.6	162	16.4	89	25	14.5	13.7	10	26.3	5.5	85	17	46.7
	M	916	83.4	824	83.6	90	147	85.5	16	28	73.7	3.1	414	83	45.2
PSYCH	F	194	63.8	148	63	76.3	16	57.1	8.2	16	76.2	8.2	50	60.2	25.8
	M	110	36.2	87	37	79.1	12	42.9	10.9	5	23.8	4.5	33	39.8	30
SOC	F	494	49.8	193	51.5	39.1	23	41.1	4.7	17	43.6	3.4	120	56.6	24.3
	M	498	50.2	182	48.5	36.5	33	58.9	6.6	22	56.4	4.4	92	43.4	18.5
VET	F	146	44	143	43.6	97.9	0	0	0	4	44.4	2.7	122	44.7	83.6
	M	186	56	185	56.4	99.5	1	100	0.5	5	55.6	2.7	151	55.3	81.2
Total	F	10577	41.5	8803	42.3	83.2	438	36.2	4.1	478	45.3	4.5	5668	44.6	53.6
	M	14886	58.5	12003	57.7	80.6	772	63.8	5.2	577	54.7	3.9	7033	55.4	47.2

Figure 2. Percentage point differences in high intensity collaboration (>75% articles published in international collaboration in the scientist's individual publication portfolio for 2009–2018) for all Polish internationally productive university professors, by gender, collaboration type (four panels), and ASJC discipline.



Note: Vertical axes: positive results indicate female advantage in a given ASJC discipline, negative results indicate male advantage, and 0 indicates exactly the same distribution of collaboration by gender.

The propensity to collaborate internationally deserves a separate treatment. Overall, collaboration rates are low for all intensity levels (a finding which is in line with findings for Poland at the highly aggregated level: 35.8% of Scopus-indexed articles in 2018 were written in international collaboration, which is the lowest rate among the European Union member states).

In our sample, less than a half of female (45.4%) and male (47.4%) scientists had at least one paper published in internationally collaboration in the decade examined (total 46.6%; see Table 6 in Electronic Supplementary Material or ESM). At this low level of international collaboration intensity, in the three largest disciplines, the low rate of international collaboration was higher for male scientists by as much as 8.3 p.p. for MED and 4.8 p.p. for ENG (for AGRI, it is 2.6 p.p.). The largest male advantage is noted for disciplines such as PSYCH, PHARM, ENER, and BIO (10.8 p.p.). For HUM, SOC, and ECON, the three disciplines with the lowest rates of international collaboration, the male advantage is notable. HUM definitely has the lowest rate: only 8.3% female scientists and 11.9% male scientists have published at least one article in international co-authorship in the decade studied. At this low level of collaboration intensity, female advantage occurs in only four disciplines: two medium-sized (BUS and CHEMENG) and two small (DEC and DENT) disciplines. There are four disciplines in which international collaboration reaches the highest levels (in the range of 60%–70%): large disciplines such as BIO, CHEM, MATER, and PHYS as well as the small discipline IMMU. Male scientists are highly likely to collaborate internationally at this low level of intensity in CHEM (75%), PHYS (74.8%), and BIO (71.2%).

At a medium level of collaboration intensity (>50% articles published in international collaboration in the scientist's individual publication portfolio for 2009–2018), male advantage is overwhelming, with only four disciplines in which there is female advantage (again BUS, CHEMENG, DEC, and IMMU). However, as expected, collaboration rates dropped drastically compared to the low level of collaboration intensity to an average of 8.4% for female scientists and 10.8% for male scientists. Only in one discipline, at least a quarter of both male and female scientists attained this medium intensity level (PHYS: 31.0% and 25.3%, respectively) and only in several disciplines, at least 15% of scientists attained this level of intensity (females in BUS, DEC, and MATH; males in PSYCH).

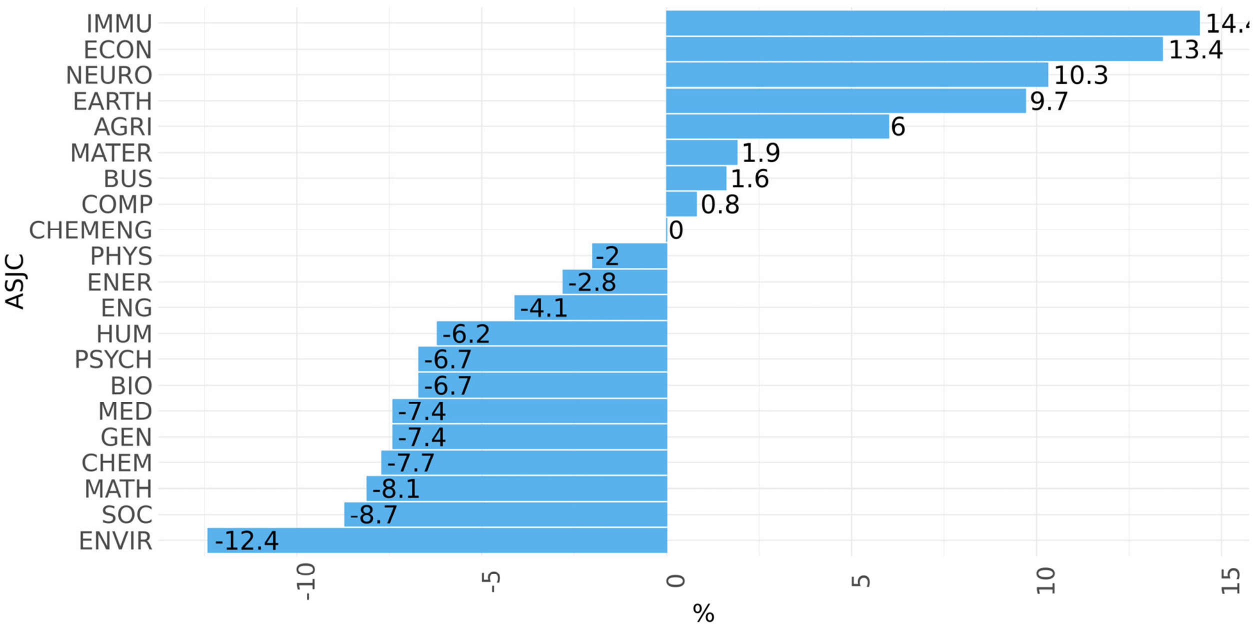
At the highest level of intensity (>75% articles published in international collaboration in the scientist's individual publication portfolio for 2009–2018), international collaboration rates dropped by half to 4.1% for females scientists and 5.2% for male scientists (Table 6 in ESM). For international collaboration, the female advantage was noted only for 9 disciplines and the male advantage was noted only for 13 disciplines, with CHEMENG exhibiting no gender difference (Figure 2, panel 4).

In addition, female advantage was noted for nine disciplines. For the three largest disciplines, the gender disparity is marginal. The largest male advantage was noted for MATH, PSYCH, and PHYS but the gender disparity is the largest for the two very small disciplines of IMMU and DEC with female advantage. It is only in selected disciplines that over 10% of scientists reach this high intensity international collaboration level: interestingly, it is BUS in the general cluster of soft fields (both females and males, 11.6% and 10.8%) and PHYS in the general cluster of hard fields (both females and males, 13.7% and 16.0%). In addition, this collaboration intensity was found for males in PSYCH and females in DEC.

To summarize, gender disparities are different for the four collaboration types analyzed. While the propensity to engage in general, national, and institutional collaboration is higher for female Polish scientists, the propensity to collaborate internationally is higher for male Polish scientists. However, as analyzed in detail above, there are substantial cross-disciplinary gender differentiations. For each collaboration type, there are specific disciplines in which the above overall picture does not fit the picture disaggregated to the level of disciplines.

Finally, gender disparities in high-intensity international collaboration at the disaggregated level of disciplines can also be examined from another perspective: disciplines with female advantage would be those in which the share of female scientists involved in high-intensity international collaborations would be higher than the overall share of female scientists in these disciplines. In the case of no gender disparities, the percentages of all females and females with high intensity in international collaboration must be exactly the same. Figure 3 indicates that there are eight such disciplines, including one of the three largest (AGRI), two middle-sized soft disciplines (BUS and ECON), the middle-sized hard disciplines of COMP, EARTH, and MATER, and two other very small disciplines (IMMU and NEURO). Apart from CHEMENG (with no gender disparity viewed from this angle), an advantage for males was found for all other disciplines, including in large disciplines such as MED and ENG, soft disciplines such as HUM and SOC, and hard disciplines such as PHYS, MATH, BIO, and CHEM. In this section, a small-sized discipline (n = 481 scientists) of chemical engineering, as studied from several angles, clearly emerges as a prototypical discipline with no gender disparities in international research collaboration.

Figure 3. Percentage differences between the overall share of female scientists and the share of female scientists involved in high-intensity international collaboration, all Polish internationally productive university professors, by ASJC discipline (three disciplines omitted: low counts) (in %).



4.2. Gender Distribution in International Collaboration: Cross-disciplinary Differences

In the next stage of analysis, it is useful to examine cross-disciplinary distribution differences in international collaboration by gender using boxplots. The first one is for all internationally productive Polish university professors (both collaborating and non-collaborating internationally—a total of 25,463—Figure 13) and only university professors collaborating internationally (that is, with at least a single article written in international collaboration—11,854 or 46.6%—within the decade studied). Thus, the international collaboration examined here is one that is beyond the three selected intensity levels. Scientists in disciplines such as PHYS, followed by BIO and CHEM, for both males and females, represent the highest average level of international collaboration (in their individual publication portfolios). For example, for male scientists in PHYS, 50% of authors have a share of at least 25% in internationally co-authored articles and 50% of authors have at most a 25% share of this publication type in their individual portfolios. In all three cases, the median value for males is higher than the median value for females. In certain disciplines, a median of zero for female scientists can be contrasted with a higher-than-zero median for male scientists (ENER and HEALTH), with the reverse being the case for DEC; further, for HUM, ECON, and SOC in the cluster of soft fields and for DEN in the cluster of hard fields, all quantiles up to the third quartile equal zero for both males and females. This effectively implies, as evident from Figure 4, that all observations in these disciplines are outliers— that is, there are a small number of scientists with atypically high shares of internationally co-authored articles in their portfolios within these disciplines.

The average level of intensity of international collaboration (at the level of individual scientists) by gender in the case of internationally collaborating scientists is presented in Figure 5. The median level reached approximately 50% for males in such disciplines in the cluster of soft fields as HUM, BUS, PSYCH, and SOC and in disciplines in the cluster of hard fields such as COMP, PHYS, and MATH. For females, the median reached similar levels in the same cluster of soft disciplines (HUM, BUS, PSYCH, and SOC) and in COMP, but not in PHYS and MATH. However, it is important to note the low numbers of scientists collaborating internationally in most disciplines (from 1,875 in MED and 1,362 in AGRI to 14 in DENT). As a general pattern, the average shares of internationally collaborative papers in the portfolios of male scientists are higher or equal to those of female scientists.

Figure 4. Distribution of international collaboration percentages in individual scientists' publishing portfolios for 2009-2018, all Polish internationally productive university professors (N = 25,463, both collaborating and non-collaborating internationally), by gender and ASJC discipline (in %).

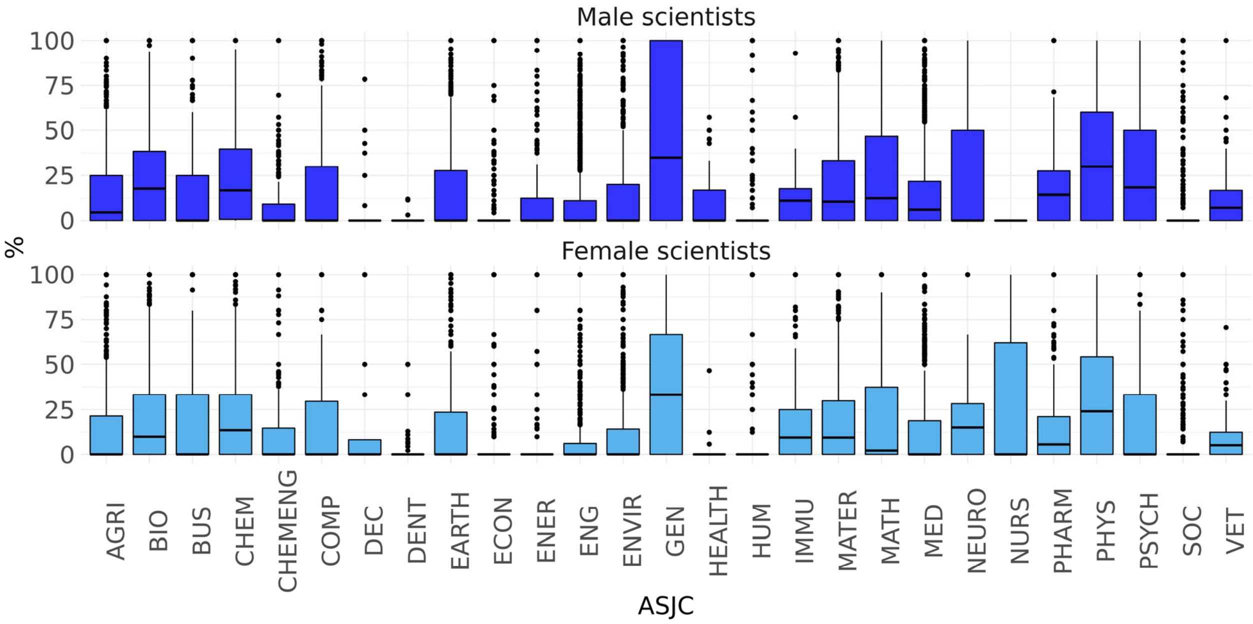
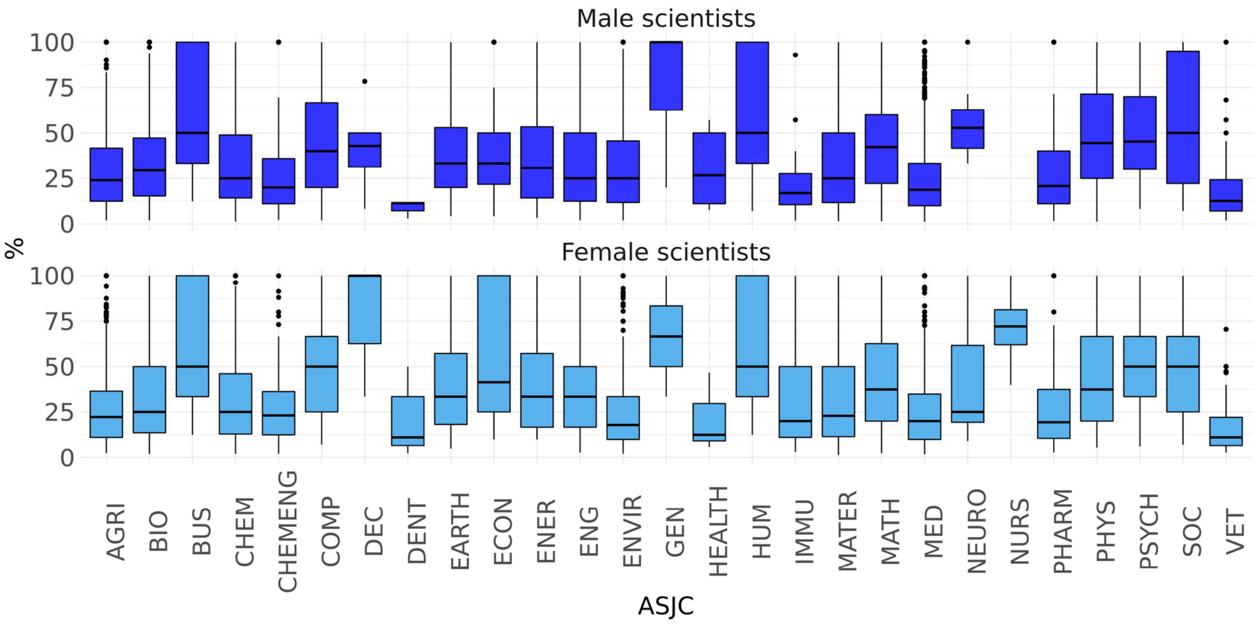


Figure 5. Distribution of international collaboration percentages in individual scientists' publishing portfolios for the period 2009–2018, internationally collaborating university professors only (N = 11,854; low collaboration intensity), by gender and ASJC discipline (in %).

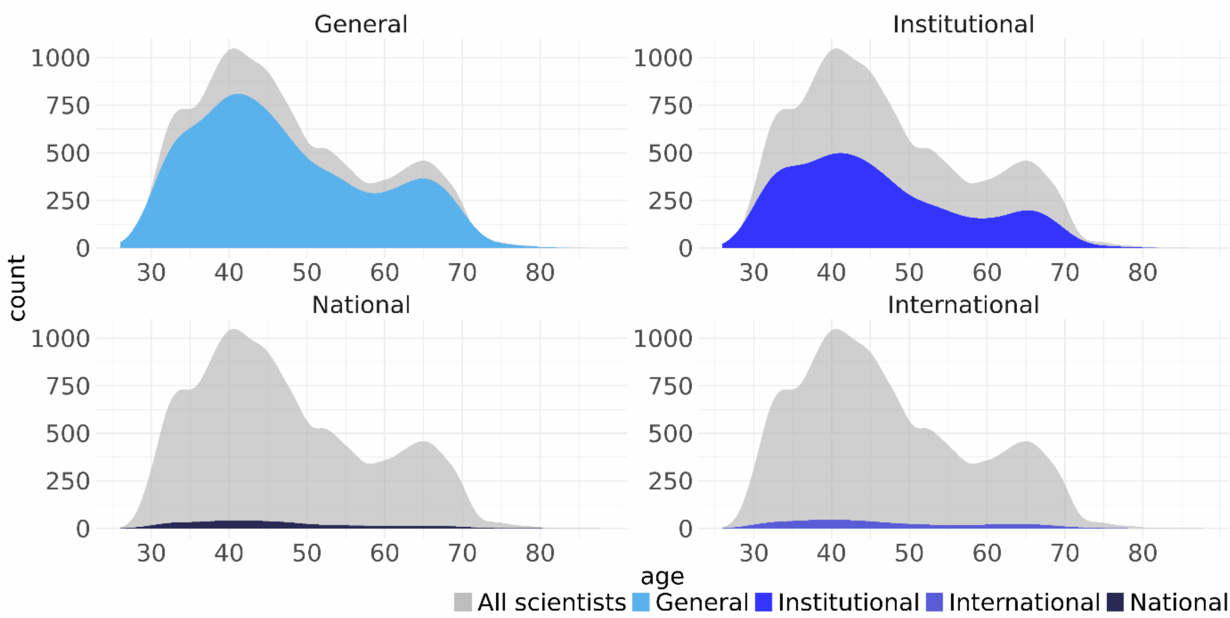


4.3. Gender, Collaboration, and Age Distribution

Gender disparity in international collaboration has been linked with age and Polish scientists are not an exception: female and male scientists differ substantially in their collaborative behavior by age. At the low intensity level, general collaboration is undertaken by the overwhelming majority of scientists of almost all ages; it is almost the same case for institutional collaboration. The difference is that institutional collaboration is less prevalent for scientists aged 40–45 and 60–65 years. Both national and international collaboration show the same pattern, with the largest number of scientists involved in national collaboration while they are aged 40–50 years and involved in international collaboration while aged 35–45 years.

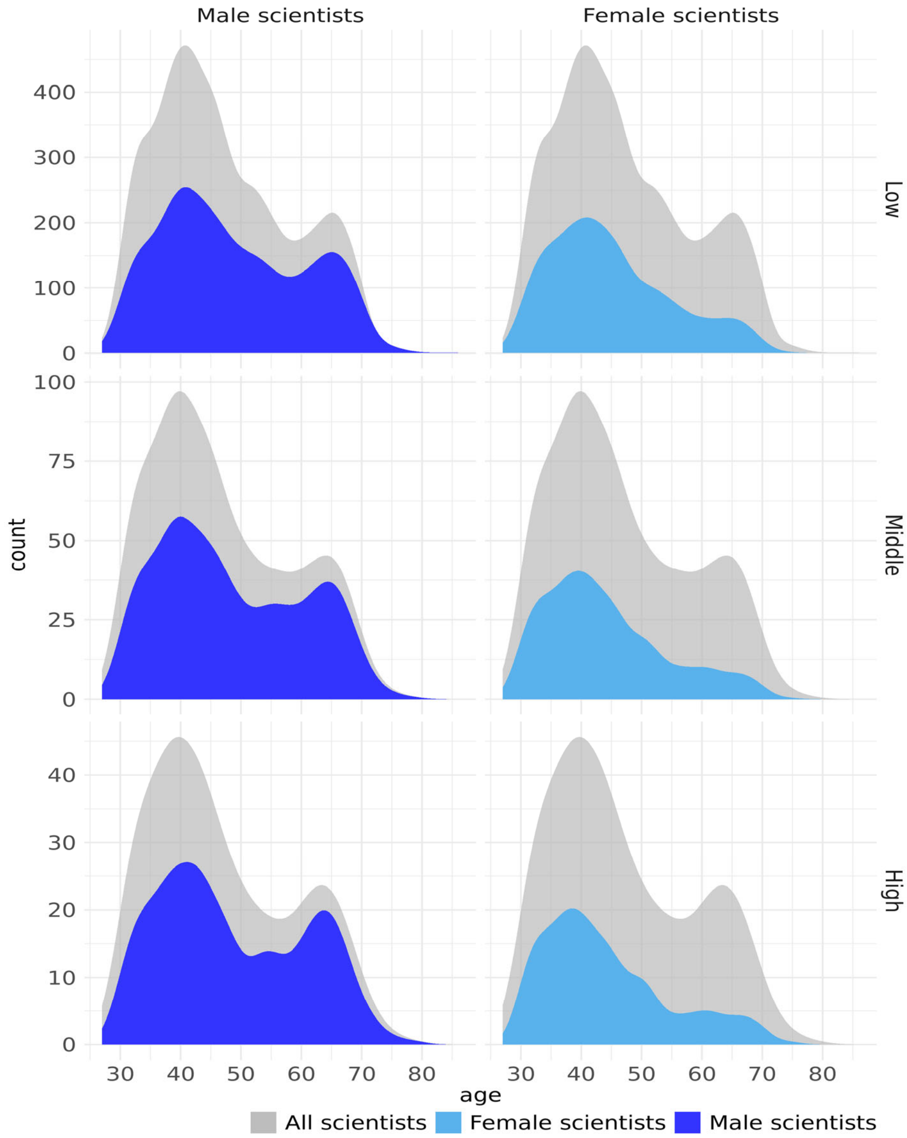
At the medium intensity level, general collaboration is equally present, institutional collaboration is slightly less prevalent, but both national and international collaboration are considerably less present, with visible participation of scientists aged 35–50 in national collaborations and 35–45 and 60–65 in international collaborations. The same pattern applies to the high-intensity level, with the peak of national collaboration for scientists being approximately 40 years and the two peaks of international collaboration for scientists being 40 and 65 years (Figure 6).

Figure 6. Age distribution of all internationally productive Polish university professors by collaboration type and professors involved in high-intensity collaboration only.



From a gender disparity perspective, interestingly, different levels of collaboration intensity indicate structurally similar patterns for international collaboration (albeit with different counts). At all three intensity levels, there are fewer females involved in collaboration and the peak for females is approximately 40 years, after which the number of females involved in international collaboration decreased drastically (Figure 7).

Figure 7. Age distribution by gender and collaboration intensity, all Polish internationally productive university professors, international collaboration only (low collaboration intensity, n = 11,854; medium collaboration intensity, n = 2,490; high collaboration intensity, n = 1,210).



Until about the age of 40, the gender differences in international research collaboration were marginal. The major cross-gender finding is that the number of male scientists involved in international collaboration after the age of 40 decreases, stops at about 60 (in the case of low intensity level, panel 1) and about 50 (in the cases of medium and high intensity levels, panels 2 and 3), and increases at about 60 years of age for another five years (until the age of 65, which is the retirement age for males outside the academic sector; full professors retire at the age of 70). Graphically, there are two peaks for male scientists, one for those aged about 40 and another for those aged about 65; however, there is only one peak for females at about 40 years of age. The pattern of gender disparities is exactly the same for all three international collaboration intensity levels. Although the patterns could also be tested for each discipline separately, this has not been done here on account of space limitations.

Consequently, it appears that is difficult for female scientists in the Polish higher education system to maintain a comparable level of international research collaboration with male scientists after the age of 40, regardless of the collaboration intensity level. At the highest level studied, we compared 438 female scientists with 772 male scientists (from our sample of 25,463 scientists). The major gender disparity in international research collaboration was found to occur in the case of scientists in their 50s and 60s: an ever-growing number (and proportion) of males who undertake international collaborations in the Polish academic science system is offset by a shrinking number (and proportion) of internationally collaborative females. There may be numerous reasons for this, but the social role of taking care of children and grandchildren and special early retirement arrangements for female scientists might be the main ones.

4.4. A Model Approach: Logistic Regression Analysis

In our modeling approach, the strength of the joint effect of traditional predictors of international research collaboration was tested. In this section, international research collaboration is understood as a low-intensity collaboration: having at least one article published in international collaboration in one's individual publication portfolio in the study period of 10 years from 2009 to 2018.

An analytical linear logistic model was constructed based on research literature, particularly predictive models built in Cummings and Finkelstein (2012), Rostan et al. (2014), Sooryamoorthy (2014), and Finkelstein and Sethi (2014). Three models were built: Model 1 for all scientists, Model 2 for male scientists, and Model 3 for female scientists. Estimating the odds ratios of being scientists defined as "internationally collaborative" was based on a set of independent variables: age, gender (reference category: female), academic position (reference category: assistant professor or the lowest position in our study), institutional type (reference category: the university or the institutional type with the lowest share of internationally collaborative articles, with traditional comprehensive universities having the highest share of them), productivity (the total number of articles indexed in the Scopus database in the study period 2009–2018), academic disciplines (defined as dominating Scopus ASJC categories, reference category: HUM, arts, and humanities, or the reference category with the lowest share of internationally collaborative articles in Poland). Importantly, all data in the models come from the integrated database—that is, originally from an official administrative and biographical database and a Scopus journal publication and citation database; consequently, the data are as objective as they can be (no self-declared data as in academic profession surveys are used in this model, thereby making its results more robust). The total number of observations used in the model was 25,463 (10,577 or 41,5% females and 14,886 or 58.5% males). The occurrence of potential multicollinearity was tested using an inverse correlation matrix. From among the variables analyzed, age turned out to be significantly correlated with a vector of other independent variables; however, age was entered into the model with the awareness that the estimate efficiency of this parameter was reduced.

For all scientists (Model 1), the model fit rather well to the data, as Nagelkerke's R^2 = 0.36. The data indicated that being a male scientist increases the chances of being internationally collaborative by merely 12.4% on average compared with female scientists (all other parameters being equal, see the details in Table 5). Interestingly, each year of age decreases that chance by 1.7% on average. In general, this is in line with descriptive statistics, as depicted in Figure 7. The implication is that younger scientists are more prone to collaborating internationally. Being an associate professor increases the odds by approximately one-fifth compared with being assistant professor, and being a full professor increases that chance by half.

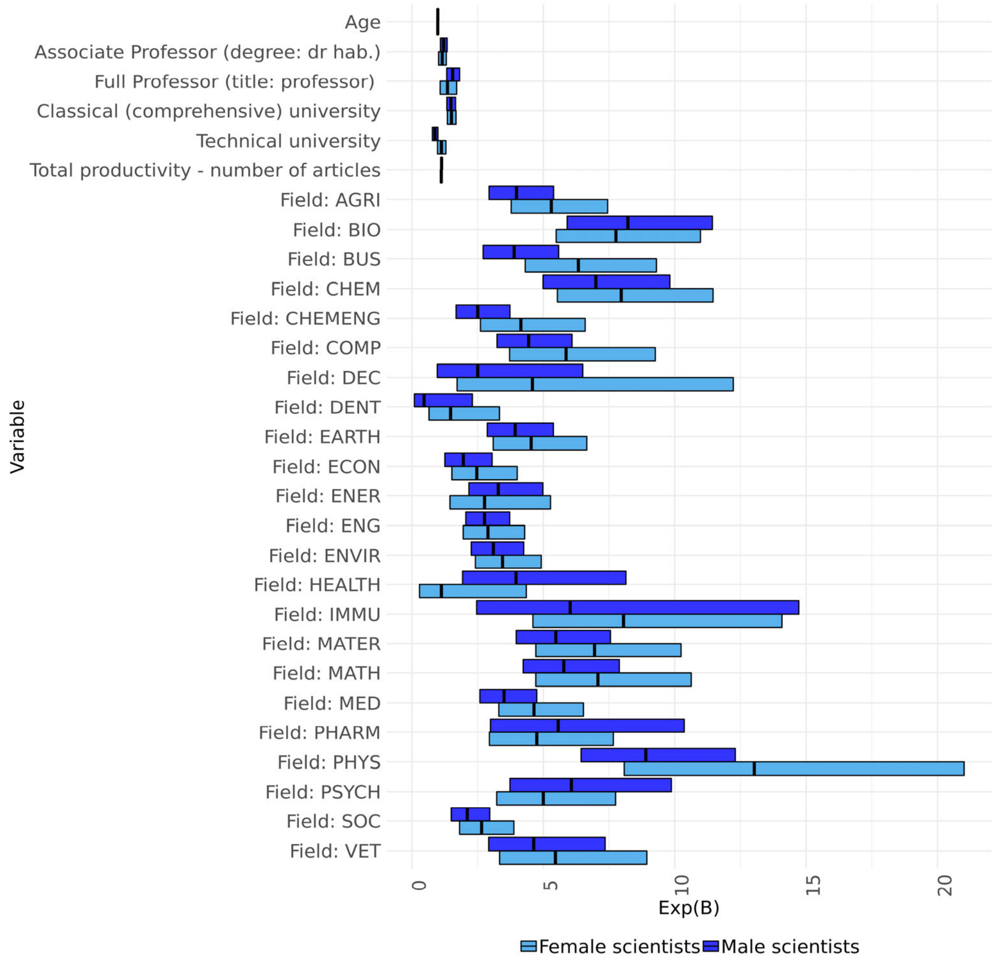
Further, working in a classical (comprehensive) university increases the odds by half compared with working in a university (defined as a non-comprehensive, such as university of economics or medical university). There was no differences in the odds for technical universities in the same comparison (comprehensive vs. non-comprehensive). The higher the total individual productivity, the greater the odds of being internationally collaborative. Each (Scopus-indexed) article published increases the odds by as much as 12.3%. Almost all disciplines (except dentistry) are characterized by significantly larger odds compared with arts and humanities. The most internationally collaborative disciplines are physics and astronomy (PHYS, with scientists 10 times more likely to collaborate than academics in arts and humanities), biochemistry, genetics and molecular biology, and chemistry (BIO and CHEM, both over seven times more likely). The least internationally collaborative disciplines are economics, econometrics, and finance (ECON); engineering (ENG); and health professions (with scientists between twice and three times more likely to collaborate than academics in arts and humanities).

Global literature on gender disparities in international research collaboration indicates that separate regression models for each gender might be useful. Certain variables might have a much stronger effect on only male or only female scientists, while others do not differentiate by gender. In our case, apparently, being a male full professor increases the odds of being internationally collaborative by half, while for females it increases the odds by approximately one-third. Further, publishing in the field of business, management, and accounting (BUS) increases the odds three times for males compared with five times for females (Models 2 and 3). However, prior to any further analysis, a hypothesis of a statistically significant difference between parameter values in the two models was tested. This hypothesis was tested by comparing 95% confidence intervals of logistic regression parameters for Model 2 (male scientists) and Model 3 (female scientists). If the intervals overlapped to any extent, it implied that the difference between parameter values is not significantly different from zero. This effectively implies that that the parameter values in the population studied are equal to each other and no gender difference can be shown. As we can see in Figure 8, confidence intervals for each parameter in regression models are indeed overlapping. In a very specific Polish case, a modeling approach to examining gender disparities in international research collaboration must result in a reliance on a single holistic model (Model 1) with gender as an independent variable.

Table 5. Odds ratio estimates of being internationally collaborative, three logistic regression models.

	Model 1	Model 2	Model 3
	All scientists R^2=0.355	Male scientists R^2=0.381	Female scientists R^2=0.320
Age	0.983***	0.984***	0.982***
Male	1.124***		
Associate Professor (degree: Dr. hab.)	1.183***	1.206***	1.152*
Full Professor (title: professor)	1.511***	1.55***	1.352**
Classical (comprehensive) university	1.496***	1.485***	1.504***
Technical university	0.962	0.876*	1.122
Total productivity - number of articles	1.123***	1.128***	1.118***
Field: AGRI	4.552***	3.977***	5.304***
Field: BIO	7.588***	8.218***	7.763***
Field: BUS	5.029***	3.889***	6.333***
Field: CHEM	7.419***	6.998***	7.963***
Field: CHEMENG	3.145***	2.503***	4.146***
Field: COMP	4.764***	4.441***	5.866***
Field: DEC	3.322***	2.504	4.584**
Field: DENT	1.039	0.461	1.471
Field: EARTH	4.226***	3.928***	4.537***
Field: ECON	2.174***	1.956**	2.468***
Field: ENER	3.23***	3.286***	2.763**
Field: ENG	2.873***	2.762***	2.893***
Field: ENVIR	3.257***	3.098***	3.447***
Field: HEALTH	2.97***	3.961***	1.121
Field: IMMU	6.946***	6.021***	8.051***
Field: MATER	6.086***	5.475***	6.948***
Field: MATH	6.357***	5.779***	7.077***
Field: MED	3.989***	3.505***	4.645***
Field: PHARM	4.653***	5.565***	4.752***
Field: PHYS	10.041***	8.899***	13.027***
Field: PSYCH	5.146***	6.068***	5***
Field: SOC	2.334***	2.104***	2.652***
Field: VET	4.999***	4.634***	5.46***
Constant	0.111***	0.128***	0.106***

Figure 8. 95% confidence intervals of logistic regression parameters for Model 2 (male scientists) and Model 3 (female scientists).



5. Summary of Findings, Discussion, and Conclusions

This research reveals substantial gender disparities in the collaboration patterns of Polish scientists. For the first time, such differences are systematically explored from a large-scale bibliometric perspective using “The Polish Science Observatory” database maintained by the authors. A detailed examination of our administrative, biographical, publication, and citation database of all Polish internationally productive university professors (N = 25,463, including 14,886 male and 10,577 female scientists; 159,943 articles written in the decade 2009–2018) leads to a number of conclusions.

First, while female scientists exhibit a higher rate of general, national, and institutional collaboration, male scientists exhibit a higher rate of international collaboration. Gender differences are statistically significant for all four major research collaboration types. This finding is critically important in explaining gender disparities in terms of impact, productivity, and access to large grants in view of the fundamental role of international collaboration in global science in comparison to any other collaboration type (Wagner, 2018; Gazni et al., 2012; Larivière et al., 2011).

However, second, an aggregated picture of gender disparities in international collaboration hides a much more nuanced picture of gender disparities by disciplines. There are substantial cross-disciplinary gender differentiations in international research collaboration (and in the three other collaboration types examined, as Thelwall and Maflahi 2019 show for national collaboration) and there are specific disciplines—notably computer science (COMP); business, management and accounting (BUS); economics, econometrics, and finance (ECON); agricultural and biological sciences (BIO); and earth and planetary sciences (EARTH) in which the above overall picture of male advantage does not fit the picture disaggregated to the level of the disciplines, with female advantage in these disciplines. Our findings support the general conclusions drawn in Abramo et al. (2013), which focused on Italian scientists: (1) male scientists exhibit higher collaboration rates in only one collaboration type—international collaboration; in all other collaboration types (general, national, and institutional), female scientists are more collaborative; (2) there is no one-fits-all answer to the question of gender disparity in international collaboration: differences by disciplines are fundamental and hidden in aggregated data; and (3) the power of individual-level data (with the scientist as a unit of analysis) is underestimated, and data sets for entire populations of scientists in other national systems are required to further explore collaboration patterns.

Third, we examined international research collaboration at three separate intensity levels (low, medium, and high), with male scientists dominating in international collaboration at each of them. While 47.4% of male scientists collaborate internationally at the low intensity level, 10.8% at the medium intensity level, and 5.2% at the high intensity level, for female scientists the rates are 45.4%, 8.4%, and 4.1%, respectively. However, interestingly, at each intensity level, there are specific disciplines with females collaborating more than males: for example, at the high intensity collaboration level, female scientists have higher collaboration rates in nine disciplines, including the above. There are also eight disciplines, including one of the three largest (AGRI, representing more than one in ten Polish scientists), two middle-sized soft disciplines (BUS and ECON), and three middle-sized hard disciplines of COMP, EARTH, and MATER, in which there are no gender disparities from any other perspective. In these disciplines, females representing high-intensity international collaboration are overrepresented compared with males. Moreover, a small-sized discipline (n = 481 scientists) of chemical engineering emerges as a discipline with no gender disparities in international research collaboration. All these disciplines together are populated by approximately 8,000 internationally productive university professors or by almost one-third of all those in our sample.

At the same time, however, the analysis of the average distribution of international collaboration percentages or rates in individual scientists' publishing portfolios by gender in the case of internationally collaborating scientists (N = 11,854) indicates that the rates are higher (or equal) for male scientists in all disciplines. A general pattern is that the lower the intensity level of international collaboration, the smaller the gender disparities. The highest gender disparities occur for most internationally collaborative scientists.

Gender disparity in international research collaboration in Poland is linked with age. Our major cross-gender finding is that the number of male and female scientists involved in international collaboration increases with age until the peak of about 40 years; after the age of 40, it decreases and then again increases at about 60 years for another 5 years, but only for male scientists. There are two peaks for male scientists (about 40 years and about 65 years) but only one peak for females (about 40), and the pattern is the same regardless of the collaboration intensity. The Polish system does not maintain international involvement of female scientists after 40 but does so for male scientists. Until about 40 years of age, the gender disparities in international research collaboration are marginal; thereafter, these grow, substantially increasing for scientists in their 50s and 60s when increasing numbers (and proportion) of internationally collaborative males are accompanied by shrinking numbers (and proportion) of internationally collaborative females.

The following conclusion can be drawn from linear logistic models: in the Polish case, a single holistic model with gender as an independent variable works better than separate models for the two genders, as indicated by the overlapping of 95% confidence intervals for the two separate models. Somehow surprisingly, being a male scientist increases the odds of being internationally collaborative by merely 12.4% (the odds would increase for being "highly collaborative" as the dependent variable). Further, age, academic position, institutional type, total productivity, and working in selected disciplines are significant. Age decreases the odds, as expected (each year of age by 1.7% on average). The likelihood for being internationally collaborative increases by half for full professors and by approximately one-fifth for associate professors, as it does for scientists working in comprehensive universities. The higher the individual total productivity, the greater the odds of being internationally collaborative (each article published increasing the odds by as much as 12.3%). Further, compared with arts and humanities, the likelihood also abruptly increases for scientists from the physics and astronomy fields (as much as 10 times) as well as from the biochemistry, genetics, molecular biology, and chemistry fields (7 times).

The current study is comprehensive and examines international collaboration in the context of all other collaboration types. By using a dataset with a combination of administrative, biographical, and bibliometric data for all internationally productive Polish scientists, the study goes beyond bibliometrics. By its methodological stance of using the individual scientist as the unit of analysis, the study avoids the pitfalls of aggregation and over-reliance on highly productive, highly internationalized scientists present in every system. As the unit of analysis in our study is a single scientist, the role of a female scientist with five internationally collaborative articles (out of 10) is exactly the same as the role of the one with 50 publications (out of 100): for both observations, the international collaboration intensity is 50%.

Our future research directions include wider (geographical coverage, possibly a global approach) and more fine-grained analyses of gender disparities in international research collaboration (hundreds of academic sub-disciplines, major institutional types, major strata of scientists by productivity and impact). Ideally, we could link large-scale survey data with the big data sources and treat both as complementary rather than substitute, as is done in a majority of the ongoing research in the area.

References

Abramo, G., D'Angelo, C.A., Murgia, G. (2013) Gender differences in research collaboration. *Journal of Informetrics*.7: 811-822.

Abramo, G., D'Angelo, C.A., & Caprasecca, A. (2009) The contribution of star scientists to overall sex differences in research productivity. *Scientometrics* 81(1): 137-156.

Abramo, G., D'Angelo, C.A., Di Costa, F. (2019) A gender analysis of top scientists' collaboration behavior: evidence from Italy. *Scientometrics.* Published online: 30 May 2019.

Abramo, G., D'Angelo, C.A., Solazzi, M. (2011) The relationship between scientists' research performance and the degree of internationalization of their research. *Scientometrics* 86: 629-643.

Abramo, G., D'Angelo, C.A., Di Costa, F. (2018) The collaboration behavior of top scientists. *Scientometrics.* On-line first: 29 November 2018.

Abramo., G., Aksnes, D.W., D'Angelo, C.A. (2019) Comparison of research productivity of Italian and Norwegian professors and universities. arXiv:1907.02043 [cs.DL]

Abramo., G., C. A. D'Angelo, M. Solazzi (2011b) Are researchers that collaborate more at the international level top performers? An investigation on the Italian university system. *Journal of Informetrics* 5: 204-213.

Ackers, L. (2008) Internationalization, mobility, and metrics: A new form of indirect discrimination? *Minerva* 46: 411-435.

Aksnes, D. W., Rørstad, K., Piro, F. N., & Sivertsen, G. (2011). Are female researchers less cited? A large scale study of Norwegian researchers. *Journal of the American Society for Information Science and Technology* 62(4): 628-636.

Aksnes, D.W., Piro, F.N., Rørstad, K. (2019) Gender gaps in international research collaboration: A bibliometric approach. *Scientometrics.* Published online: 13 June 2019.

Antonowicz, D. (2016) Digital players in an analogue world: Higher education in Poland in the post-massification era. In B. Jongbloed & H. Vossensteyn (eds.) *Access and expansion post-massification. Opportunities and barriers to further growth in higher education participation* (pp. 63–81). London: Routledge.

Antonowicz, D., Kwiek, M., & Westerheijden, D. F. (2017). The government response to the private sector expansion in Poland. In H. de Boer, J. File, J. Huisman, M. Seeber, M. Vukasovic, & D. F. Westerheijden (eds.) *Policy analysis of structural reforms in higher education* (pp. 119–138). Dordrecht: Springer.

Araújo E.B., Araújo N.A.M., Moreira A.A., Herrmann H.J., Andrade J.S. Jr. (2017) Gender differences in scientific collaborations: Women are more egalitarian than men. *PLoS ONE* 12(5): e0176791. https://doi.org/10.1371/journal.pone.0176791

Bieliński, J. & Tomczyńska, A. (2018) The ethos of science in contemporary Poland. *Minerva*. On-line first: https://link.springer.com/article/10.1007/s11024-018-9365-1.

Boschini, A., & Sjögren, A. (2007) Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics 25*(2): 325-365.

Bozeman, B., & Gaughan, M. (2011) How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy* 40(10): 1393-1402.

Cummings, W. K. & Finkelstein, M. J. (2012) *Scholars in the Changing American Academy. New Contexts, New Rules and New Roles.* Dordrecht: Springer.

Enamorado, T., Fifield, B., Imai, K., (2019) Using a Probabilistic Model to Assist Merging of Large-Scale Administrative Records, *American Political Science Review* (2019) 113, 2, 353-371.

Etzkowitz, H., Kemelgor, C., & Uzzi, B. (2000) *Athena Unbound: The Advancement of Women in Science and Technology*. Cambridge: Cambridge University Press.

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(1): 113-141. |

Fellegi I. P., Sunter A. B., (1969) A theory for record linkage. *Journal of the American Statistical Association* 64: 1183-1210.

Finkelstein, M. & Sethi, W. (2014) Patterns of faculty internationalization: A predictive model. In F. Huang, M. Finkelstein & M. Rostan (eds.) *The Internationalization of the Academy. Changes, Realities and Prospects* (pp. 237-258). Dordrecht: Springer.

Fox, M. F. (1991). Gender, environmental milieu, and productivity. In H. Zuckerman, J. Cole, & J. Bruer (eds.) *The outer circle: Women in the scientific community* (pp. 188-204). New York: W.W. Norton and Company.

Fox, M.F., Realff, M.L., Rueda, D.R., & Morn, J. (2017) International research collaboration among women engineers: Frequency and perceived barriers, by regions. *Journal of Technology Transfer* 42(6): 1292-1306.

Frehill, L. M., & Zippel, K. (2010). Gender and international collaborations of academic scientists and engineers: Findings from the survey of doctorate recipients, 2006. *Journal of the Washington Academy of Sciences* 97(1): 49-69.

Gazni, A, Sugimoto, C.R, and Didegah, F. (2012) Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology* 63(2): 323-335.

Halevi, G. (2019) Bibliometric studies on gender disparities in science. In W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (eds.) *Springer Handbook of Science and Technology Indicators* (pp 563-580). Cham: Springer.

Harron K., Dibben C., Boyd J., Hjern A., Azimaee M., Barreto M. L., Goldstein H. (2017) Challenges in administrative data linkage for research., *Big Data & Society*, July–December 2017: 1-12.

Herzog T. N., Scheuren F. J., Winkler W. E., 2007, *Data Quality and Record Linkage Techniques*, Dordrecht: Springer.

Jadidi, M., Karimi, F., Lietz, H., & Wagner, C. (2018) Gender disparities in science? Dropout, productivity, collaborations, and success of male and female computer scientists. *Advances in Complex Systems* 21(3-4): 1750011.

Jaro, M. A. (1989) Advances in record linkage methodology as applied to the 1985 census of Tampa Florida. *Journal of the American Statistical Association*. 84 (406): 414-420.

Jung, J., Kooij, R., & Teichler, U. (2014) Internationalization and the new generation of academics. In F. Huang, M. Finkelstein and M. Rostan (eds.) *The internationalization of the academy. Changes, realities and prospects* (pp. 207-236). Dordrecht: Springer.

Kosmulski, M. (2015) Gender disparity in Polish science by year (1975–2014) and by discipline. *Journal of Informetrics* 9: 658-666.

Kulczycki, E., Korzeń, M., & Korytkowski, P. (2017) Toward an excellence-based research funding system: Evidence from Poland. *Journal of Informetrics* 11(1): 282-298.

Kwiek, M. (2012). Changing Higher Education Policies: From the Deistitutionalization to the Reinstitutionalization of the Research Mission in Polish Universities. *Science and Public Policy*. Vol. 39. Issue 5. 641-654.

Kwiek, M. (2015a). The internationalization of research in Europe. A quantitative study of 11 national systems from a micro-level perspective. *Journal of Studies in International Education*, *19*(2), 341-359.
Kwiek, M. (2015b). The unfading power of collegiality? University governance in Poland in a European comparative and quantitative perspective. *International Journal of Educational Development* 43, 77-89.
Kwiek, M. (2016). The European research elite: A cross-national study of highly productive academics across 11 European systems. *Higher Education*, *71*(3), 379-397.
Kwiek, M. (2018a). International research collaboration and international research orientation: Comparative findings about European academics. *Journal of Studies in International Education*. 22(2). 1-25.
Kwiek, M. (2018b). Academic top earners. Research productivity, prestige generation and salary patterns in European universities. *Science and Public Policy*. 45(1). 1–13.
Kwiek, M. (2018c). High Research Productivity in Vertically Undifferentiated Higher Education Systems: Who Are the Top Performers?. *Scientometrics.* 115(1). 415–462.
Kwiek, M. (2019a). *Changing European Academics. A Comparative Study of Social Stratfication, Work Patterns and Research Productivity.* London and New York: Routledge.
Kwiek, M. (2019b). "Social Stratification in Higher Education: What It Means at the Micro-Level of the Individual Academic Scientist". *Higher Education Quarterly.* Vol. 73. Issue 4. 419-444.
Kyvik, S., & Aksnes, D. W. (2015) Explaining the increase in publication productivity among academic staff: A generational perspective. *Studies in Higher Education*, 40: 1438-1453.
Larivière, V., Sugimoto, C. R., Chaoquin, N., Gingras, Y., & Cronin, B. (2013) Global gender disparities in science. *Nature* 504: 211-213.
Larivière, V., Vignola-Gagné, 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.
Latour B. & Woolgar S. (1986) *Laboratory Life. The Construction of Scientific Facts*. Princeton: Princeton University Press.
Leahey, E. (2006). Gender differences in productivity. Research specialization as a missing link. Gender & Society 20(6): 754-780.
Long, J. S. (1990). The origins of sex differences in science. Social Forces 68(4): 1297-1315.
Maddi, A., Larivière, V., Gingras, Y. (2019) Man-woman collaboration behaviors and scientific visibility: does gender affect the academic impact in economics and management? *Proceedings of the 17th International Conference on Scientometrics & Informetrics* (pp. 1687-1697). September 2-5, 2019.
Mayer, S. J., & Rathmann, J. M. K. (2018) How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions. *Scientometrics* 117(3): 1663-1693.
McDowell, J. M., Larry, D., Singell, Jr., & Stater, M. (2006) Two to tango? Gender differences in the decisions to publish and coauthor. *Economic Inquiry* 44(1): 153-168.
McPheron, M., Smith-Lovin, L., and Cool, J.M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27: 415-444.
Melkers, J., & Kiopa, A. (2010) The social capital of global ties in science: The added value of international collaboration. *Review of Policy Research* 27(4): 389-414.
Nielsen, M. W. (2016) Gender inequality and research performance: Moving beyond individual-meritocratic explanations of academic advancement. *Studies in Higher Education* 41(11): 2044-2060.

O'Brien, K.R., Hapgood, K.P. (2012) The academic jungle: ecosystem modeling reveals why women are driven out of research. *Oikos* 121: 999-1004.
Ostrowicka, H., & Stankiewicz, L. (2018) The truths of business and the lies of academia: The order of discourse on higher education in Poland. *Higher Education Research & Development.* Online first: https://doi.org/10.1080/07294360.2018.1545746.
Piro, F. N., Rørstad, K., & Aksnes, D. W. (2016) How do prolific professors influence the citation impact of their university departments? *Scientometrics* 107(3): 941-961.
Rhoten, D., & Pfirman, S. (2007) Women in interdisciplinary science: Exploring preferences and consequences. *Research Policy* 36(1): 56-75.
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: 317-333.
Rostan, M. (2015) International aspects of academic work and career at the beginning of the twenty-first century. In U. Teichler and W.K. Cummings (eds.) *Forming, recruiting, and managing the academic profession* (pp. 241-258). Dordrecht: Springer.
Rostan, M., Ceravolo, F.A. & Metcalfe, S. A. (2014) The internationalization of research. In F. Huang, M. Finkelstein, & M. Rostan (eds.), *The internationalization of the academy. Changes, realities and prospects* (pp. 119-144). Dordrecht: Springer.
Shauman, K. A., & Xie, Y. (1996) Geographic mobility of scientists: Sex differences and family constraints. *Demography* 33(4): 455-468.
Siemienska, R. (2007) The puzzle of gender research productivity in Polish universities. In R. Siemienska and A. Zimmer (eds.), *Gendered career trajectories in academia in cross-national perspectives* (pp. 241-266). Warsaw: Scholar.
Sooryamoorthy, R. (2014) Publication productivity and collaboration of researchers in South Africa: New empirical evidence. *Scientometrics* 98: 531-545.
Thelwall, M., Maflahi, N. (2019) Academic collaboration rates and citation associations vary substantially between countries and fields. arXiv:1910.00789.
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.
Urbanek, P. (2018) Reform of the higher education system in Poland from the perspective of agency theory. *European Journal of Higher Education.* Online first, December 20, 2018, https://doi.org/10.1080/21568235.2018.1560344.
Vabø, A., Padilla-Gonzales, L.E., Waagene, E., Naess, T. (2014) Gender and faculty internationalization. In F. Huang, M. Finkelstein & M. Rostan (eds.), *The Internationalization of the Academy. Changes, Realities and Prospects* (pp. 183-206). Dordrecht: Springer.
Villanueva-Felez, A., Molas-Gallart, M., & Escribá-Esteve, A. (2013) Measuring personal networks and their relationship with scientific production. *Minerva* 51: 465-483.
Wagner, C.S. (2018) *The Collaborative Era in Science. Governing the Network.* Cham: Palgrave Macmillan.
Wang, Y.S., Lee, C.J., West, J.D, Bergstrom, C.T., and Erosheva, E.A. (2019) Gender-based homophily in collaborations across a heterogeneous scholarly landscape. arXiv:1909.01284 [stat.AP].
Winkler W. (1990) String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. *Proceedings of the Section on Survey Research Methods*, *American Statistical Association* 354-359.
Winkler W. 2006. Overview of record linkage and current research directions, *Research Report Series (Statistics #2006-2),* US Census Bureau, Washington.
Wolszczak-Derlacz J. & Parteka, A. (2010) *Scientific Productivity of Public Higher Education Institutions in Poland. A Comparative Bibliometric Analysis.* Warsaw: Ernst and Young.

Wuchty, S., Jones, B.F., and Uzzi, B. (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827): 1036-1039.
Yemini, M. (2019) International research collaborations as perceived by top-performing scholars. *Journal of Studies in International Education.* Online first, November 9, 2019, pp. 1-16.
Zeng, X. H. T., Duch, J., Sales-Pardo, M., Moreira, J. A. G., Radicchi, F., Ribeiro, H. V., et al. (2016) Differences in collaboration patterns across discipline, career stage, and gender. *PLoS Biology,* 14(11): e1002573.
Zippel, K. (2017) *Women in Global Science.* Stanford: Stanford University Press.

Electronic Supplementary Material

Table 6. Percentage differences in international collaboration; all Polish internationally productive university professors; by collaboration intensity (low, medium, and high), gender, and discipline (in %).

		Scientists: Total		Scientists: minimum 1 article (Low intensity collaboration)			Scientists: > 50% articles (Middle intensity collaboration)			Scientists: > 75% articles (High intensity collaboration)		
		N	% col	N	% col	% row	N	% col	% row	N	% col	% row
AGRI	Female	1444	53.4	710	52.1	49.2	88	46.6	6.1	44	59.5	3
	Male	1258	46.6	652	47.9	51.8	101	53.4	8	30	40.5	2.4
BIO	Female	1068	60	645	56	60.4	119	55.1	11.1	48	53.3	4.5
	Male	712	40	507	44	71.2	97	44.9	13.6	42	46.7	5.9
BUS	Female	372	52.1	125	55.6	33.6	56	54.4	15.1	43	53.8	11.6
	Male	342	47.9	100	44.4	29.2	47	45.6	13.7	37	46.3	10.8
CHEM	Female	756	51.3	514	48.8	68	103	48.8	13.6	34	43.6	4.5
	Male	719	48.7	539	51.2	75	108	51.2	15	44	56.4	6.1
CHEM-ENG	Female	185	38.5	73	44.2	39.5	7	38.9	3.8	5	38.5	2.7
	Male	296	61.5	92	55.8	31.1	11	61.1	3.7	8	61.5	2.7
COMP	Female	170	16.5	61	15.6	35.9	18	13.8	10.6	13	17.3	7.6
	Male	860	83.5	331	84.4	38.5	112	86.2	13	62	82.7	7.2
DEC	Female	24	44.4	6	46.2	25	4	80	16.7	4	80	16.7
	Male	30	55.6	7	53.8	23.3	1	20	3.3	1	20	3.3
DENT	Female	57	76	11	78.6	19.3	0	0	0	0	0	0
	Male	18	24	3	21.4	16.7	0	0	0	0	0	0
EARTH	Female	385	33.4	149	30.9	38.7	40	31.3	10.4	28	43.1	7.3
	Male	769	66.6	333	69.1	43.3	88	68.8	11.4	37	56.9	4.8
ECON	Female	186	49.1	36	44.4	19.4	13	56.5	7	10	62.5	5.4
	Male	193	50.9	45	55.6	23.3	10	43.5	5.2	6	37.5	3.1
ENER	Female	82	27.8	17	19.8	20.7	5	20.8	6.1	4	25	4.9
	Male	213	72.2	69	80.2	32.4	19	79.2	8.9	12	75	5.6
ENG	Female	501	14.9	133	13	26.5	26	14	5.2	10	10.9	2
	Male	2857	85.1	894	87	31.3	160	86	5.6	82	89.1	2.9
ENVIR	Female	848	50.5	347	49.2	40.9	36	38.3	4.2	16	38.1	1.9
	Male	832	49.5	358	50.8	43	58	61.7	7	26	61.9	3.1
HEALTH	Female	23	34.3	3	14.3	13	0	0	0	0	0	0
	Male	44	65.7	18	85.7	40.9	1	100	2.3	0	0	0
HUM	Female	527	49.8	44	41.1	8.3	18	41.9	3.4	17	43.6	3.2
	Male	531	50.2	63	58.9	11.9	25	58.1	4.7	22	56.4	4.1
IMMU	Female	90	75.6	57	74	63.3	14	87.5	15.6	9	90	10
	Male	29	24.4	20	26	69	2	12.5	6.9	1	10	3.4
MATER	Female	495	33.9	309	33.7	62.4	67	33.7	13.5	29	35.8	5.9
	Male	967	66.1	608	66.3	62.9	132	66.3	13.7	52	64.2	5.4
MATH	Female	259	25.2	130	23.2	50.2	44	23.2	17	13	17.1	5
	Male	767	74.8	430	76.8	56.1	146	76.8	19	63	82.9	8.2
MED	Female	1920	53.7	934	49.8	48.6	110	48.9	5.7	44	46.3	2.3
	Male	1654	46.3	941	50.2	56.9	115	51.1	7	51	53.7	3.1
PHARM	Female	169	66.5	96	61.5	56.8	14	60.9	8.3	3	50	1.8
	Male	85	33.5	60	38.5	70.6	9	39.1	10.6	3	50	3.5
PHYS	Female	182	16.6	133	16.3	73.1	46	13.9	25.3	25	14.5	13.7
	Male	916	83.4	685	83.7	74.8	284	86.1	31	147	85.5	16
PSYCH	Female	194	63.8	72	54.5	37.1	28	53.8	14.4	16	57.1	8.2
	Male	110	36.2	60	45.5	54.5	24	46.2	21.8	12	42.9	10.9
SOC	Female	494	49.8	105	47.5	21.3	31	38.8	6.3	23	41.1	4.7
	Male	498	50.2	116	52.5	23.3	49	61.3	9.8	33	58.9	6.6
VET	Female	146	44	87	40.8	59.6	1	25	0.7	0	0	0
	Male	186	56	126	59.2	67.7	3	75	1.6	1	100	0.5
Total	Female	10577	41.5	4797	40.5	45.4	888	35.7	8.4	438	36.2	4.1
	Male	14886	58.5	7057	59.5	47.4	1602	64.3	10.8	772	63.8	5.2