Man-Woman Collaboration Patterns in Science: Lessons from a Study of 25,000 University Professors

Centre for Global Higher Education (CGHE),

University of Oxford, December 8, 2020

Professor Marek Kwiek
Center for Public Policy Studies
University of Poznan, Poland
kwiekm@amu.edu.pl

Introduction (1)

- Science as a largely **collaborative enterprise**.
- Traditional picture is: (male and female) scientists collaborating **internationally, nationally, institutionally or not collaborating**. This in *not* my topic .
- My topic is collaboration **across genders –** not countries and institutions. **All-male, all-female,** and **mixed-sex collaboration** and implications of the patterns.
- The dominating view in literature is that, on average **males collaborate more often with males**. And **females** collaborate more often **with females**.
- I am **testing** this hypothesis using a large-scale dataset and new variables.

Introduction – Homophily (1)

- **The homophily principle** maintains that "**similarity breeds connection**": consequently, **personal networks** are **homogeneous** with regard to many sociodemographic and personal characteristics (such as age, ethnic origin, class origin, wealth, education; here: **gender**).
- Male-male/female-female/male-female collaborations.
- **Positively**: homophily in science is reported to **simplify communication**, enhance the **predictability** of behavior, entail **reciprocity** in collaboration, and **increase trust** between collaborating parties (McPherson et al., 2001; Kegen, 2013).
- **Negatively**: homophily in science is reported to "**limit people's social worlds in a way that has powerful implications** for the **information** they receive, the **attitudes** they form, and the **interactions** they experience" (McPherson, Smith-Lovin, & Cook, 2001).
- If science is **increasingly collaborative**, then **the homophily principle increasingly matters!**
- **Research collaboration** in science (or gender co-authorship patterns) provide **fertile ground to test the homophily principle**.

Introduction – Homophily (2)

- According to the homophily principle, "**birds of a feather flock together**" (McPherson et al., 2001).

- If homophily means "**the tendency of people to choose to interact with similar others**," then gender-based homophily in this research means **Polish male scientists disproportionately co-authoring with other male scientists.** (And Polish female scientists disproportionately co-authoring with other female scientists).

- While the **behavior of collaborators** might be more predictable and collaboration potentially less **costly** and less **risky**, **gender homophily might exclude women** from **powerful informal networks**.

Introduction – Homophily (3)

- **Embeddedness in (informal) academic social networks** is **crucial** both for doing research and for achieving a successful career.
- "**Networks matter**. Producing high-quality work is not sufficient for research to **gain the attention** of the widest number of scholars or **have the greatest impact**" (Maliniak et al., 2013, p. 918).
- **Collaboration patterns found** so far suggest that men tend to co-author with men and women with women. **Female** scientists are reported to exhibit **stronger gender homophily than male scientists**: females collaborate more often with females than males with males.

  (Ghiasi et al., 2018; Potthoff & Zimmermann, 2017; Lerchenmueller et al., 2019; Kegen, 2013; Wang et al., 2019; Boschini & Sjögren, 2007; Jadidid, Karimi, Lietz, & Wagner, 2018; Lerchenmueller et al., 2019).
- Evidence from co-authorship patterns indicates that **team formation in academic publishing is not gender-neutral**: there is powerful **gender sorting** in team formation (Boschini & Sjögren, 2007).
- **Team formation in academia** (publishing with co-authors) is **voluntary**: researchers team up when they think that **they are better-off** collaborating than publishing alone.
- The teams formed (or the articles published) reflect **individual tastes and perceptions of the returns to collaboration** (as well as the costs of coordination).

Introduction – Homophily, Networks, & Academic Careers

- However, collaboration types **influence career progress of males and females differently**!
- **Excessive gender homophily among women**, while supportive for early-career females, may also **harm** their careers.
- Women may place themselves at a **disadvantage** when collaborating disproportionately with other women because "women tend to be part of **less resource-rich and influential networks** or because women's work may **receive less attention** than men's" (Lerchenmueller et al., 2019).
- This is **not the Polish case**, though!
- Polish female scientists tend to *avoid* publishing **exclusively** or **massively** with other female scientists (at all levels of their careers and for all age groups).

The Gender Context in Science (1)

- "**Research collaboration**" and "**women in science**" have been widely studied for about half a century.
- However, the **gender context of academic science has changed substantially** (Halevi, 2019; Larivière et al. 2013):
  - **more female scientists** are entering and remaining in the higher education sector every decade (Elsevier, 2018) and
  - **female scientists** are increasingly occupying **high academic positions** (Madison & Fahlman, 2020)
  - **female scientists** are present in ever **greater proportions** (Zippel, 2017) and
  - **female scientists** are present in an increasing number of **disciplines** (Diezmann & Grieshaber, 2019).
- **Females are no longer "newcomers" to science**: massively involved, fully legitimate participants in global science.

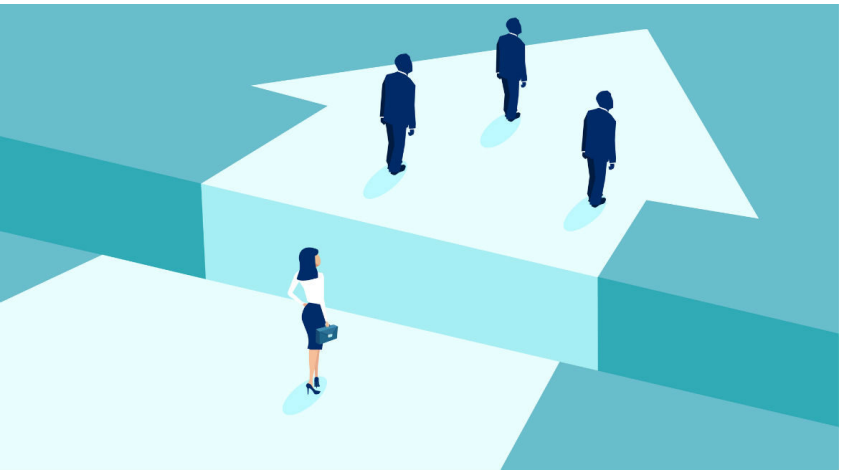
Major points:

- **Science has gender. Publications have gender. And both can be determined.**
- New bibliometric literatures applying the **various gender-determination methods** to authors and authorships bring **new data-driven insights** (Halevi, 2019; Elsevier, 2020).
- Gender disparities in science have been changing and literatures have become **much less based on anecdotal and localized studies**.

The Gender Context in Science, Poland (2)

- In **Poland**, females constitute **a substantial, highly productive, and highly internationalized** part of the academic workforce.
- Formerly communist European countries generally exhibit **greater gender parity than the world and the OECD averages**.
- Poland has a **high proportion of full professors** (29.82%, 2018).
- Females are almost half (46.10%) of the entire full-time academic workforce.
- **Female participation about or exceeds 50% in** 10/24 ASJC Scopus disciplines studied (STEM & non-STEM):
  - social sciences and humanities (traditionally);
  - business, economics, and econometrics;
  - agricultural and biological sciences;
  - medicine;
  - chemistry;
  - biochemistry, genetics and molecular biology; and
  - psychology.

The Gender Context in Science: The Darker Side (3)



Female scientists globally, compared with males, still:

- occupy **more junior positions** with **lower salaries**,
- receive **fewer citations**,
- are more often in **non-tenure-track** and **teaching-only positions**,
- receive **less grant money**,
- are **promoted more slowly**,
- are less likely to be listed as either **first or last author on a paper,**
- tend to be **less involved in international collaboration**,
- **suffer from "biased attention"** to their work.

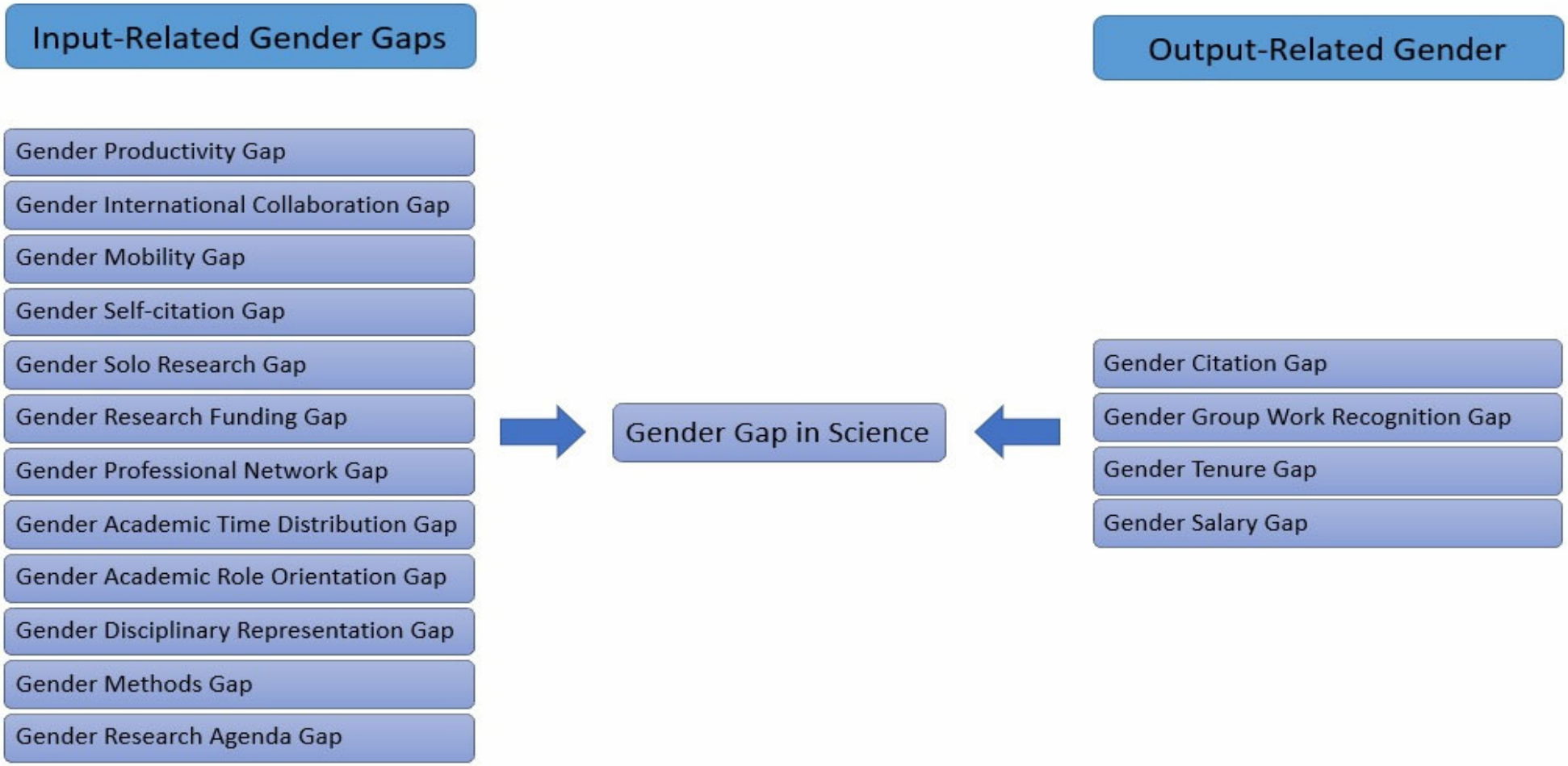
(see Holman & Morandin, 2019; Halevi, 2019; Larivière et al., 2013; Larivière et al., 2011; Aksnes, Rørstad, Piro, & Sivertsen, 2011; Aksnes, Piro, & Rørstad, 2019; Huang et al., 2020; Maddi, Larivière, & Gingras, 2019; Potthoff & Zimmermann, 2017; Fell & König, 2016; van den Besselaar & Sandström, 2016; Nielsen, 2016; Lerchenmueller, Hoisl, & Schmallenbach, 2019; Elsevier, 2020 and Elsevier, 2018).

The Gender Context in Science: Citation Gap (3)

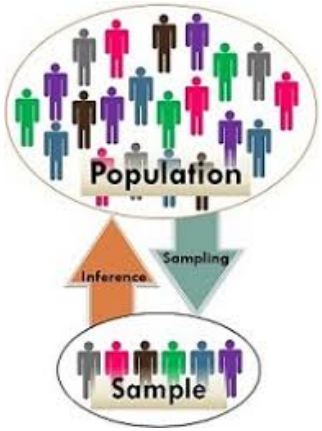
For instance, the **gender citation gap:**

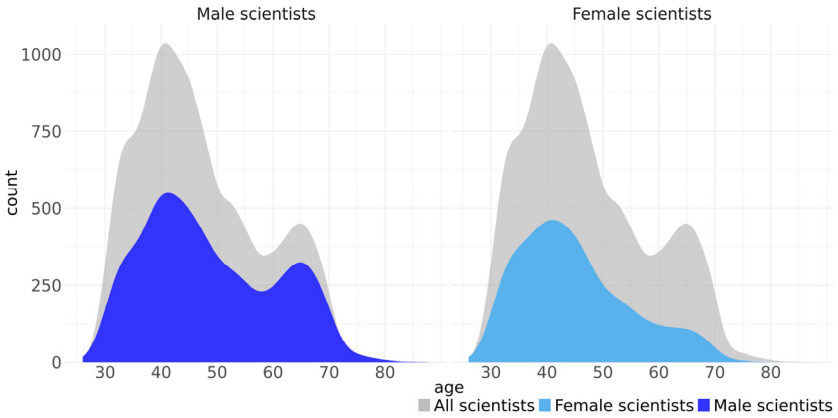
- The **gender citation gap matters** because citations are one of the chief metrics used in academia to **evaluate a scholar's performance and influence** and to distribute resources.
- Citations are used as a "**reward currency in science**" upon which **decisions on all major aspects** of an academic career are often based (Ghiasi et al.. 2018).
- Moreover, gender influences **the attribution of credit for group work**.
- **Co-authorship matters differently for tenure** for men and women. Women are less likely to receive tenure **the more they co-author** (Sarsons et al. 2020).
- Should (Polish) **female scientists excessively (or exclusively)** collaborate with females, they would be **losing** in terms of their **career progression**. Will be checked!

The Gender Context in Science: Gender Gaps (5)



Sample



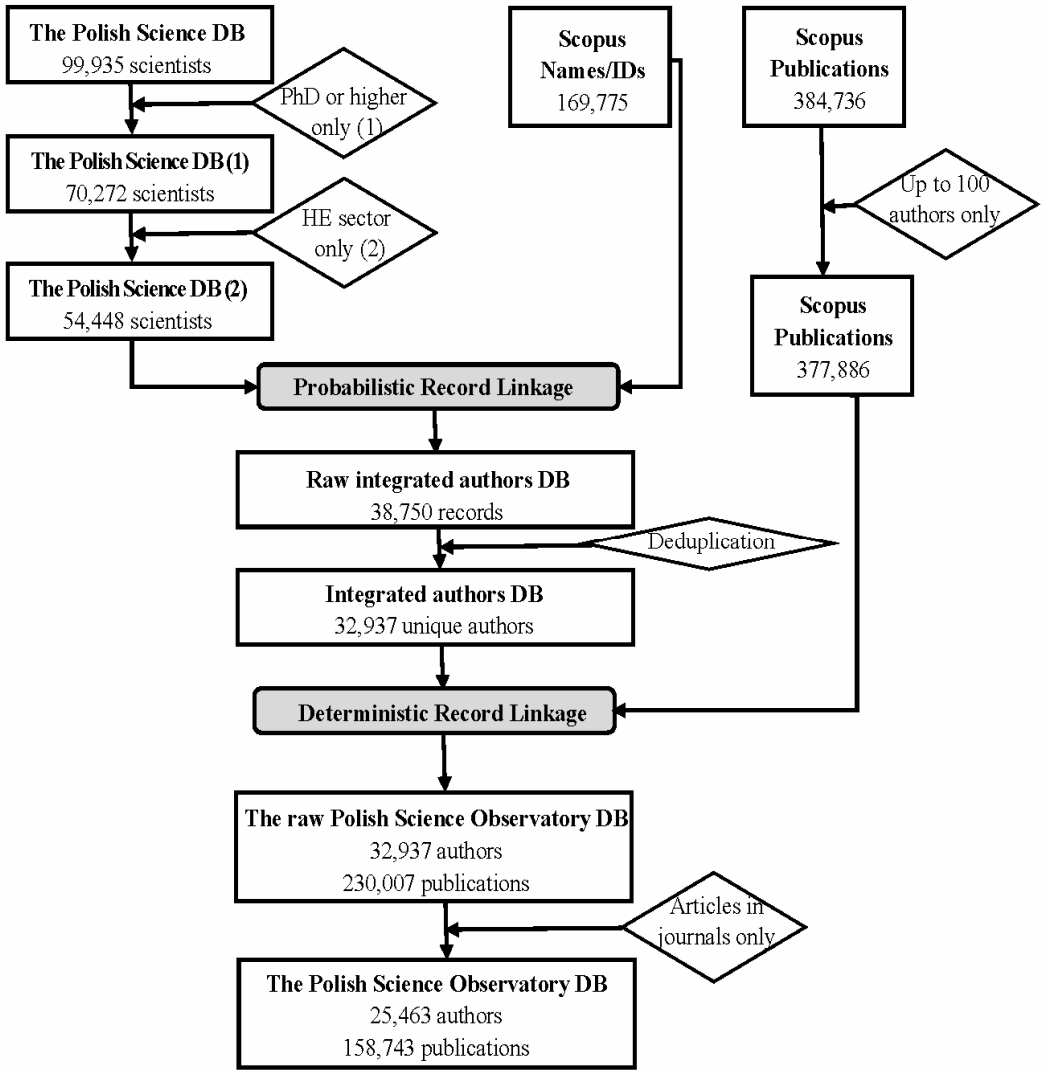


- N = 25,463 scientists

  (14,886 males, 10,577 females) (58.5% and 41.5%)
- **All the university professors holding at least a doctoral degree** and employed in 85 research-involved universities, grouped into 27 disciplines with all their Scopus-indexed publications (158,743 articles).
- **Female participation** in the academic profession **decreases with age.**
- **Female scientists clustered in lower academic positions**: half of all assistant professors, a quarter of full professors (48% and 24%).
- 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%).

Dataset

- Two large databases of different natures were merged (Wojciech Roszka, a co-author):
- Database I: an **official national administrative and biographical register** of all Polish academic scientists **(100,000scientists).**
- Database II: the **Scopus database (400,000 publications, 2009-2018)**.
- Merged to create "**The Observatory of Polish Science**."
- The key procedure: to appropriately **identify authors with their different individual IDs in the two databases** and to provide them with a new ID in the integrated database.
- **Probabilistic methods of data integration** used. The computation was made using the fastLink R package (version 0.6.0).
- The **main steps** in merging the databases - Figure 1:



How this research differs from previous studies?

(1) We examined **every internationally publishing Polish male and female scientist** and all (Scopus-indexed) publications from a decade (2009–2018).

(2) We had **100% gender determination** for all scientists in the system (rather than probability thresholds in gender determination).

(3) We defined what we termed the "**individual publication portfolio" for every Polish scientist** to examine their same-sex collaboration ratio.

(4) Our **unit of analysis** was the **gender-defined individual scientist** rather than the **individual publication**, with their specific distribution of male/female authorships.

(5) Most importantly, we used **a comprehensive, fully integrated biographical, administrative, publication, and citation database, "The Observatory".**

- This research goes **beyond traditional bibliometric studies** by **combining** the following:
  (1) data **routinely inaccessible** to large-scale studies: the **age** of all scientists and their **academic positions**, and
  (2) data **routinely accessible** in bibliometric studies, such as journal prestige, disciplines, and institutional type.

Findings (1) - Gender

Table 1. The median the same-sex collaboration ratio by gender.

	Same-sex collaboration
Male	0.500
Female	0.153
Total	0.333
Z	-44.291
p-value	<0.001

- While most **previous literature** highlights that **women are much more likely to have a female collaborator** (leading to **excessive gender homophily** in female publishing), our findings **do not support** this disparity in patterns for females.
- Using **an individual scientist** as **the unit of analysis**, we calculated the **proportion** of same-sex publications among collaborative articles **within the individual publication portfolio** of every Polish scientist.
- **Same-sex collaboration ratio:** the percentage of same-sex collaboration articles (male-male, female-female) among all collaborative articles in an individual publication portfolio.
- **For male scientists collaborating only with male scientists**, the ratio is 1. Analogously, the ratio of **0 is equivalent to conducting no same-sex collaboration.**

Findings (2) – Age and Gender Homophily

Findings (3) – Age Groups and Gender Homophily

Figure 2. The same-sex collaboration ratio: distribution by age groups and gender.

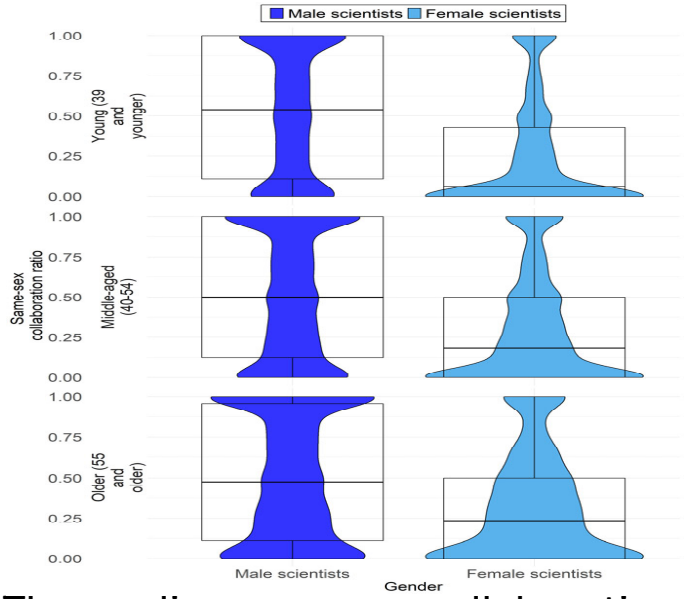


Table 2. The median same-sex collaboration ratio by age group and gender.

	Male	Female	Total	Z	p-value
Young (39 and younger)	0.5396	0.0625	0.2727	-29.676	<0.001
Middle-aged (40–54)	0.5000	0.1818	0.3333	-28.163	<0.001
Older (55 and older)	0.4762	0.2353	0.3750	-15.696	<0.001
Total	0.5000	0.1538	0.3333	-44.291	<0.001

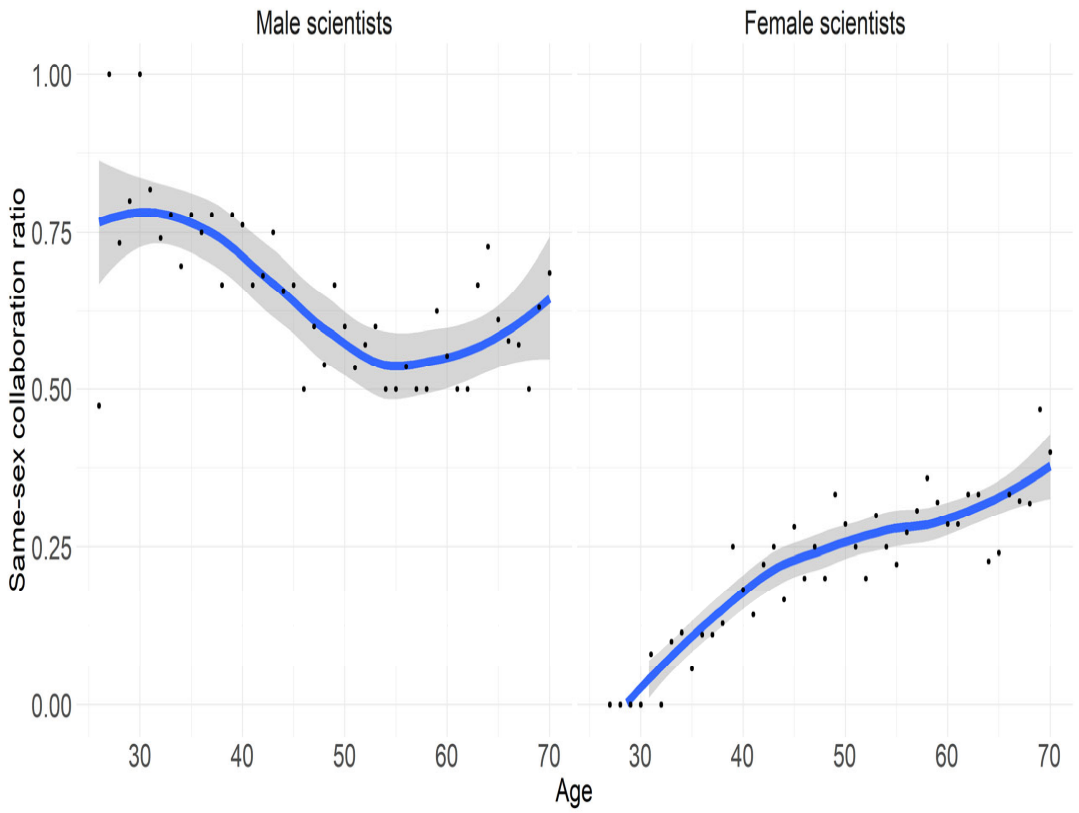
Across all age groups:

(1) **Male scientists tend to collaborate with male scientists**.

(2) **Female scientists tend *not* to collaborate with other female scientists**.

- Gender homophily in team formation seems to occur **with male scientists** - but **not with female scientists.**
- This finding is **not in line with previous research**: **female** scientists were expected to collaborate more often with **females** than males with males).

Findings (4) – Age and Gender Homophily



- **The year-by-year approach: two opposite trends** for both genders. (The **dots** represent the median value of the same-sex collaboration ratio **for each year of age**).
- **The difference in collaboration patterns for young scientists** by gender (an age group with **equal participation**) is especially interesting.
- **Young females** tend *not* to collaborate with females.
- **Older females** still tend to collaborate primarily with males.

Findings (5) – Academic Positions & Gender Homophily

Findings (6) – Academic Positions & Gender Homophily

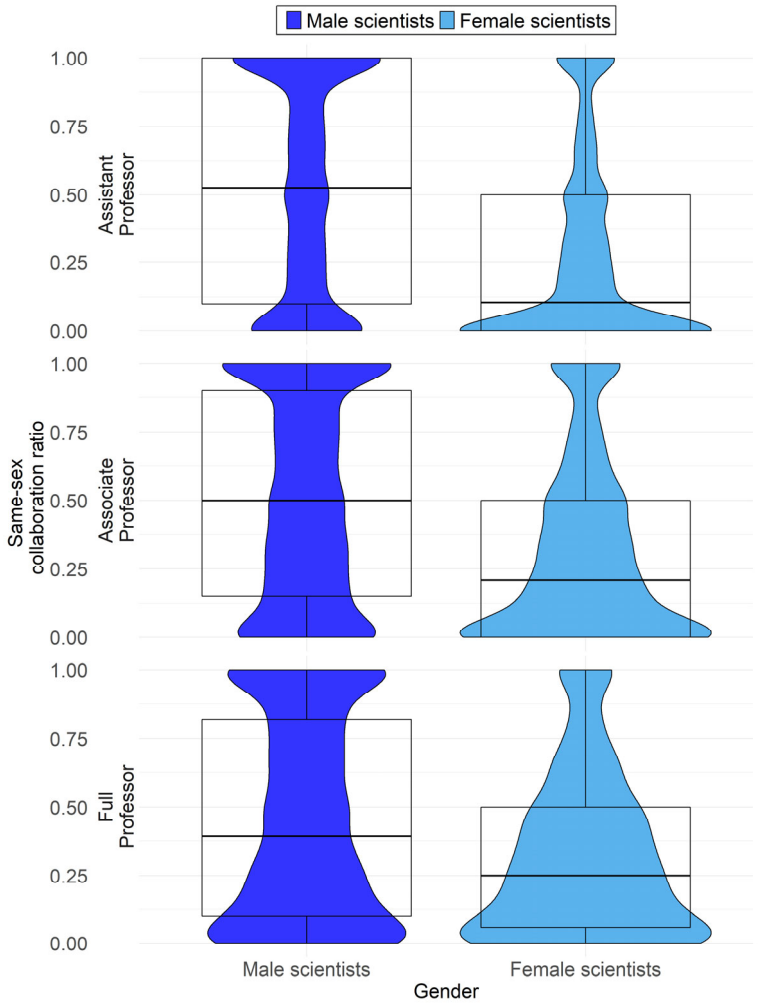


Figure 4. The same-sex collaboration ratio: distribution by academic position and gender (boxplots and violin plots combined).

- **All-female** collaboration (often discussed in literature) is **marginal**.
- **All-male** collaboration is **pervasive**.
- The gender patterns in publishing are **stable not only across age groups**—but also across **academic positions**.

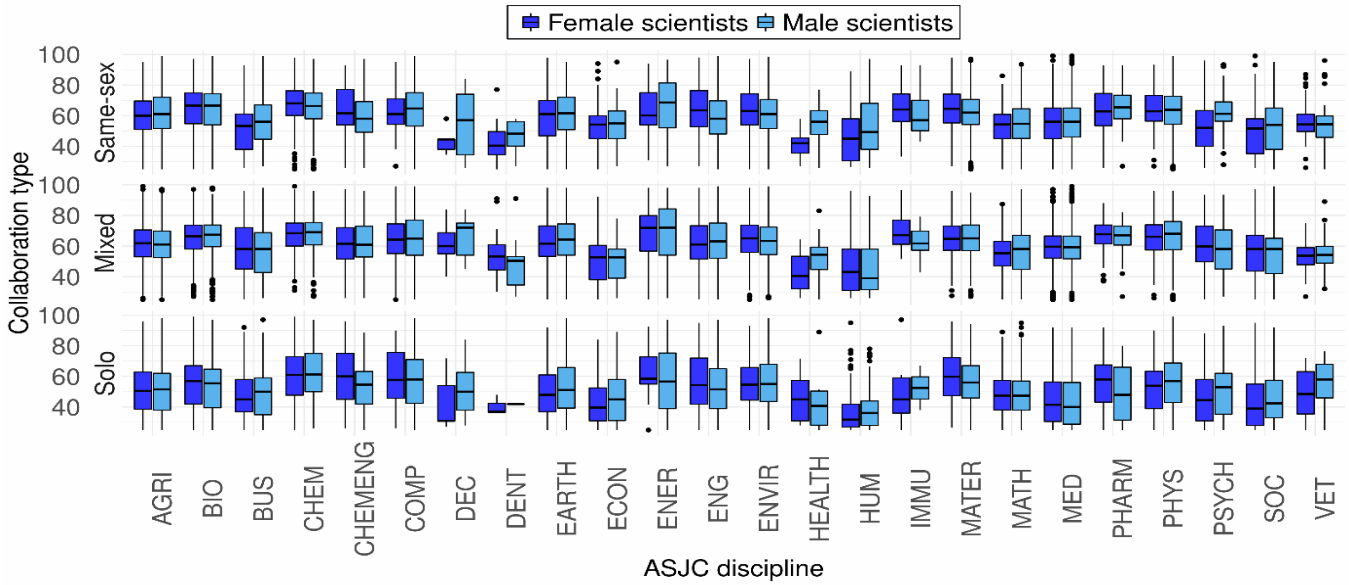
Findings (7) – Journal Prestige & Gender Homophily

- Both the **quantity and quality of output** in academia are relatively easily measured (standard limitations) - **articles** are published in **journals of different ranks (Scopus)**.
- The scientists have **their own unique individual publication portfolios**.
- Publications are translatable into **average individual prestige** via Scopus citation metrics.
- **Average prestige** which represents **the median prestige value for all publications** written by a scientist in 2009–2018 for three categories of publications (**same-sex, mixed-sex, and solo publications**).

Findings (8) – Journal Prestige & Gender Homophily

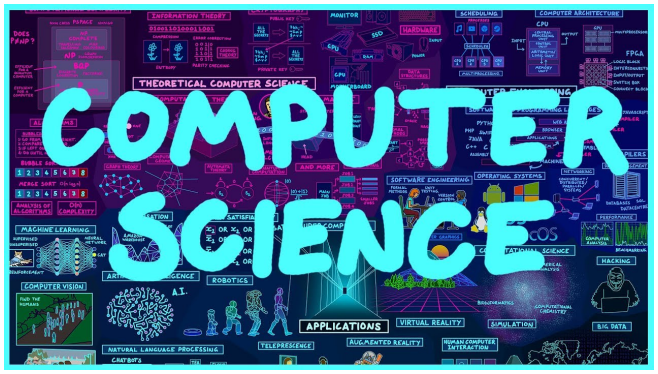
Table 3. The median prestige level distribution (by percentile from 0–99) of publications by major gender collaboration type and gender.

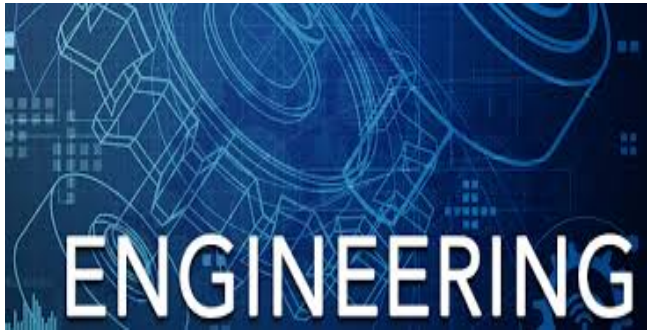
	Mixed-sex collaboration	Same-sex collaboration	Solo research (zero collaboration)
Male	62.50	59.17	50.00
Female	62.20	58.00	46.50
Total	62.42	58.27	48.50
Z	-1.497	-5.981	-5.121
p-value	0.134	<0.001	<0.001

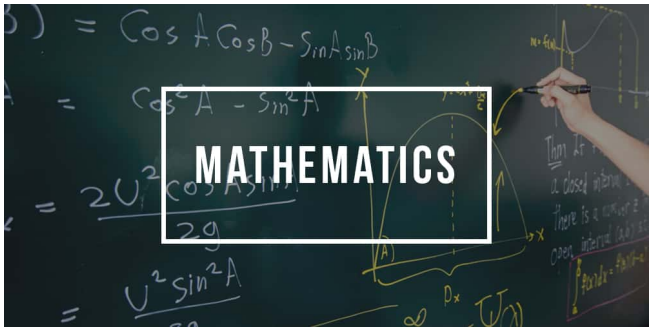


- Articles written in **mixed-sex collaboration** are published in more prestigious journals than those written in **same-sex collaboration.** (Consistent with previous literature).
- For each discipline, **solo research** is characterized by the lowest prestige level.

Findings (9) – Disciplines & Gender Homophily





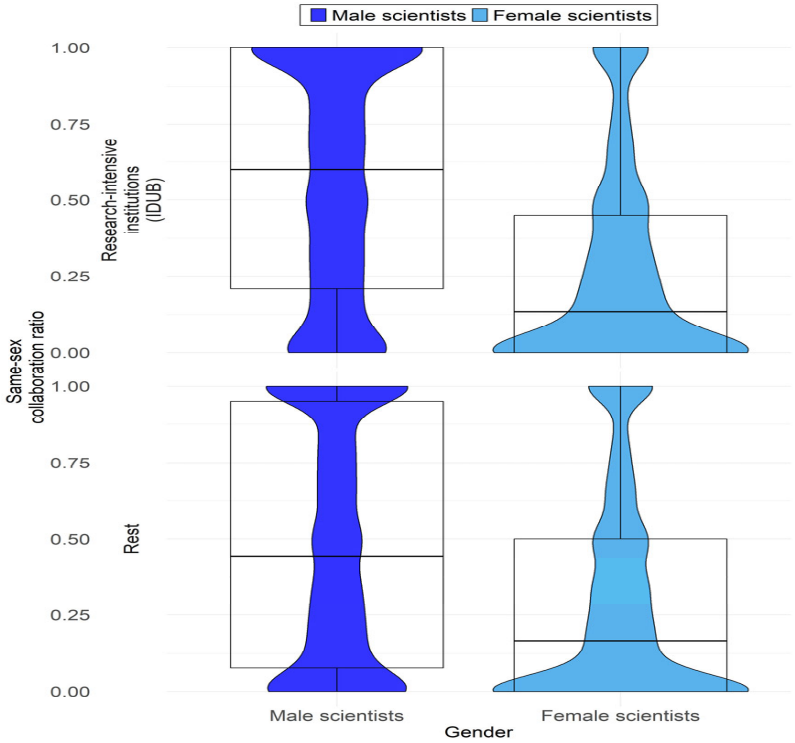


- In the case of the **male-dominated fields** of **computer science, engineering, and mathematics**, the same-sex collaboration ratio for males is very high.
- At least **half of male scientists** in these disciplines **collaborate *exclusively* with males**.
- At least **half of females do not collaborate with females *at all*.**

Findings (10) – Institutional Type & Gender Homophily

Table 4. The median of the same-sex collaboration ratio by institutional type and gender.

Institutional type	Male	Female	Total	Z	p-value
Research-intensive (IDUB)	0.6000	0.1348	0.4138	-30.717	<0.001
Rest	0.4444	0.1667	0.2857	-31.992	<0.001
Total	0.5000	0.1538	0.3333	-44.291	<0.001



- **For males, the proportion of all-male collaboration** is **higher** in **research-intensive** institutions (than the already high proportion for all institutions).
- **For females**, in contrast, the proportion of **all-female collaboration** is **lower** in **research-intensive institutions** (than the already low proportion for all institutions).
- **Males** are **more likely to collaborate with males** in research-intensive institutions (than in all institutions). Also females are also more likely to collaborate with males!
- Gender homophily in research-intensive institutions is thus **stronger for males** and **weaker for females** (than in the rest of the system).
- **A stronger institutional research focus** may generally induce collaboration **with male** scientists!

Findings (11) - A Modeling Approach: a Fractional Logit Regression Model

- Finally, we moved from descriptive statistics and **two-dimensional analysis** to **modeling**,
- We used a regression model for a fractional dependent **variable—a fractional logit regression model** (Papke & Woolridge, 1996).
- This model designed **for variables bounded between zero and one** (as with our dependent variable: the same-sex collaboration ratio).
- We estimated the **strength and direction of predictors** of conducting same-sex collaboration.
- **Being male** and **working in a male-dominated discipline** are the two most influential predictors, followed by working in a **research-intensive university**.

CONCLUSIONS

Conclusions

- Male-female collaboration practices in research were **tested against the homophily principle**: **similarity** in Poland **indeed breeds connection** between individual scientists.
- However, in the **Polish case**, this is **true only for male scientists!**
- While **forming collaborative research teams** Polish females tend *not* to publish with other females. They seem to massively **prefer male co-authors!**
- **Team formation** may be more **intuitive** (and result from the **dominant social norms** in academia rather than from solid individual **publishing strategies).**
- **Social norms** may **influence publishing patterns: predominantly same-sex publishing for young males** and **predominantly mixed-sex publishing** for young females.
- This, in time, may **contribute to the reduction of the gender productivity, citation, and promotion gaps** in Polish science. Global literature suggests that **these gaps may widen** if females excessively co-author and form professional networks mostly with females.

Future research avenues

(1) Adding a **comparative cross-national perspective**.

Moving to a global study, to our parallel "**Observatory of OECD Science**" dataset (27.4 million publications, 2009–2018, 1,674 research-active institutions, 11.1 authors). And

(2) Studying **global trends over time**.

Moving from a cross-sectional study ("individual publication portfolios" come from a single decade) to **a longitudinal study** (the portfolios come from the 1990s, 2000s, and 2010s).

Finally, lessons – do's and don'ts

- **Collaborate** in research (especially in mixed-sex teams).
- Consider **prestige generation** (if any) by each your publication.
- Consider **collaboration gains and collaboration costs** in various **collaboration types**: to whom the **publication credit** goes? leading the team or just joining the team? What journal?
- Mixed-sex teams open you the **access to informal networks** (otherwise unavailable).
- **Gender homophily** does **not hurt males' careers** – but **may hurt** females' careers.
- **All-male strategies** are good for males, **all-female strategies** may be detrimental for females.
- Publications are **not sex-neutral**: **all-female papers** are go to different journals than **all-male papers**.
- Do not contribute to the **long list of gender gaps** through your research – the gaps are **detrimental** to science overall!


THANK YOU
ANY QUESTIONS?