Young Male and Female Scientists: A Quantitative Exploratory Study of the Changing Demographics of the Global Scientific Workforce

Marek Kwiek
(1) Professor and Director, Center for Public Policy Studies (CPPS), AMU University of Poznan, Poznan, Poland
(2) Visiting Professor, German Center for Higher Education Research and Science Studies (DZHW), Berlin, Germany
kwiekm@amu.edu.pl, ORCID: orcid.org/0000-0001-7953-1063

Lukasz Szymula
(1) Faculty of Mathematics and Computer Science and (2) Center for Public Policy Studies (CPPS), AMU University of Poznan, Poland
lukasz.szymula@amu.edu.pl, ORCID: orcid.org/0000-0001-8714-096X

Abstract

In this study, the global scientific workforce is explored through a large-scale, generational, and longitudinal approach. We examine 4.3 million nonoccasional scientists from 38 OECD countries publishing in 1990–2021. Our longitudinal interest is in the changing distribution of young male and female scientists over time across 16 STEMM disciplines. We unpack the details of the changing scientific workforce using ten 5-year age groups within each discipline. The usefulness of global bibliometric data sources in analyzing the scientific workforce along the four dimensions of gender, age, discipline, and time is tested. Traditional aggregated data about scientists in general hide a nuanced picture of the changing gender dynamics within and across disciplines and age groups. For instance, the pivotal role of medicine in the global scientific workforce is highlighted, with almost half of all scientists (45.98%) in the OECD area being primarily involved in medical research, and more than half of female scientists (55.02%) being disciplinarily located in medicine. Limitations of bibliometric datasets are explored and global studies are compared with national-level studies. The methodological choices and their implications are shown, and new opportunities for how to study scientists globally are discussed.

Keywords:
Global scientific workforce; gender distribution; academic age distribution; OECD economies; bibliometric data; longitudinal research

1. Introduction

We explore the changing demographics of the global scientific workforce from the combined perspectives of age, gender, and academic discipline. Our approach is large scale, generational, and longitudinal: we examine 4.3 million nonoccasional scientists publishing over the past three decades (1990–2021). Our longitudinal interest is in the changing distribution of male and female scientists over time across different academic age groups—especially of young scientists with no more than 10 years of publishing experience—across 16 STEMM disciplines. The present research focuses on 38 OECD countries.

Large-scale and longitudinal approaches to study the differences in academic careers by gender, age, and discipline have been used only recently, accompanied by increasing access to digital national and global, commercial, and noncommercial workforce, administrative, and bibliometric databases, such as Web of Science, Scopus, and Microsoft Academic Graph (MAG), as well as Academic Analytics and Digital Bibliography and Library Project (DBLP) for the USA and CRISTIN for Norway or POL-on for Poland; Boekhout et al., 2021; Elsevier, 2020; King et al., 2017; Kwiek & Roszka, 2021a; Larivière et al., 2013; Nielsen & Andersen, 2021; Nygaard et al., 2022; Robinson-Garcia et al., 2020; Savage & Olejniczak, 2021; Way et al., 2017; Zhang et al., 2022). However, a generational approach to changing the global scientific workforce—especially age distribution by specific age groups—has not been applied. Currently, the participation of men and women in science can be studied longitudinally with a previously unattainable level of detail across countries, institutions, disciplines, and academic journals, as well as across age and seniority groups. Publications and their authors can be examined through temporal, topical, geographic, and network analyses or connected to time, themes, places, and other scientists (Börner, 2010, pp. 62–63). In our study, we have followed Huang et al. (2020), who reconstructed the complete publication history of over 1.5 million scientists to examine gender inequality in scientific careers globally (83 countries, 13 disciplines), Boekhout et al. (2021), who traced the publication careers of about 6 million male and female scientists in 1996–2018, and King et al. (2017), who examined 1.5 million research papers from the JSTOR bibliometric database to show gender differences in self-citation rates across disciplines and time.

Other examples of recent influential large-scale and sometimes longitudinal studies of global academic careers and their publishing, collaboration, or impact patterns are as follows: Robinson-Garcia et al. (2020) examined 71,000 publications from PLoS journals with 350,000 distinct authors to profile scientists across three task specializations and the changes in their career stages. Larivière et al. (2013) studied global gender disparities in science, using 5.5 million papers and 27.3 million authorships, showing that, globally, women account for fewer than 30% of fractionalized authorships and are similarly underrepresented regarding first authorships. Nielsen and Andersen (2021) studied the rise in global citation inequality, with a small stratum of elite scientists accruing increasing citation shares based on a dataset of 4 million authors and 26 million papers. Finally, Ioannidis et al. (2014), in their study of the "continuously publishing core" of the global scientific workforce, based on 15.2 million publishing scientists from 1996–2011, showed that less than 1% of scientists published their research each year in the studied 16-year period, accounting for as much as 41.7% of all papers.

Also, large-scale, national-level studies of academic careers in the USA have been increasingly precise in terms of gender, discipline, and age determination. For instance, Way et al. (2017) examined the traditional "rapid rise, gradual decline" narrative about productivity patterns, showing that this pattern holds for only 20% of individual faculty (while for the remaining 80%, there is a rich diversity of patterns). Using a DBLP dataset of 200,000 publications and career trajectories of

2,453 tenure-track faculty CV data, they showed how much diversity is hidden behind average academic career trajectories, creating inaccurate pictures of productivity patterns. Similarly, using the Academic Analytics commercial database, Savage and Olejniczak (2021) showed that the career publication activity of US scientists does not follow the traditional "peak-and-decline" pattern described in earlier studies.

Using a combination of data sources such as Academic Analytics, Web of Science, and the NSF Survey of Graduate Students and Postdoctorates in Science and Engineering, Zhang et al. (2022) showed that the disproportionate productivity of scientists in US elite institutions can be largely explained by their substantial labor advantage: their better access to externally funded graduate and postdoctoral labor. They used a matched pair design in which one midcareer researcher in the pair moved to a working environment with more available labor while the other moved to an environment with less available labor, with detailed productivity data for 78,000 faculty across 25 scientific disciplines. The association of institutional prestige with greater productivity was explained by greater available funded labor, which drove larger group sizes, thereby increasing group productivity (Zhang et al., 2022, p. 6). Studying Web of Science data for 1990 to 2010, Boothby et al. (2022, p. 9) showed that an average of 10% of US researchers leave academic science each year, and these researchers were in the very early career stages.

Huang et al. (2020) focused on gender differences in publishing career lengths and dropout rates in the USA; they used a career length matching design to study the relationship between career length and total productivity (412,770 female authors were matched with 412,778 male authors). A large proportion of observed gender gaps were rooted in gender-specific dropout rates and subsequent gender gaps in publishing career length and total productivity (Huang et al., 2020, p. 4615). Although gender, age, and discipline variables were used, the changing demographics of the global scientific workforce over time were not examined. Boekhout et al. (2021) showed an increasing trend in the percentage of women starting their careers as publishing researchers, from 33% in 2000 to 40% recently. Instead of considering entire publication careers (as in Huang et al. 2020), the authors compared the productivity of male and female scientists in specific years in their careers, showing that male scientists have a consistently higher publication productivity than female scientists, regardless of the year in which they started their career and period in their career, with differences in the range of 20–35% (full counting) and 25–40% (fractional counting) in favor of male productivity (Boekhout et al., 2021, p. 9; for an overview, see Halevi, 2019).

The present paper examines what we can know—based on available global data sources of the bibliometric type—about the changing demographics of the scientific workforce globally. We wanted to explore how useful the potential global data sources can be in analyzing the scientific workforce along the combined four dimensions of gender, age, discipline, and time. We tested how demographic transformations of the global science profession (including the global academic profession) can be measured using new data sources, hence transgressing the traditional approach in which national statistics from national statistical offices are aggregated, as in the OECD, UNESCO, and the European Union scientific workforce datasets.

We contribute to the discussion of the advantages and disadvantages of using global publication and citation databases—or "structured" Big Data (Holmes, 2017; Salganik, 2018; Selwyn, 2019)—in global academic profession studies in which the data on gender, age, and disciplines have traditionally been available almost exclusively cross-sectionally, mostly on a small national scale and increasingly on a small international comparative scale through survey research. We unpack details of the changing scientific workforce using ten 5-year age groups within each discipline from a longitudinal perspective.

We have changed the unit of analysis: the individual scientist (with their unique identity) rather than individual publication (with its unique identity) is the focus. Although a bibliometric data source is used (Scopus raw data provided to us by Elsevier's ICSR Lab through a multiyear collaboration agreement), our focus is on scientists and their attributes rather than publications and their properties.

1.1. Research Questions

Our four research questions regarding publishing and nonoccasional scientists are as follows: (1) What is the global gender distribution of scientists across disciplines, and especially, how are male and female scientists disciplinarily located? (2) What is the global age distribution of scientists across disciplines and gender, and especially, how are young and old male and female scientists disciplinarily located? (3) How do the global gender and age distributions of scientists across disciplines change over time, especially for young versus old male and female scientists? (4) How is the participation in science of female scientists changing over time and across disciplines, and what are the disciplinary participation trends?

2. Data and Methods

2.1. Data

The major characteristics of the longitudinal study population for 1990–2021 (4,314,666 scientists, including 1,645,860, or 38.15% female) are presented in Table 1. The major characteristics of the cross-sectional study subpopulation for 2021 (1,502,792 scientists, including 579,399, or 38.55% female) are presented in Table 2. Our population was constructed as follows (we refer to the population rather than the sample because we have all scientists, with their attributes, as units of analysis): First, to determine the number of scientists, unique authors of publications (type: journal article, conference paper in a book or a journal) who published their works in 1990–2021 were selected. For this selected group of authors, the years of their research activities were determined. The resulting set of scientists was then narrowed down according to a package of five restrictions: (1) an OECD country, (2) a STEMM discipline, (3) gender (binary approach: man or woman), (4) a nonoccasional status in science: a minimum scientific output defined as three publications throughout the scientist's career (lifetime), and (5) academic age, or the time passed since the first publication, in the 1–50 years range.

The minimum output in lifetime publication history allowed us to limit our population to nonoccasional scientists, that is, scientists functioning in the scientific community more than accidentally. Additionally, scientists with one or two publications in the Scopus database are more likely to result from mistakes made by author name disambiguation algorithms (see Boekhout et al., 2021, p. 3). Generally, in terms of author name disambiguation, Scopus is more accurate than Web of Science (Sugimoto & Larivière, 2018, p. 36). Then, for each scientist, academic experience in full years, beginning in the year of the first publication of any type, was determined. For each year of a scientist's research activities, the length of their academic experience and membership in the corresponding academic age group were determined. We used a population for 1990–2021 for longitudinal analyses, a subpopulation for 2021 for a cross-sectional analysis, and the two subpopulations for 2000 and 2021 for analyses comparing two points in time. Figure 1 summarizes the population's design.

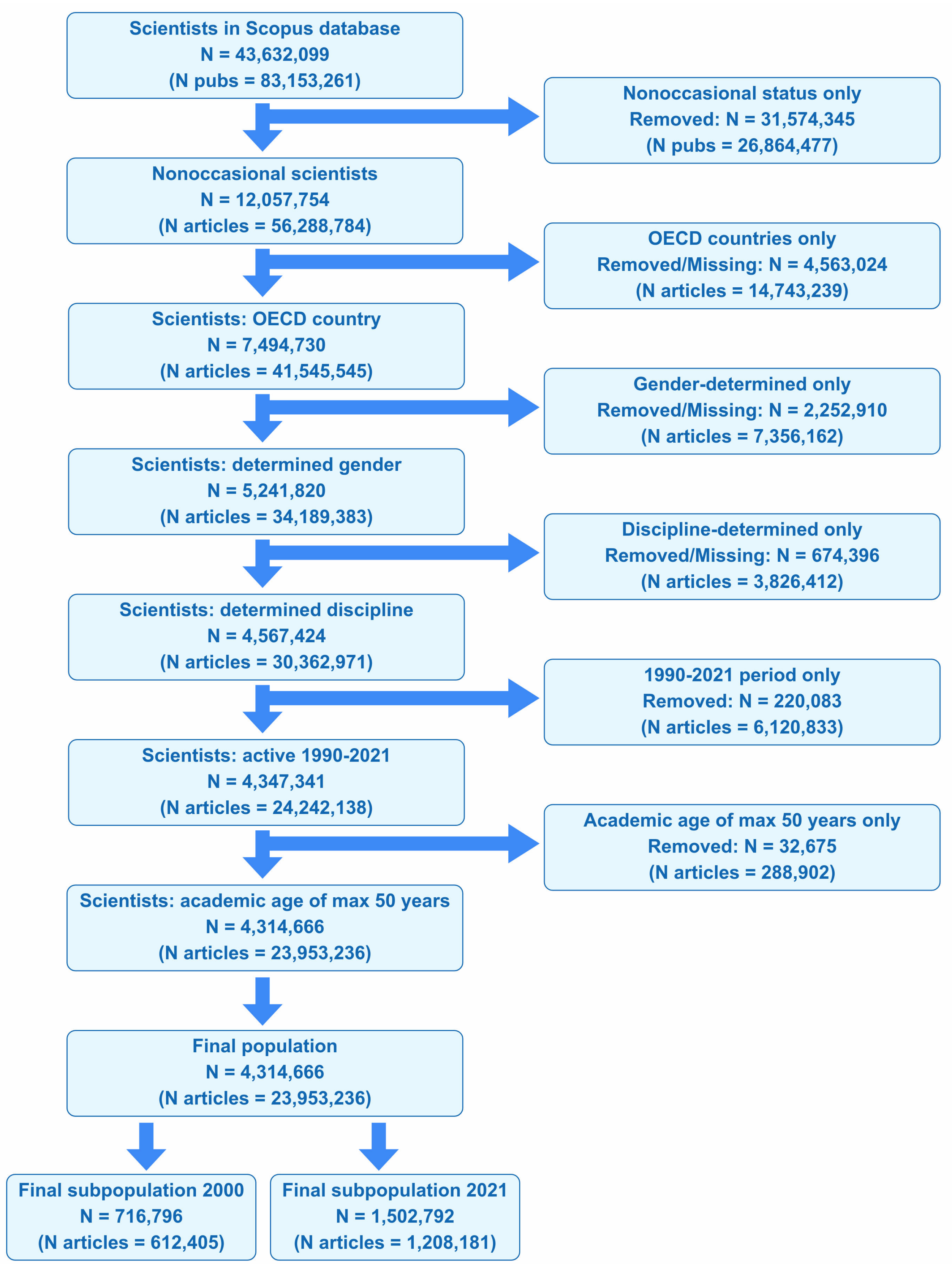

Figure 1. Flowchart: stages in constructing the population and the two subpopulations.

Table 1. The population for 1990–2021: major characteristics.

		Female scientists			Male scientists			Total		
		n	row %	col %	n	row %	col %	n	row %	col %
Discipline	Total	1,645,860	38.15	100	2,668,806	61.85	100	4,314,666	100	100
	AGRI	104,805	39.98	6.37	157,318	60.02	5.89	262,123	100	6.08
	BIO	328,806	46.26	19.98	381,963	53.74	14.31	710,769	100	16.47
	CHEM	87,608	30.16	5.32	202,843	69.84	7.60	290,451	100	6.73
	CHEMENG	4,294	23.06	0.26	14,330	76.94	0.54	18,624	100	0.43
	COMP	16,191	16.59	0.98	81,414	83.41	3.05	97,605	100	2.26
	EARTH	34,042	27.62	2.07	89,221	72.38	3.34	123,263	100	2.86
	ENER	3,255	19.09	0.20	13,793	80.91	0.52	17,048	100	0.40
	ENG	24,992	11.52	1.52	191,978	88.48	7.19	216,970	100	5.03
	ENVIR	35,867	38.35	2.18	57,661	61.65	2.16	93,528	100	2.17
	IMMU	26,805	53.24	1.63	23,547	46.76	0.88	50,352	100	1.17
	MATER	26,227	26.16	1.59	74,043	73.84	2.77	100,270	100	2.32
	MATH	11,915	20.15	0.72	47,206	79.85	1.77	59,121	100	1.37
	MED	836,890	45.44	50.85	1,005,040	54.56	37.66	1,841,930	100	42.69
	NEURO	40,961	47.20	2.49	45,819	52.80	1.72	86,780	100	2.01
	PHARM	15,641	41.35	0.95	22,183	58.65	0.83	37,824	100	0.88
	PHYS	47,561	15.44	2.89	260,447	84.56	9.76	308,008	100	7.14
OECD country (TOP 10)	USA	540,501	39.73	32.84	819,882	60.27	30.72	1,360,383	100	31.53
	Japan	92,601	19.28	5.63	387,599	80.72	14.52	480,200	100	11.13
	Germany	118,509	33.49	7.20	235,312	66.51	8.82	353,821	100	8.20
	UK	116,285	39.49	7.07	178,187	60.51	6.68	294,472	100	6.82
	Italy	119,688	50.36	7.27	117,960	49.64	4.42	237,648	100	5.51
	France	93,770	42.07	5.70	129,110	57.93	4.84	222,880	100	5.17
	Canada	68,983	42.75	4.19	92,393	57.25	3.46	161,376	100	3.74
	Spain	71,656	48.13	4.35	77,233	51.87	2.89	148,889	100	3.45
	Australia	50,652	44.79	3.08	62,425	55.21	2.34	113,077	100	2.62
	South Korea	19,886	19.32	1.21	83,038	80.68	3.11	102,924	100	2.39

2.2. Methods

In this section, we present the five basic procedures to unambiguously define the attributes of the scientists in our population. We initially used raw data for 2020 and before, based on the Scopus database version dated 18 August 2021. The raw data were made available to us by Elsevier under an agreement with the ICSR Lab. Finally, the Scopus database version for 2021 and before, dated 21 October 2022 was used.

To obtain the results at the aggregate level, the operation in the ICSR Lab relied on the use of the Databricks environment, which allowed for managing and executing cloud computing with Amazon EC2 services. The scripts to generate the results were written using the PySparkSQL library. The work on obtaining the results proceeded in two steps. The first step was to work on 1% of the Scopus database data with the snapshot date 18 August 2021 (from ICSR Lab: 1% of the data volume based on a set of 20,000 publications between 2010 and 2018 and including all publications cited by and citing these publications) using a cluster in standard mode with Databricks Runtime version 11.2, including Apache Spark technology in version 3.3.0, Scala 2.12 and i3.2xlarge instance with 61 GB Memory, 8 Cores, one to four workers for worker type and i3.xlarge instance with 30.5 GB Memory, 4 Cores for driver type. Test runs of the scripts covered 1% of the data, with the goal of optimizing the time and cost of the performed calculations.

After reviewing the correctness of the scripts, the final run was performed. The operation was carried out on a 100% Scopus database with a snapshot date 21 October 2022 using cluster in standard mode with Databricks Runtime version 11.2 ML with Apache Spark technology version 3.3.0, Scala 2.12, and instance i3.2xlarge with 61 GB Memory, 8 Cores, one to six workers for worker type and instance c4.2xlarge with 15 GB Memory, 4 Cores for Driver type. The execution time for the entire script took 1.13 hours; this operation was launched on November 22, 2022.

Table 2. The subpopulation for 2021: major characteristics

		Female scientists			Male scientists			Total		
		n	row %	col %	n	row %	col %	n	row %	col %
Academic age group	Total	579,399	38.55	100	923,393	61.45	100	1,502,792	100	100
	5 and less	148,749	46.26	25.67	172,795	53.74	18.71	321,544	100	21.40
	6–10	149,875	43.47	25.87	194,936	56.53	21.11	344,811	100	22.94
	11–15	102,419	40.52	17.68	150,366	59.48	16.28	252,785	100	16.82
	16–20	71,335	36.73	12.31	122,878	63.27	13.31	194,213	100	12.92
	21–25	45,297	32.74	7.82	93,052	67.26	10.08	138,349	100	9.21
	26–30	30,302	28.86	5.23	74,698	71.14	8.09	105,000	100	6.99
	31–35	17,736	24.83	3.06	53,682	75.17	5.81	71,418	100	4.75
	36–40	8,432	20.58	1.46	32,541	79.42	3.52	40,973	100	2.73
	41–45	3,833	17.27	0.66	18,357	82.73	1.99	22,190	100	1.48
	46–50	1,421	12.35	0.25	10,088	87.65	1.09	11,509	100	0.77
Discipline	AGRI	42,657	40.13	7.36	63,645	59.87	6.89	106,302	100	7.07
	BIO	92,185	43.27	15.91	120,854	56.73	13.09	213,039	100	14.18
	CHEM	22,450	30.21	3.87	51,862	69.79	5.62	74,312	100	4.94
	CHEMENG	1,287	24.98	0.22	3,865	75.02	0.42	5,152	100	0.34
	COMP	6,449	18.20	1.11	28,986	81.80	3.14	35,435	100	2.36
	EARTH	14,446	27.87	2.49	37,390	72.13	4.05	51,836	100	3.45
	ENER	1,527	20.28	0.26	6,004	79.72	0.65	7,531	100	0.50
	ENG	9,029	13.82	1.56	56,326	86.18	6.10	65,355	100	4.35
	ENVIR	14,688	40.15	2.54	21,892	59.85	2.37	36,580	100	2.43
	IMMU	6,949	50.03	1.20	6,940	49.97	0.75	13,889	100	0.92
	MATER	10,257	27.09	1.77	27,601	72.91	2.99	37,858	100	2.52
	MATH	4,653	20.02	0.80	18,590	79.98	2.01	23,243	100	1.55
	MED	318,792	46.14	55.02	372,166	53.86	40.30	690,958	100	45.98
	NEURO	13,873	43.76	2.39	17,833	56.24	1.93	31,706	100	2.11
	PHARM	3,190	45.98	0.55	3,748	54.02	0.41	6,938	100	0.46
	PHYS	16,967	16.53	2.93	85,691	83.47	9.28	102,658	100	6.83
OECD country (TOP 10)	USA	176,646	40.63	30.49	258,155	59.37	27.96	434,801	100	28.93
	Japan	22,331	18.15	3.85	100,695	81.85	10.90	123,026	100	8.19
	Germany	36,659	32.19	6.33	77,212	67.81	8.36	113,871	100	7.58
	Italy	51,171	49.21	8.83	52,821	50.79	5.72	103,992	100	6.92
	UK	40,328	38.88	6.96	63,392	61.12	6.87	103,720	100	6.90
	France	31,657	39.74	5.46	47,996	60.26	5.20	79,653	100	5.30
	Spain	29,067	46.89	5.02	32,925	53.11	3.57	61,992	100	4.13
	Canada	24,022	42.36	4.15	32,685	57.64	3.54	56,707	100	3.77
	Australia	21,160	44.49	3.65	26,396	55.51	2.86	47,556	100	3.16
	South Korea	7,903	19.31	1.36	33,034	80.69	3.58	40,937	100	2.72

2.2.1. Gender determination

To obtain the gender of the scientists in the population, the gender data established by the ICSR Lab platform was first used (N_{author}=34,596,581). Then, only scientists who had a defined gender (man/woman) with a gender probability score greater than or equal to 0.85 were included (N_{author}=21,508,029). To assign gender to an author, the ICSR Lab used Elsevier's solution, which used the Namsor tool. Determination of gender was based on three characteristics: author's first name, author's last name, and author's first country. The author's first country was determined based on the author's dominant country in their first publication year, based on output in the Scopus database. For authors who had more than one dominant country, the observation was not assigned a value. The Namsor tool returned gender and gender probability score (Elsevier, 2020, pp. 122–123).

2.2.2. Discipline determination

To obtain the dominant discipline of scientists in the population, a set of publications from the Scopus database was used (N_{pub}=85,585,123; N_{author}=43,632,099). Publications were from 2021 and before and were restricted by source and type of publication: (1) journal article and (2) conference paper in a book or journal (N_{pub}=60,987,987; N_{author}=36,379,221). From the table of publications, the columns with publications' identifiers, authors' identifiers, and cited references were selected. Each cited reference ($N_{citedreference}$=1,434,621,669) was accompanied by its discipline as assigned by the discipline of the journal in which it appeared. The disciplines assigned to a cited reference were based on the four-digit ASJC code used by the Scopus database. To switch to a two-digit classification, unique disciplines were selected, based on the first two digits of the four-digit value. Then, for each author, the number of cited references was counted for all disciplines referenced by the author, excluding the "multidisciplinary" discipline. For each author, the discipline with the highest number of cited references (modal value) was selected. A table containing the author's identifier and their dominant discipline was obtained. For the described summary, there could have been cases in which an author had several dominant disciplines or no disciplines (included N_{author}=26,706,031). Here, authors who had more than one dominant discipline or no discipline were removed from the table (removed N_{author}=9,673,190). Authors were removed, among other reasons, because the cited references from their articles may have referred to journals outside the Scopus database, or there was an equal number of cited references to different disciplines. Subsequently, the table was restricted to only authors with an assigned discipline from the STEMM group, and the final number was (N_{author}=24,425,447).

2.2.3. Determining the country of affiliation

Publications were from 2021 or earlier and were restricted by source and type: (1) journal article and (2) conference paper in a book or journal (N_{pubs}=60,987,987; N_{author}=36,379,221). From the table of publications, columns with publications' identifiers, authors' identifiers, and countries for each author of the publication were selected. Then, for each author, the number of countries that the scientist indicated in all their publications was counted. For each author, the country with the highest number of references (modal value) was selected. For the described summary, there may have been cases in which an author had several countries (included N_{author}=31,332,750). For this purpose, authors who had more than one country or no countries were removed from the table (removed N_{author}=5,046,471). The table was then filtered to include scientists from 38 OECD countries. The final number was (N author=19,296,388).

2.2.4. Determining scientists' nonoccasional status

Under the proposed definition, a nonoccasional scientist has at least three research articles (as defined above) in their output. The publications were from 2021 or before and were limited by the same source and type of publication as above (N_{pubs}=60,987,987; N_{author}=36,379,221). Columns containing publications' identifiers and authors' identifiers were selected from the table of publications. For each author, the number of publications was counted. The table was then filtered to include scientists who had a minimum of three publications (N_{author}=12,057,755).

2.2.5. Determining academic age

Finally, to obtain the academic age of the scientists in the population, the same set of publications from the Scopus database was used, and the publications were from 2021 or before. Author identifiers and year of publication were selected from the table. For each author, the year of the first and last publication (of any type) was determined. Then, the number of years of authors' research activities (distance from the first to last publication in years) was calculated according to the formula: year of the last publication – year of the first publication + 1. Authors who had more than 50 years of research activities were removed from the table (included N_{author}=43,568,252; removed N_{author} = 63,847). Then, for the authors included in the study (N_{author}=4,314,666; i.e., the final population) that contained the years of academic activity defined for publications, the academic age in a given publication year was determined according to the following formula: publication year – year of first publication + 1.

Based on the value of academic age, an author was assigned to an age group according to 10 ranges: 5 and less, 6–10, 11–15, 16–20, 21–25, 26–30, 31–35, 36–40, 41–45, and 46–50.

2.2.6. List of STEMM disciplines

We focused on all 16 STEMM disciplines, as defined by the journal classification system used in the Scopus database (All Science Journal Classification, ASJC): AGRI, agricultural and biological sciences; BIO, biochemistry, genetics, and molecular biology; CHEMENG, chemical engineering; CHEM, chemistry; COMP, computer science; EARTH, earth and planetary sciences; ENER, energy; ENG, engineering; ENVIR, environmental science; IMMU, immunology and microbiology; MATER, materials science; MATH, mathematics; MED medicine, NEURO, neuroscience; PHARM, pharmacology, toxicology, and pharmaceutics; and PHYS, physics and astronomy.

3. Results

3.1. General Results

Although the analysis of the changing numbers of male and female scientists over time may be distorted by the inability to distinguish between an expansion in numbers of scientists and in numbers of journals indexed in large bibliometric datasets, in contrast, the changing relative presence of female scientists is traceable. Although the increasing number of publishing scientists over time correlated with the increasing coverage in Scopus, the percentages of publishing male and female scientists were independent of the journal coverage. Consequently, while the number of publishing scientists changing over time was not a reliable measure of the changing women's participation in global science, the percentages of male scientists and female scientists adequately reflected the changes in the global academic workforce. Consequently, we refer to numbers of male and female scientists only in 2021; in all other cases (trends 1990–2021; comparison of 2000 and 2021), we refer to their percentages.

At the age-aggregated level, almost half (45.98%) of the whole global scientific workforce was engaged in medical research in 2021. There were 690,958 (nonoccasional, publishing) scientists involved in medicine MED, and the second largest discipline was biochemistry, genetics, and molecular biology BIO, with more than three times fewer publishing scientists (213,039), followed by agricultural and biological sciences AGRI (106,302) and physics and astronomy PHYS (102,658), which were seven times smaller. Figure 2 (left top) provides a snapshot view of where the current research has been located and how publishing scientists have been distributed among disciplines and gender.

In total, across all age groups, in 2021, there were 1.5 million scientists, 923,000 men, and 579,000 women (38.55%). Figure 2 (top right) shows where (nonoccasional, publishing) female scientists were globally concentrated in terms of countries: the USA (with 176,600 women), followed by Italy (51,200), the UK (40,300), Germany (36,700), France (32,000), and Spain (29,000). Two-thirds of female scientists (63.09%) publishing in 2021 were located in these six countries.

The concentration of female scientists was even steeper across disciplines: more than half of female scientists (55.02%) were located in medicine (MED) and 1 in 7 (15.91%) in biochemistry, genetics, and molecular biology (BIO). About 70% (70.93%) of female scientists were concentrated in these two disciplines (see Tables 12 and 13 in the Electronic Data Appendices [EDA] for details). At the same time, the highest share of female scientists was in immunology and microbiology (IMMU), where half of all scientists were female (50.03%), followed by MED, PHARM, NEURO, BIO, ENVIR, and AGRI, all higher than 40% (Figure 2, bottom left). The lowest share of female scientists—less than 20% or around it—was observed in ENG, followed by PHYS, COMP, MATH, and ENER.

3.2. The Global Distribution of Male and Female Scientists: A Cross-Sectional View (2021)

3.2.1. Scientists by age groups, gender, and disciplines: Two complementary approaches

To study the gender distribution of the scientific workforce by age group, we used two complementary approaches we termed "horizontal" and "vertical."

(1) **A horizontal approach**: Analyzing the gender distribution of scientists horizontally within the same age groups. For each discipline, for each of the ten 5-year age groups, the percentages of male and female scientists totaled 100%.

(2) **A vertical approach**: Analyzing the gender distribution of scientists vertically—separately male and separately female scientists—across all age groups. For each discipline, there was 100% of male and 100% of female scientists, differently distributed across the 10 age groups.

3.2.1.1. A horizontal approach

Disciplines at a single point in time (2021) were populated across disciplines by scientists of different age groups and genders. Figure 3 shows the percentage of female scientists across disciplines by age group. We generally observed results of a huge inflow of female scientists (who are present in 2021) to most disciplines in the past years and decades: for younger generations working in 2021, the percentages of female scientists were substantially higher than for older generations.

Two distinctive clusters of disciplines clearly emerged regarding the gender composition of very young scientists (age group: 5 years and less) in 2021:

(1) "**Young female dominated**" disciplines: with at least 50% of very young female scientists. The share of female scientists in the youngest age group was more than 50% (IMMU, PHARM, NEURO, MED, AGRI, BIO). These disciplines showed a high and increasing share of female scientists for younger age groups. The discipline most open to female scientists in the past years and decades was IMMU (59.04%; Figure 3). More than 8 in 10 female scientists worked within these six disciplines (480,346, or 82.90%).

(2) "**Young male dominated**" disciplines: with less than 50% of very young female scientists. These disciplines showed both a high (but not exceeding 50%) share of very young female scientists and a low and stable share of very young female scientists (e.g., COMP 20.65%, ENG 17.74%, MATH 22.79%, and PHYS 21.43%, Figure 3). The discipline most closed to female scientists in the past years and decades was ENG. Almost 2 in 10 female scientists worked in these 10 disciplines (99,053, or 17.1%; see the details in Table 14 in EDA).

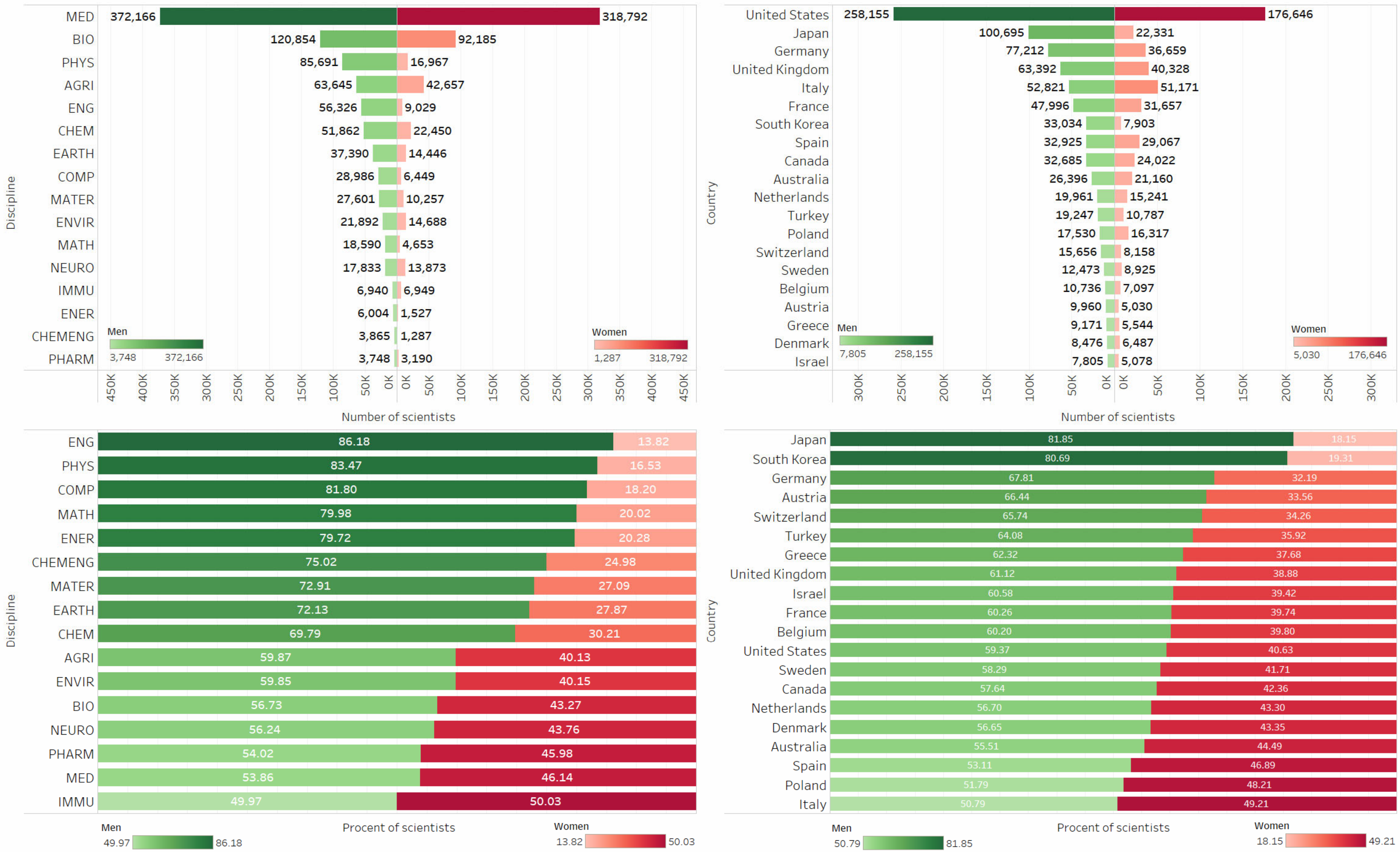

Figure 2. The number of publishing nonoccasional STEMM scientists by discipline and gender (left top) and by country (20 biggest systems only) and gender (right top). The share by discipline and gender (left bottom) and by country (20 biggest systems only) and gender (right bottom) (in %), 2021 (N = 1,502,792)

Generally expecting ever more female scientists across all STEMM disciplines moving up the age groups, we assessed ongoing changes based on a snapshot (2021), especially examining the youngest age groups. MED and BIO showed a structure in which, for every successive lower age group in 2021, a higher share of female scientists was observed. PHYS, COMP, and MATH, traditionally male-dominated disciplines, in contrast, showed a stable structure in which, for every successive lower age group in 2021, a similar (or only slightly higher) share of female scientists was observed. These two contrasting demographic patterns showed different inflows of young female scientists to disciplines: huge and increasing versus small and stable.

The current global disciplinary distribution of young women in science is consequential for gender parity in science in the future, despite high attrition among young scientists generally and young female scientists in particular (1 in 10; see Boothby et al., 2022). The current youngest cohort will be middle-aged cohorts within a decade, and current oldest cohorts will disappear from the publishing enterprise, with new challenges for disciplines continuously heavily male dominated.

3.2.1.2. A vertical approach

In contrast, using a vertical approach to changing gender composition within disciplines, we examined male and female scientists separately: within each discipline, the distribution of all male and all female scientists was studied by age groups (Figure 4).

In nine disciplines, most female scientists were located in the two young age groups or their academic experience was no more than 10 years (Figure 4). Young female scientists dominated (> 50%) among all female scientists disciplines like CHEM or MED, in which as much as 68.19% of all female scientists were located. Thus, the inflow of (publishing nonoccasional) female scientists in the past decade or so in these disciplines has been massive. The lowest share of young female scientists among all female scientists—or the weakest inflow (< 40%)—was for COMP and MATH. In all disciplines combined (Total), the share of young female scientists among all female scientists reached 51.54%, and the share of young male scientists among all male scientists was considerably lower and reached 39.82%. The emergent picture supports narratives of ever more young women in science: of all the women currently present, more than half had no more than 10 years of publishing experience.

Medicine (MED) was the largest discipline (691,000 or 45.98% of all scientists in 2021), in which more than half of all female scientists (55.02%) were currently disciplinarily located. Horizontally (Figure 3), more than half of very young scientists in MED were female (53.42%), and more than half of young scientists were female (51.36%). With every older age group, the share of female scientists in this group decreased, going below 40% for the age group 21–25 and about 20% for the age group of 41–45. Vertically (Figure 4), more than half of all female scientists in MED (53.33%) were in the age group of 10 or less years (of academic experience). In the three youngest age groups, the share of female scientists among all female scientists was higher than that of male scientists among all male scientists; for all other age groups, the case was the opposite.

Figure 3. Ever-increasing participation of women in younger generations of scientists, with a few exceptions. Horizontal approach: distribution of publishing nonoccasional STEMM scientists by discipline, age group, and gender (row percentages: 100% horizontally), 2021 (N = 1,502,792)

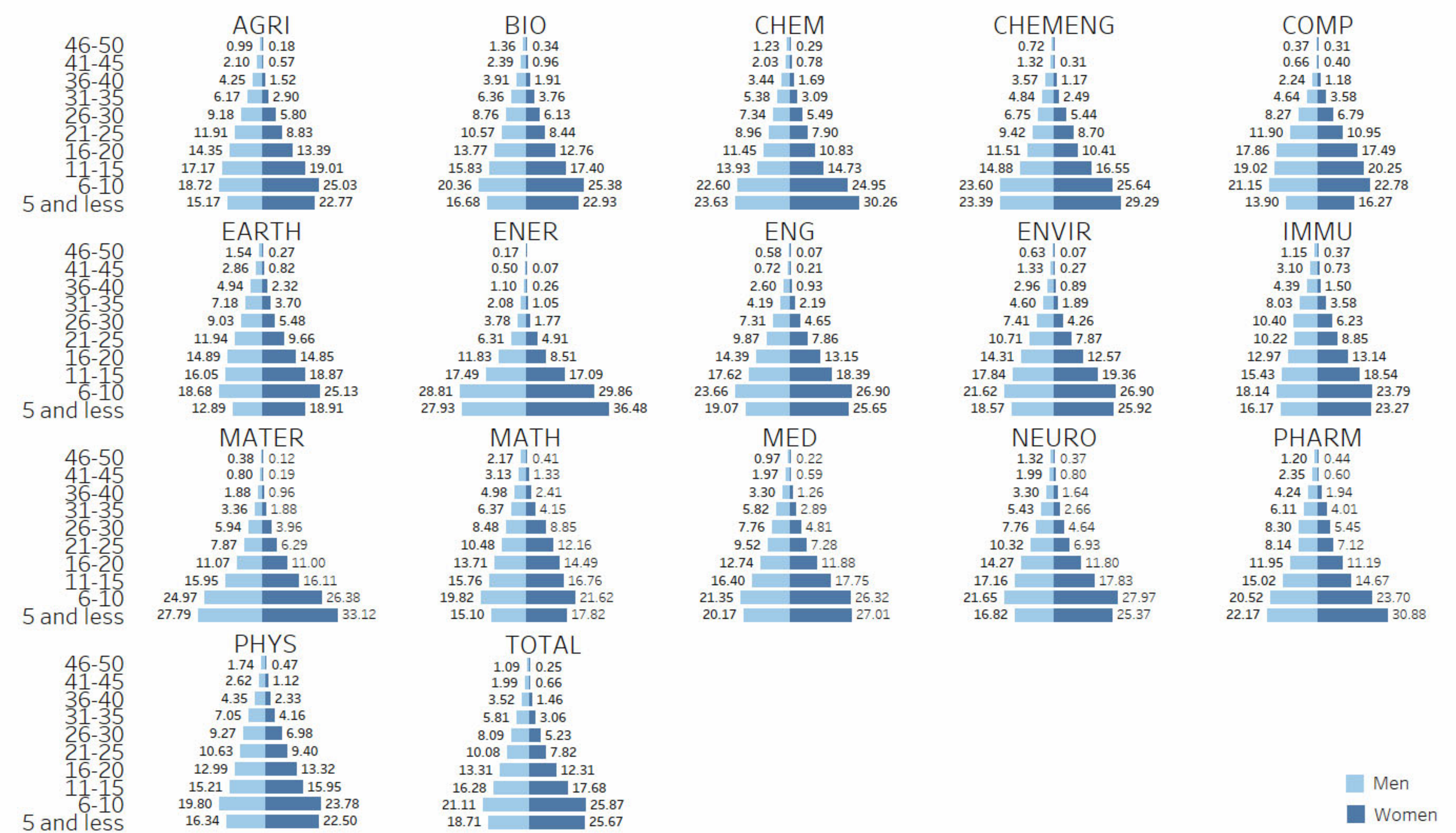

Figure 4. In most disciplines, the majority of women belong to the two youngest age groups. Vertical approach: distribution of publishing nonoccasional STEMM scientists by discipline, age group, and gender (column percentages: 100% vertically, for all age groups combined), 2021 (N = 1,502,792)

3.3. Female Scientists by Disciplines: Trends 1990–2021

We analyzed the changing participation of women in science over time to test the claim that the inflow of female scientists into science over the past three decades was powerfully differentiated by discipline.

The number of individual scientists used here to examine the trend over time was 4.3 million (61.85% male and 38.15% female, Table 1). We studied the trend of the percentage of female scientists present in global science in 1990–2021. Our analysis used a linear trend in the form of y = at + b. In the equation, *b* is where the line intersected the "y axis" and *a* denotes the slope of the line. The slope describes how steep a line is by using a positive or negative value. The slope of *a* indicates the average change from year to year, and *b* is the intercept indicating the level of the phenomenon in the zero period (preceding the first year of analysis).

In some disciplines, women's participation in science was high and growth was strong (MED and PHARM), or it was high and growth was weak (BIO); in other disciplines, their participation was low and growth was strong (AGRI, CHEMENG). Finally, in a cluster of disciplines, the share was low and growth was weak: generally math-intensive COMP, ENG, MATH, and PHYS (Figure 5). For the two disciplines with the lowest share of female scientists in both 1990 and 2021—ENG and PHYS—the increase in shares was substantial, but the share was still comparatively low: 13.81% for ENG and 16.53% for PHYS. For all disciplines combined, the increase was substantial, from 22.16% to 38.55% (Table 3).

In all disciplines, the percentage of female scientists has been increasing year after year. At the same time, the rate of this increase has varied. In three disciplines, the slope was 0.71-0.81: ENVIR 0.81 (95% confidence interval for slope was 0.79–0.84, see slope values and their 95% confidence intervals for percentage of women, trend line by academic disciplines), AGRI 0.81 (0.70–0.75), and MED 0.71 (0.68–0.73). There were nine disciplines for which the slope was 0.42–0.61. The lowest increase of female scientists was in disciplines for which the slope was smaller than or equal to 0.33: MATH, COMP, PHYS, and ENG. The changes over time are shown in Figure 5.

In general, all slopes were significantly different from 0 and positive, meaning that there was an upward trend in percentages of female scientists in all disciplines over time. The models' fit was high (with R^2 more than 0.94), with 11 models that best fit the empirical data with an R^2 of 0.97–0.99 and the others with R^2 values of 0.945–0.969. For all models, a p-value < 0.0001 was obtained. The average R^2 for all models was 0.975.

The analysis of the confidence intervals of slopes indicated specific groups average growth rates per year. Each discipline had a different time (in years) of a one percentage point increase in the percentage of female scientists (Table 4). The fastest growth occurred for ENVIR (1.24 years), AGRI (1.37), and MED (1.41). A slightly longer increase of one percentage point was observed for nine disciplines (1.64–2.39 years). For the lowest slope (0.27–0.33), the increase in the percentage of female scientists by one percentage point took 3.03 years for MATH, 3.55 for COMP and PHYS, and 3.69 for ENG.

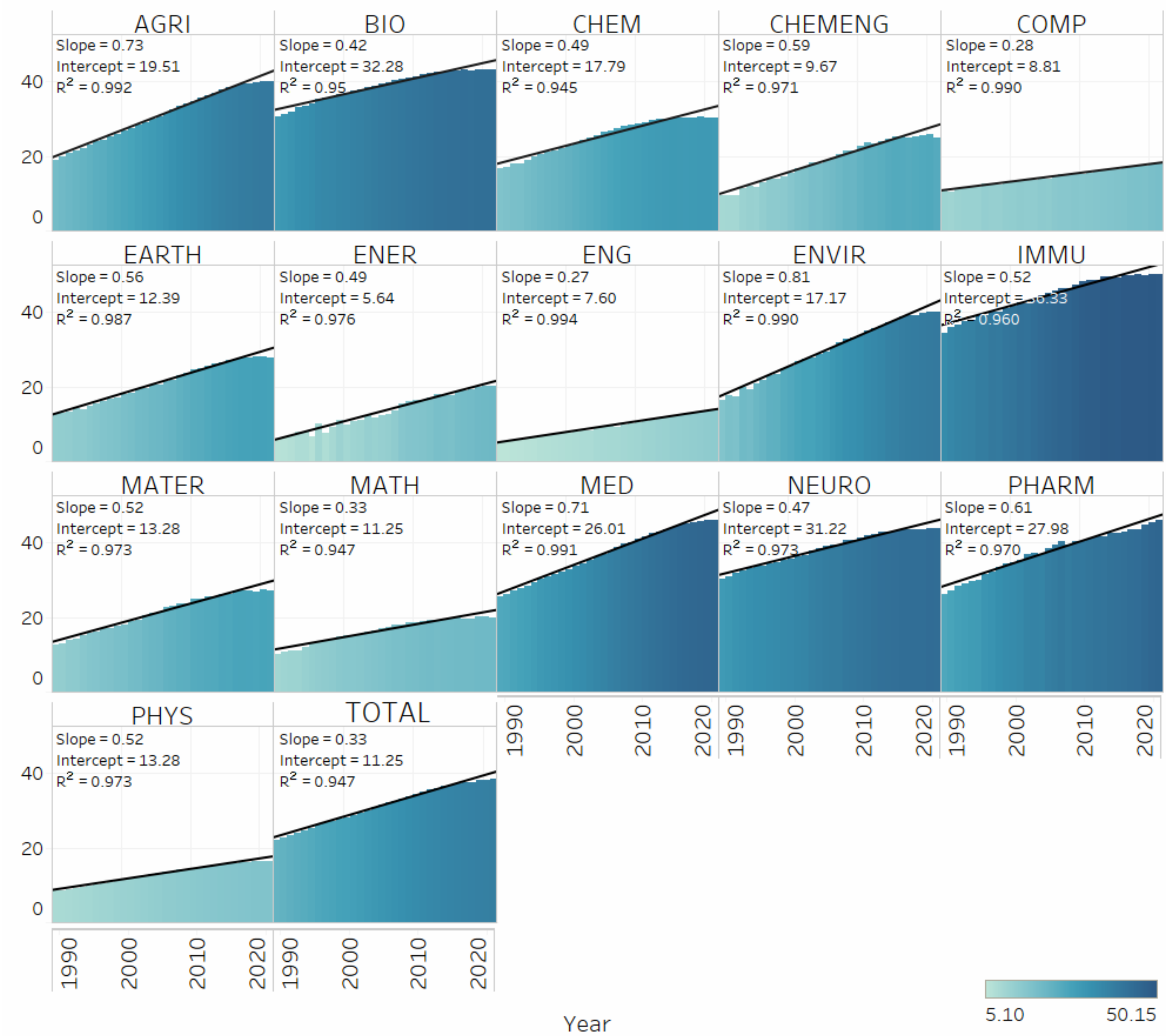

Figure 5. Different starting points and growth in participation of women in science. The trend in the percentage of female scientists by discipline, 1990–2021 (N = 4,314,666)

Table 3. Percentage of female scientists by discipline, 1990–2021 (in five-year intervals).

Discipline	1990	1995	2000	2005	2010	2015	2020	2021
AGRI	19.06	22.90	26.68	30.67	34.62	38.31	40.08	40.13
BIO	30.48	34.10	36.86	39.51	41.37	42.82	43.23	43.27
CHEM	16.68	19.77	22.68	26.34	28.72	30.27	30.36	30.21
CHEMENG	9.35	11.64	15.37	18.95	22.55	25.23	25.80	24.98
COMP	10.81	11.64	13.17	13.82	15.29	16.36	17.87	18.20
EARTH	12.75	14.77	17.83	20.64	24.63	27.09	28.09	27.87
ENER	5.93	6.73	9.71	12.30	16.31	17.74	20.15	20.28
ENG	5.10	6.22	7.71	9.12	10.30	12.15	13.59	13.82
ENVIR	16.32	20.94	25.67	29.46	33.82	37.88	39.96	40.15
IMMU	34.50	39.13	41.54	45.13	48.25	49.32	50.15	50.03
MATER	12.62	15.81	18.11	21.30	24.96	26.24	27.28	27.09
MATH	10.06	12.80	15.29	16.90	18.48	19.46	20.08	20.02
MED	25.69	29.21	32.94	36.96	40.86	44.14	45.88	46.14
NEURO	30.16	33.61	35.53	39.11	41.27	43.06	43.96	43.76
PHARM	26.21	29.86	34.61	38.58	40.52	42.57	45.55	45.98
PHYS	8.50	9.71	11.50	13.09	14.58	15.74	16.47	16.53
Total	**22.16**	**25.28**	**28.06**	**31.39**	**34.31**	**36.83**	**38.24**	**38.55**

3.4. The Global Gender Distribution of Scientists by Age Groups and Disciplines: A Longitudinal View (2000 vs. 2021)

In this section, before moving to young and old scientists in more detail, we discuss how the two age pyramids (or age distributions) changed over a period of two decades. We compare the age pyramids in 2021 and 2000. Longitudinal research permits “the measurement of differences or change in a variable from one period to another” (Menard, 2002, p. 2; in our case at two distinct periods). An age pyramid is made up of a pair of bar graphs, one for men and one for women, turned on their sides and joined, where the vertical axis corresponds to age. For each of the 10 age groups in our population, the bar coming off the axis to the right represents the share of women in that group, and the bar to the left represents the share of men (see Wachter, 2014, pp. 218–221). The age pyramids for disciplines in 2021 (light blue) are superimposed on the age pyramids for 2000 (dark blue). Both age pyramids cover a different population (there are incoming and outgoing scientists in each case); however, some of the cohorts of scientists were found to be common. Scientists included were publishing between 1970 and 2021 (for 2021 data) and 1940 and 1990.

In Figure 6, we show only the percentages of male and female scientists among nonoccasional publishing authors at these two specific points in time; we disregarded the number of authors in these two years. The same sampling allocation principles were used in both cases. This approach enabled us to compare demographics (all age groups) at two points in time, leading us to zoom in on young and old scientists in the next section.

In Figure 6, we show snapshots of 2021 and 2000 by 10 age groups and gender. We display the distribution of male (left) and female scientists (right) by age group in the discipline, indicating the dynamics of change over time. While in Section 3.3 we used trend analysis to show the direction of change in the percentage of female scientists by discipline, here, we added age (or academic experience) to the analysis.

Table 4. Regression model statistics: trends in the percentage of female scientists by discipline, 1990–2021.

	Coefficient – Slope						Coefficient - Intercept				Quality Measures	
Discipline	**Value**	**Standard error**	**t-value**	**p-value**	**LB**	**UB**	**Value**	**Standard error**	**t-value**	**p-value**	**R^2**	**Standard error**
AGRI	0.73	0.012	60.513	< 0.0001	0.704	0.753	19.51	0.217	89.842	< 0.0001	0.992	0.629
BIO	0.42	0.018	23.611	< 0.0001	0.381	0.454	32.28	0.319	101.175	< 0.0001	0.949	0.924
CHEM	0.49	0.021	22.715	< 0.0001	0.443	0.531	17.79	0.387	46.011	< 0.0001	0.945	1.120
CHEMENG	0.59	0.019	31.706	< 0.0001	0.549	0.624	9.67	0.334	28.983	< 0.0001	0.971	0.966
COMP	0.28	0.005	55.445	< 0.0001	0.272	0.292	8.81	0.092	96.007	< 0.0001	0.990	0.266
EARTH	0.56	0.012	47.946	< 0.0001	0.538	0.586	12.39	0.211	58.632	< 0.0001	0.987	0.612
ENER	0.49	0.014	34.941	< 0.0001	0.466	0.524	5.64	0.255	22.089	< 0.0001	0.976	0.739
ENG	0.27	0.004	67.841	< 0.0001	0.263	0.279	7.60	0.072	105.555	< 0.0001	0.994	0.209
ENVIR	0.81	0.015	53.321	< 0.0001	0.778	0.840	17.17	0.274	62.716	< 0.0001	0.990	0.793
IMMU	0.52	0.019	26.908	< 0.0001	0.481	0.560	36.33	0.349	104.162	< 0.0001	0.960	1.010
MATER	0.52	0.016	32.913	< 0.0001	0.483	0.547	13.28	0.282	47.001	< 0.0001	0.973	0.818
MATH	0.33	0.014	23.208	< 0.0001	0.302	0.361	11.25	0.258	43.670	< 0.0001	0.947	0.746
MED	0.71	0.012	57.674	< 0.0001	0.684	0.734	26.01	0.222	117.338	< 0.0001	0.991	0.642
NEURO	0.47	0.014	32.908	< 0.0001	0.436	0.494	31.22	0.255	122.402	< 0.0001	0.973	0.738
PHARM	0.61	0.020	30.966	< 0.0001	0.568	0.649	27.98	0.355	78.940	< 0.0001	0.970	1.026
PHYS	0.28	0.006	47.373	< 0.0001	0.269	0.294	8.60	0.107	80.276	< 0.0001	0.987	0.310
Total	**0.55**	**0.010**	**56.971**	**<0.0001**	**0.535**	**0.574**	**22.66**	**0.176**	**129.059**	**<0.0001**	**0.991**	**0.508**

Table 5. Trends in the percentage of female scientists by discipline (slope, intercept, and speed of change), 1990–2021.

Discipline	Slope	Intercept	Time needed to a 1 p.p. change (in years)
ENVIR	0.81	17.17	1.24
AGRI	0.73	19.51	1.37
MED	0.71	26.01	1.41
PHARM	0.61	27.98	1.64
CHEMENG	0.59	9.67	1.70
EARTH	0.56	12.39	1.78
IMMU	0.52	36.33	1.92
MATER	0.52	13.28	1.94
ENER	0.49	5.64	2.02
CHEM	0.49	17.79	2.05
NEURO	0.47	31.22	2.15
BIO	0.42	32.28	2.39
MATH	0.33	11.25	3.02
COMP	0.28	8.81	3.55
PHYS	0.28	8.60	3.55
ENG	0.27	7.60	3.69
TOTAL	**0.55**	**22.66**	**1.82**

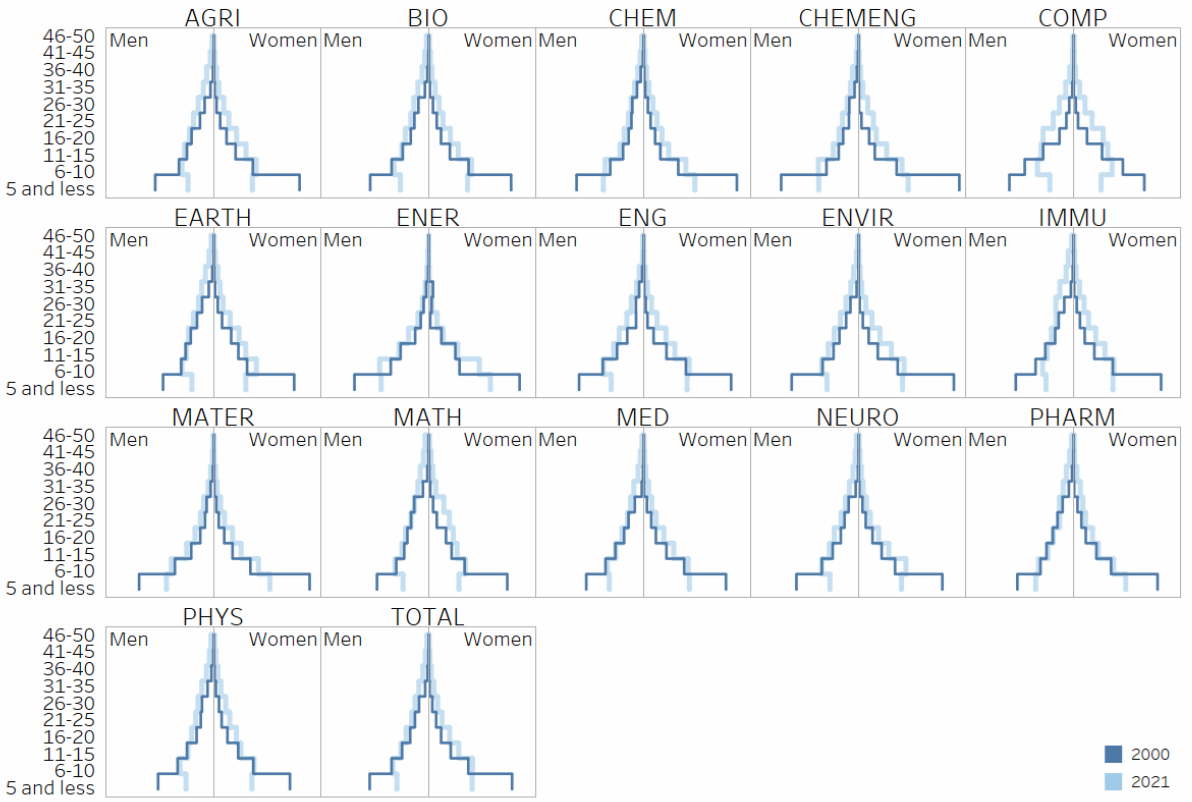

Figure 6. Shrinking percentages of the youngest male and female scientists among all male and female scientists across all disciplines. Overview of change directions in percentages, 2000 vs. 2021: vertical approach. Distribution of nonoccasional publishing STEMM scientists by discipline, age group, and gender (column percentages: 100% vertically for all age groups combined, dark blue 2000, light blue 2021) (N_{2021} = 1,502,792, N_{2000} = 716,796)

In the most general terms, for each discipline, there is a pyramid-like demographic structure of scientists where biological age is replaced with academic or professional age. As expected, for each discipline, the age pyramid narrows at the top and expands to varying degrees at the bottom. The bottom of the age pyramid shows the percentage of young male and female scientists among all male and female scientists, while the top of the age pyramid shows the percentage of older male and female scientists among all male and female scientists. The wider the bottom of the age pyramid, the higher the percentage of young scientists among all scientists; the wider the top of the age pyramid, the higher the percentage of old scientists among all scientists.

A common pattern emerged: for all disciplines in 2021, the bottom of the age pyramid (the first age group, 5 and less years) is narrower than two decades earlier for both male and female scientists (see the details in Table 9 and Table 11 in EDA). Over the period studied, the share of young female scientists among all female scientists within all disciplines has decreased dramatically compared with young male scientists with smaller decreases. The decrease may also mean that young female scientists who entered the academic workforce two decades ago stayed on in the system in 2021, increasing their shares in old age groups. However, the shrinking bottom for female scientists in 2021 compared with 2020 is clearly visible—also for all disciplines combined (Total). In terms of types of age structures in demographics (Rowland, 2014, pp. 98–107), the age structures for all disciplines in 2000 can be classified as "very young," and for 2021 as "young" or "mature."

3.5. Zooming on Young and Old Scientists

3.5.1. A cross-sectional view (2021)

Traditional gender-aggregated and age-aggregated data about scientists in general across disciplines, countries, and institutions hide a much more nuanced picture of the changing gender dynamics within and across disciplines and age groups. In this research, we examined the subpopulation of "young" and "old" scientists (academic age 10 and less years and academic age of 31–50 years, respectively).

3.5.1.1. A vertical approach

Zooming in on young scientists in 2021 vertically (Figure 7), the share of young female scientists among all female scientists exceeded 50% (51.54%), and the share of young male scientists among all male scientists was about 40% (39.82%). For most disciplines, the share exceeded 50% (10 disciplines). It was the lowest for COMP and MATH (below 40%). Most interestingly, for every discipline in the two youngest age groups, the share of young female scientists among all female scientists within a discipline was higher than the share of young male scientists among all male scientists (see the dark parts of bars on the right for each discipline in Figure 7).

Examining old scientists in 2021 vertically (Figure 8), the case was exactly the opposite: for every discipline, the share of old male scientists among all male scientists within a discipline was substantially higher than the share of old female scientists among all female scientists (see the light parts of bars on the left for each discipline in Figure 8). The highest share of old male scientists among all scientists was observed for EARTH, IMMU, MATH, and PHYS (15–17%). The share of old female scientists among all female scientists was much lower than the share of old male scientists among all male scientists (5.43% vs. 12.41%), and the difference between male and female scientists increased with each successive age group: three times higher for the 41–45 age group and four times higher for the 46–50 age group. For all disciplines, female scientists were generally younger, and male scientists were generally older.

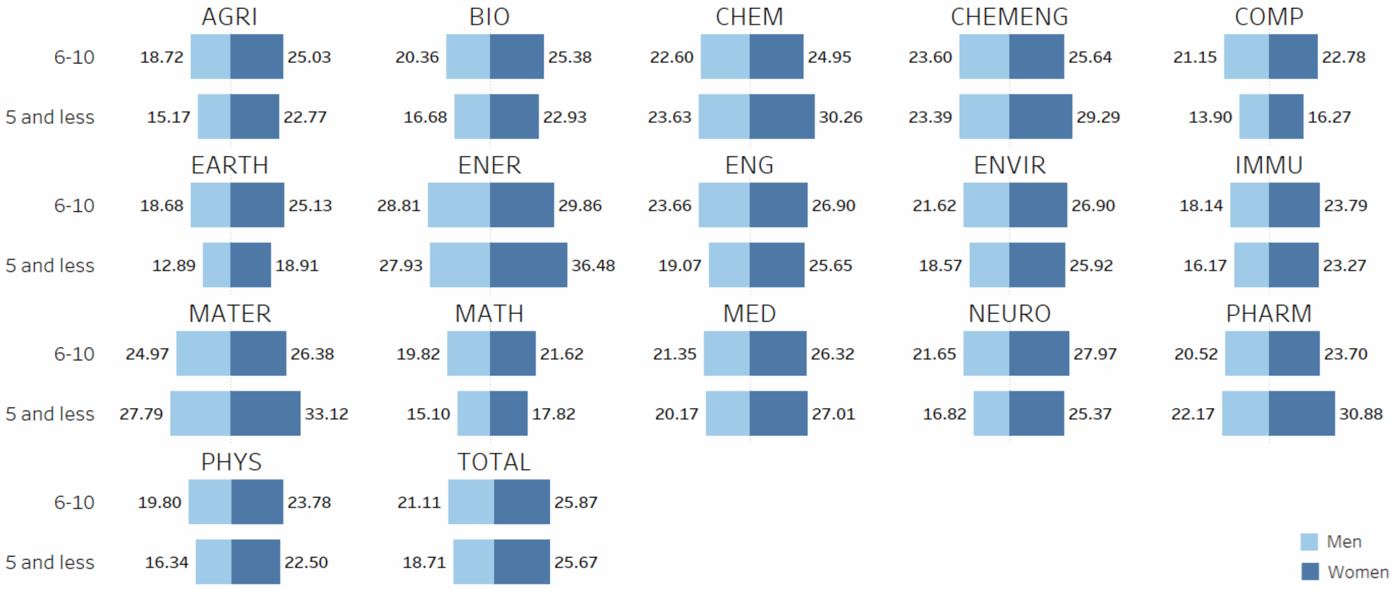

Figure 7. Higher concentration of young women than young men across all disciplines. Vertical approach: zooming in on young scientists only (academic age 10 and less): distribution of publishing nonoccasional STEMM scientists by discipline, age group, and gender (column percentages, vertically: percentage of young female scientists among all women, and young men among all men; women in dark blue), 2021 (N = 666,355)

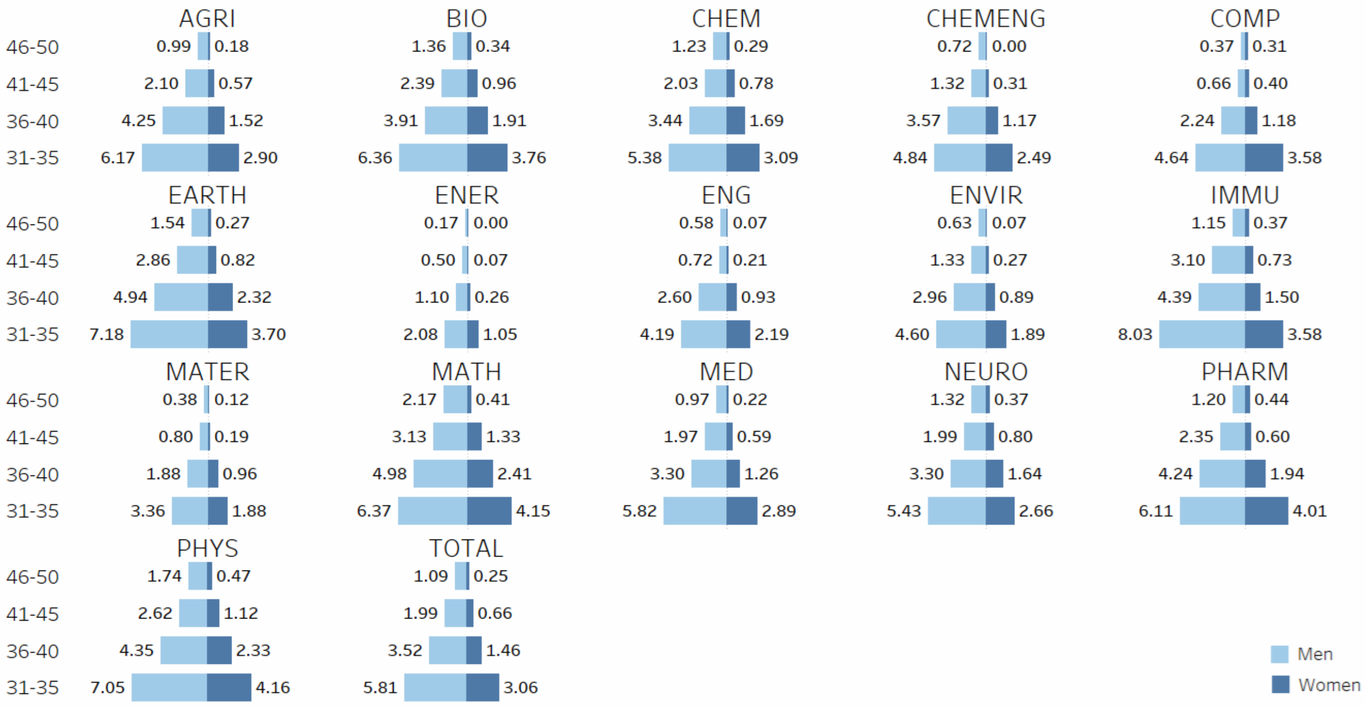

Figure 8. Higher concentrations of old men than old women across all disciplines. Vertical approach: zooming in on old scientists (academic age 31–50 years): distribution of publishing nonoccasional STEMM scientists by discipline, age group, and gender (column percentages, vertically: percentage of old female scientists among all women, and young men among all men; women in dark blue), 2021 (N = 146,090)

3.5.1.2. A horizontal approach

To show gender differences across young and old age cohorts, we compared disciplines by the share of young female scientists among all young scientists and share of old female scientists among all old scientists. We observed to what extent female scientists were present in young cohorts and how female scientists, in contrast, were absent in old cohorts across disciplines (Table 6).

Table 6. The frequencies and percentages of female scientists among publishing nonoccasional scientists by discipline in the two age cohorts (young and old), 2021.

Young scientists (10 years or less)				**Old scientists (31–50 years)**			
Discipline	**All young scientists**	**Young female scientists**	**% female scientists**	**Discipline**	**All old scientists (31–50)**	**Old female scientists**	**% female scientists**
AGRI	41,954	20,389	48.60	AGRI	10,799	2,206	20.43
BIO	89,295	44,533	49.87	BIO	23,377	6,422	27.47
CHEM	36,368	12,394	34.08	CHEM	7,582	1,313	17.32
CHEMENG	2,523	707	28.02	CHEMENG	455	51	11.21
COMP	12,678	2,518	19.86	COMP	2,642	353	13.36
EARTH	18,168	6,363	35.02	EARTH	7,205	1,026	14.24
ENER	4,420	1,013	22.92	ENER	252	21	8.33
ENG	28,808	4,745	16.47	ENG	4,864	307	6.31
ENVIR	16,557	7,758	46.86	ENVIR	2,545	458	18.00
IMMU	5,651	3,270	57.87	IMMU	1,587	430	27.10
MATER	20,664	6,103	29.53	MATER	2,097	323	15.40
MATH	8,327	1,835	22.04	MATH	3,481	386	11.09
MED	324,524	170,004	52.39	MED	60,685	15,775	25.99
NEURO	14,260	7,400	51.89	NEURO	2,903	758	26.11
PHARM	3,341	1,741	52.11	PHARM	744	223	29.97
PHYS	38,817	7,851	20.23	PHYS	14,872	1,370	9.21
TOTAL	**666,355**	**298,624**	**44.81**	**TOTAL**	**146,090**	**31,422**	**21.51**

Although among young scientists (Table 6, left panel) the share of female scientists in several disciplines was about a half (AGRI, BIO, IMMU, NEURO, PHARM, as well as MED), among old scientists (right panel), the share of female scientists was much less notable. Although for all disciplines combined for the young, the share was about 45%, for the old, it was about 20%.

However, for some disciplines, the share of old female scientists was about 10% or lower, meaning that the difference in numbers by gender was at least 10-fold (e.g., ENG, MATH, and PHYS: 6.31%, 11.09%, and 9.21%, respectively). Our data allowed us to examine age-related isolation of female scientists in global science: in old generations of scientists in these four disciplines, female scientists were not just minorities but were isolated individuals among their similar-age colleagues (e.g., globally in ENG, there were 84 female scientists working alongside 1,466 male scientists in the 36–40 age group and 307 alongside 4,557 in all four old age groups combined; similarly in PHYS, there were globally 396 female scientists working alongside 3,726 male scientists in the 36–40 age group and 1,370 alongside 13,502 in all four old age groups combined, Table 7).

So only when we moved from standard gender- and age-aggregated data to gender-disaggregated data for particular age groups could we see what global isolation in such disciplines such as mathematics, physics, and astronomy and engineering—across 38 OECD countries combined—could mean in practical terms. In many institutions, old female scientists were not merely minorities: they were tokens (or single, exemplary scientists representing all female scientists; see Kanter, 1977; on the role of micro-level departmental climates, see Fox & Nikivincze, 2021).

Table 7. Gender- and age-disaggregated data: distribution of nonoccasional publishing STEMM scientists by selected academic age groups and gender, 2021

Discipline	Gender	5 years and less	6–10 years	Total young cohorts	31–35 years	36–40 years	41–45 years	46–50 years	Total Old cohorts
AGRI	Female	9,714	10,675	20,389	1,238	647	244	77	2,206
	Male	9,652	11,913	21,565	3,925	2,702	1,338	628	8,593
BIO	Female	21,139	23,394	44,533	3,463	1,757	887	315	6,422
	Male	20,161	24,601	44,762	7,692	4,726	2,888	1,649	16,955
CHEM	Female	6,793	5,601	12,394	693	380	176	64	1,313
	Male	12,253	11,721	23,974	2,792	1,785	1,052	640	6,269
CHEMENG	Female	377	330	707	32	15	4		51
	Male	904	912	1,816	187	138	51	28	404
COMP	Female	1,049	1,469	2,518	231	76	26	20	353
	Male	4,030	6,130	10,160	1,344	648	190	107	2,289
EARTH	Female	2,732	3,631	6,363	534	335	118	39	1,026
	Male	4,820	6,985	11,805	2,686	1,848	1,069	576	6,179
ENER	Female	557	456	1,013	16	4	1		21
	Male	1,677	1,730	3,407	125	66	30	10	231
ENG	Female	2,316	2,429	4,745	198	84	19	6	307
	Male	10,739	13,324	24,063	2,362	1,466	403	326	4,557
ENVIR	Female	3,807	3,951	7,758	277	130	40	11	458
	Male	4,065	4,734	8,799	1,008	649	291	139	2,087
IMMU	Female	1,617	1,653	3,270	249	104	51	26	430
	Male	1,122	1,259	2,381	557	305	215	80	1,157
MATER	Female	3,397	2,706	6,103	193	98	20	12	323
	Male	7,670	6,891	14,561	927	520	222	105	1,774
MATH	Female	829	1,006	1,835	193	112	62	19	386
	Male	2,808	3,684	6,492	1,185	926	581	403	3,095
MED	Female	86,100	83,904	170,004	9,217	4,005	1,865	688	15,775
	Male	75,065	79,455	154,520	21,655	12,289	7,338	3,628	44,910
NEURO	Female	3,520	3,880	7,400	369	227	111	51	758
	Male	3,000	3,860	6,860	968	588	354	235	2,145
PHARM	Female	985	756	1,741	128	62	19	14	223
	Male	831	769	1,600	229	159	88	45	521
PHYS	Female	3,817	4,034	7,851	705	396	190	79	1,370
	Male	13,998	16,968	30,966	6,040	3,726	2,247	1,489	13,502
TOTAL	Female	148,749	149,875	298,624	17,736	8,432	3,833	1,421	31,422
	Male	172,795	194,936	367,731	53,682	32,541	18,357	10,088	114,668

However, the time context is important: for the same three disciplines of ENG, MATH, and PHYS, the isolation of young female scientists decreased at least twice, from 10 times for their older colleagues to 5 times (to 16.47% for ENG, 22.04% for MATH, and 20.23% for PHYS). Female scientists in young cohorts were at least twice as visible in this 2021 snapshot as female scientists in the old cohorts.

Across all disciplines, both those heavily male-dominated (such as ENG, MATH, and PHYS) and those closest to gender parity (such as MED, AGRI, and BIO), younger generations generally always had more female scientists and their higher percentages than older generations. Female scientists were more present in numbers and more present in percentages going down the 10 age groups and when moving from the cohort of old scientists to that of young scientists (Table 6).

Thus, zooming in on young (Figure 9) and old scientists (Figure 10) horizontally, for all disciplines except 6 (AGRI, BIO, IMMU, MED, NEURO, and PHARM), there were more youngest male scientists than youngest female scientists, and there were much more old male scientists than old female scientists (see the dark blue parts of bars on the right for each discipline). The smallest shares of young female scientists compared with young male scientists were for COMP, ENG, MATH, and PHYS, here in the range of 17% to 22%.

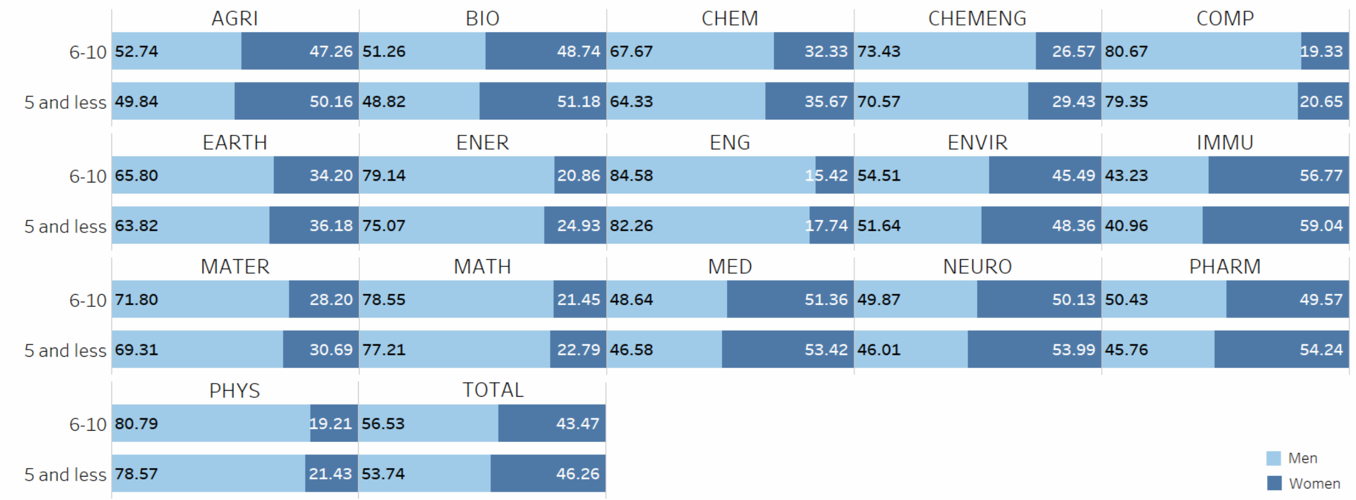

Figure 9. More young women than young men in the six disciplines. Horizontal approach: zooming on young scientists only (academic age 10 years and less). Distribution of young publishing nonoccasional STEMM scientists by discipline, age group, and gender (row percentages: 100% horizontally), 2021 (N = 666,355)

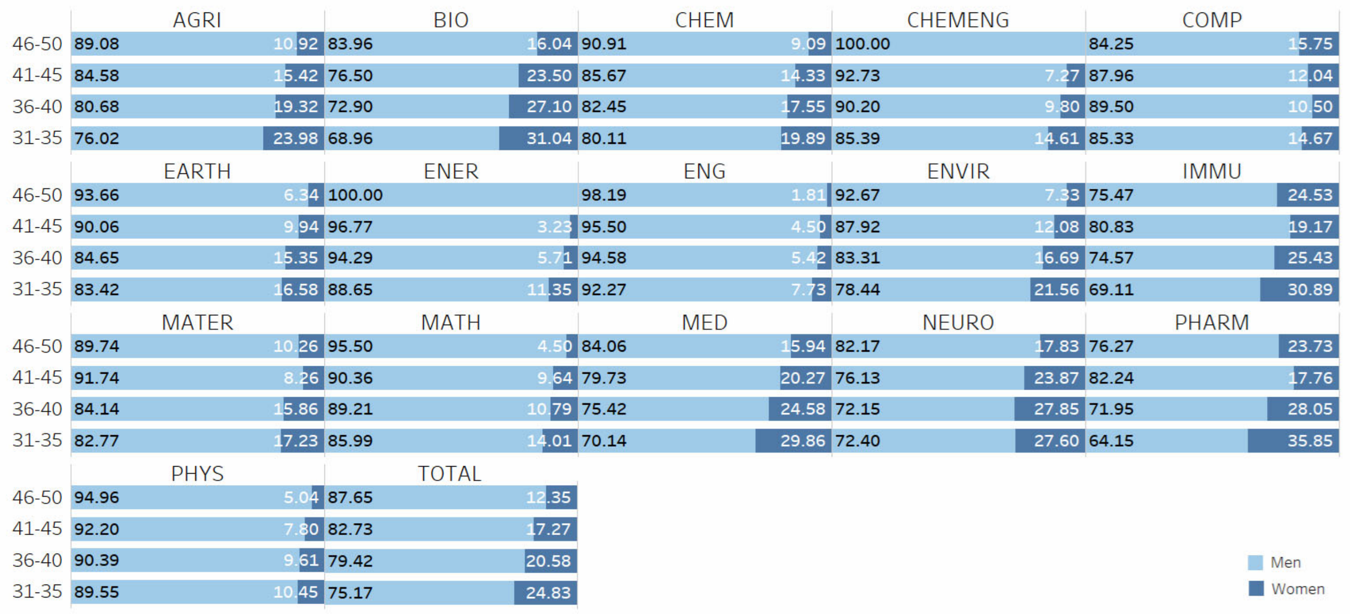

Figure 10. More old men than old women in all disciplines. Horizontal approach: zooming in on old scientists only (academic age 31–50 years). Distribution of young publishing nonoccasional STEMM scientists by discipline, age group, and gender (row percentages: 100% horizontally), 2021 (N = 146,090)

3.5.2. Zooming on young and old scientists: A longitudinal view (2000 vs. 2021)

3.5.2.1. A vertical approach

From a longitudinal perspective, comparing the shares of young male and female scientists at two points in time (2000 and 2021) within disciplines, the vertical pattern is clear: for all disciplines, the share of scientists in the youngest age group in 2000 was higher for both male and female scientists. Vertically, scientists in 2000 were younger than in 2021 (Figure 11, dark blue bars). There was a lower percentage of scientists in 2021 than in 2020 in each youngest category in each discipline, with no exceptions. The shares of the youngest scientists for all disciplines combined decreased in these two decades about two times: from 34.63% to 18.71% for male scientists and from 49.26% to 25.67% for female scientists. Regarding age group, 6–10 remained generally stable (21.80% and 21.11% for male and 23.38% and 25.87% for female scientists).

Comparing the shares of old male and female scientists in 2000 and 2021 within disciplines, the pattern is also clear: the shares of both male and female scientists in the four old age groups were much higher in 2021 than in 2020 (Figure 12, dark blue bars). There was a higher percentage of old scientists in 2021 than in 2020 in each old category in each discipline, with no exceptions. This is another dimension of the graying of the scientific workforce.

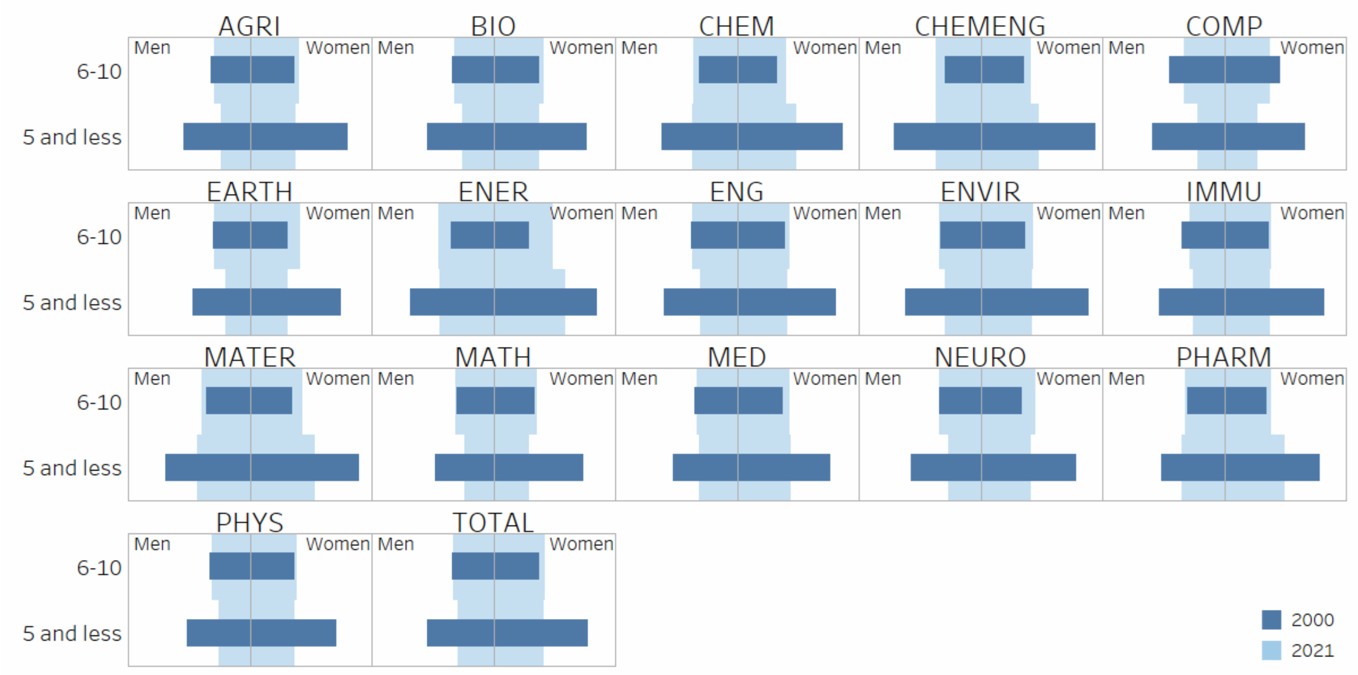

Figure 11. Shrinking base of young scientists, both men and women. Overview of percentage change directions, 2000 vs. 2021: vertical approach. Zooming in on young scientists only (academic age 10 years or less). Distribution of young publishing nonoccasional STEMM scientists by discipline, age group, and gender, 2000 (dark blue) and 2021 (light blue) (based on column percentages) (N_{2021} = 666,355, N_{2000} = 437,113)

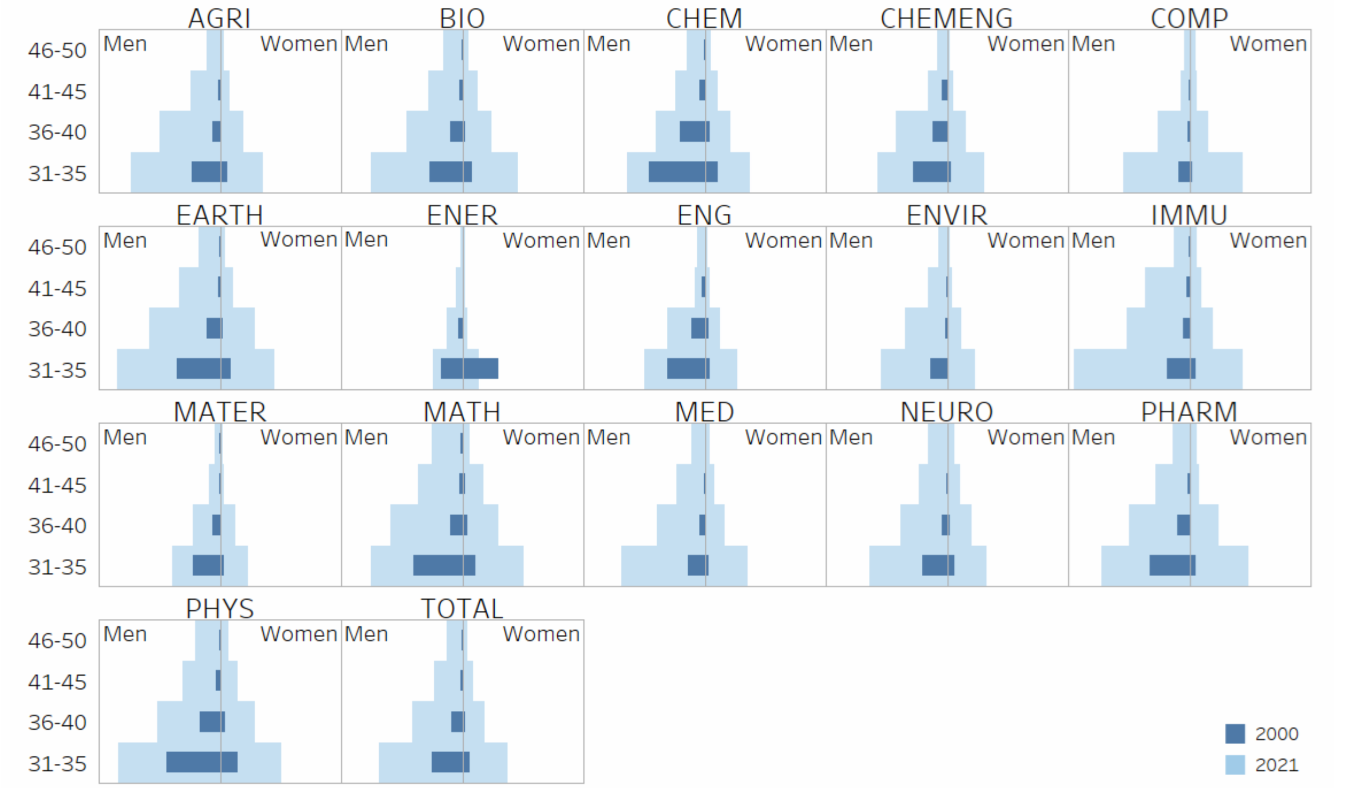

Figure 12. Expanding the base of old scientists, both men and women. Overview of change directions, 2000 vs. 2021: vertical approach. Zooming in on old scientists only: academic age of 31–50 years. Distribution of old publishing nonoccasional STEMM scientists by discipline, age group, and gender, 2000 (dark blue) and 2021 (light blue) (based on column percentages) (N_{2021} = 146,090, N_{2000} = 17,463)

3.5.2.2. A horizontal approach

A change between 2000 and 2021 from a vertical perspective needs to be complemented with a change from a horizontal perspective: the shares of male and female scientists within the young and old age groups horizontally. The direction of changes was unambiguous: for all disciplines, the share of female scientists increased for the young age groups (Figure 13), and the share of female scientists increased for the old age groups (Figure 14). The white lines show the shares of female scientists for 2000, while the dark blue bars on the right show this for 2021. For the youngest age group, for all disciplines combined, the share of female scientists increased from one-third to half (from 34.93% to 50.16%), indicating that the share of male scientists decreased from two-thirds to half (from 65.07% to 49.84%). Comparing the old age category of 31–35, the share of female scientists increased three times, from 8.12% to 23.98%.

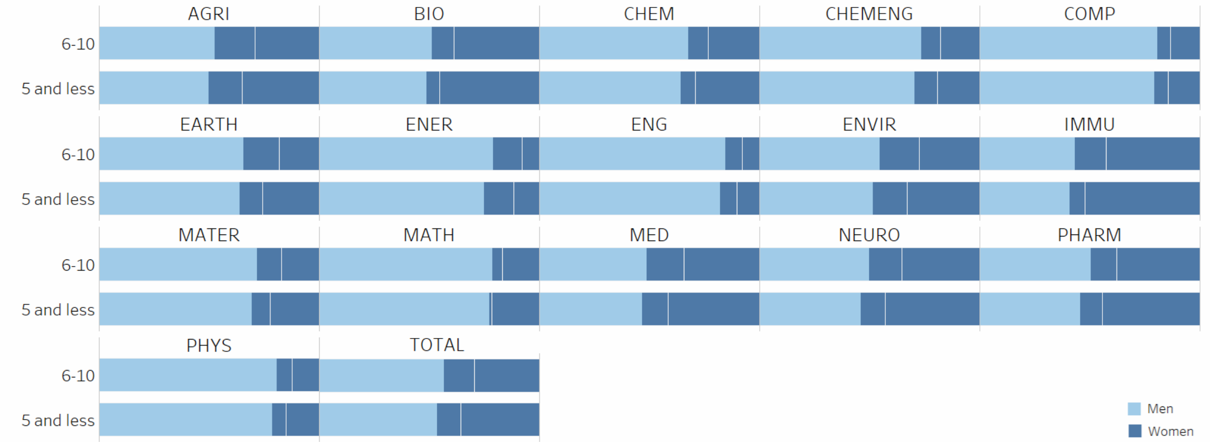

Figure 13. For all disciplines, the participation of young female scientists has increased. Overview of percentage change directions, 2000 vs. 2021: horizontal approach. Zooming in on young scientists only (academic age 10 years or less). Distribution of young publishing nonoccasional STEMM scientists by discipline, age group, and gender; dark blue percentage female scientists 2021, white lines percentage female scientists 2000 (row percentages: 100% horizontally) (N_{2021} = 666,355, N_{2000} = 437,113)

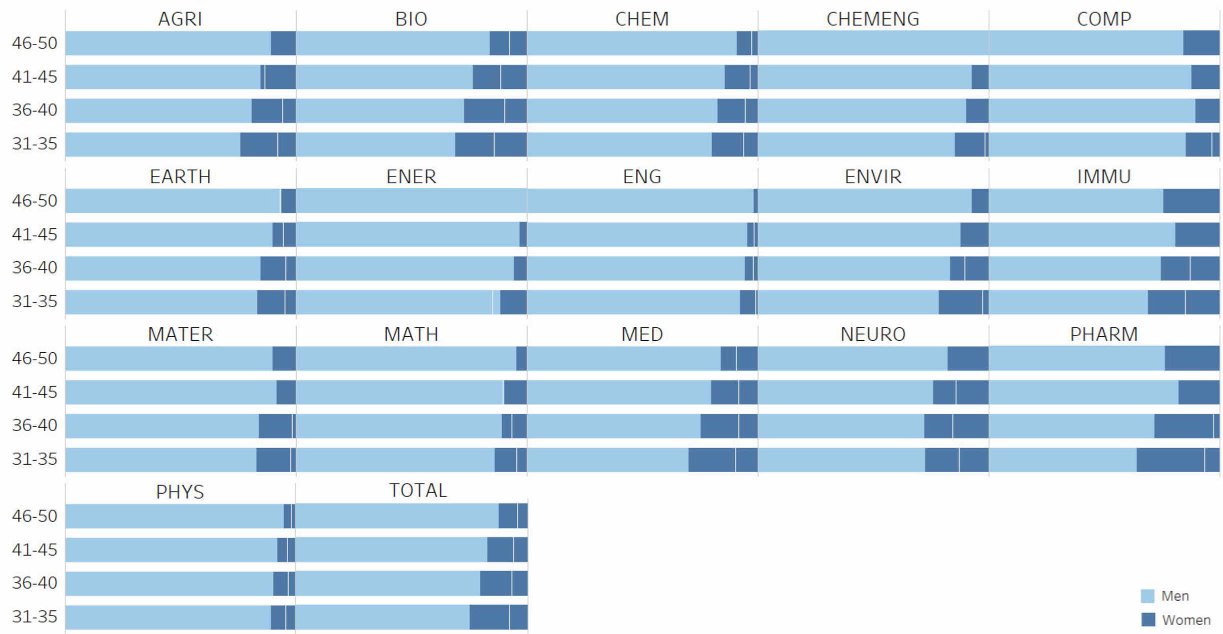

Figure 14. For almost all disciplines, the participation of old female scientists increased. Overview of the change directions, 2000 vs. 2021: horizontal approach. Zooming in on old scientists only

(academic age 31–50 years). Distribution of old publishing nonoccasional STEMM scientists by discipline, age group, and gender; dark blue female scientists 2021, white lines female scientists 2000 (row percentages: 100% horizontally) (N_{2021} = 146,090, N_{2000} = 17,463)

4. Summary, Discussion and Conclusions

We have examined the changing demographics of the global scientific workforce over the past three decades, with special emphasis on the changing participation in science of young male and female scientists. Our research was large scale (4.3 million scientists); generational (scientists were allocated to 10 academic age groups, with a major distinction between the young cohort, academic experience 10 or less years, and the old cohort, 31–50 years); and longitudinal (covering the 1990 to 2021 period and 2000 vs. 2021).

We combined two approaches to examine the four dimensions (gender, age, discipline, and time) comprehensively: in a horizontal approach, we focused on the gender distribution of scientists within the same age groups across disciplines; and in the vertical approach, we focused on the concentration of male and female scientists separately across age groups and within disciplines.

Our underlying methodological choice was to use individual scientists (with their attributes) rather than individual publications (with their characteristics) as a unit of analysis. We used raw data from the Scopus dataset because our research heavily relied on author identifiers and because Scopus provided bibliometric data with a precision of 98.1% and recall of 94.4% (Baas et al., 2020). Our study was quantitative and exploratory in nature: we asked the "what" questions without asking "why." Therefore, the present research can be complemented with further small-scale quantitative studies (based on global and national survey data) and qualitative studies based on interview and focus group methodologies (as Fox 2020 suggests in studying gender and rank). We are not aware of a similar research exercise mapping men and women in global science, specifically mapping young men and women scientists across disciplines in the context of older age groups (in terms of academic or professional experience).

The scientific workforce has been changing in terms of its gender and age composition, with different intensities in different disciplines. These changes have been ongoing and global in nature. Among the 16 STEMM disciplines, most were currently numerically dominated by men, but some were already dominated by women, and the change processes seemed to be fast in some and slow in other disciplines. A somewhat surprising finding, even in the context of the COVID-19 pandemic, was the pivotal role of medical research for the global scientific workforce, especially for women scientists: almost half of all scientists (45.98%) were defined in our methodology as doing medical research (a dominating discipline, based on cited references from lifetime publications). The concentration of female scientists was steep across disciplines: more than half (55.02%) were located in MED and 1 in 7 (15.91%) in BIO. Consequently, about 70% (70.93%) of all female scientists globally, across all science sectors, were concentrated just in these two disciplines.

The traditional narratives about some STEMM disciplines being much more heavily male dominated than others have been confirmed: women's participation in COMP, ENG, MATH, and PHYS was very low (and smaller than 20% in 2021). In most disciplines in 2021, the share of female scientists in each successive younger cohort was higher (and it was usually the highest for the youngest cohort: scientists with 5 or less years of academic experience); for COMP, ENG, MATH, and PHYS, however, the principle did not hold, with very small intercohort differences (Figure 3).

Our trend analysis of the 1990–2021 period showed that the participation of women scientists in global science increased across all disciplines, albeit with different starting points in 1990 and different intensities. For the least increasing trends, the increase in the percentage of female scientists by one percentage point took 3.03 years for MATH, 3.55 for COMP and PHYS, and 3.69 years for ENG.

However, from an age-disaggregated perspective, in 6 out of 16 disciplines, there were already more youngest female than male scientists (IMMU, PHARM, NEURO, MED, AGRI, BIO), and the discipline most open to female scientists has been IMMU (59.04%). Interestingly, more than 8 out of 10 female scientists globally worked in these six disciplines (82.90%). Across all disciplines combined, the majority of women currently involved in publishing articles were young women (with 10 years of academic experience or less).

Most interestingly, there was a higher concentration of young women than young men across all disciplines, and there was a higher concentration of old men than old women across all disciplines. For every discipline, the share of young female scientists among all female scientists within a discipline was higher than the share of young male scientists among all male scientists. For every discipline, the share of old male scientists among all male scientists within a discipline was substantially higher than the share of old female scientists among all female scientists. The patterns are clear: for all disciplines, female scientists were generally younger and male scientists generally older.

Moving from standard data to gender-disaggregated data for particular age groups, we begin to understand what the global isolation of female scientists in such disciplines as MATH, PHYS, ENG means. In these disciplines in 2021, the share of old female scientists was about 10% or less (the difference in numbers by gender was about 10-fold or higher, e.g., ENG, MATH, and PHYS: 6.31%, 11.09%, and 9.21%, respectively). In older generations, female scientists were isolated individuals among their similar-age male colleagues. The numbers show more than percentages (Table 7): for instance, in the 36–40 academic age group, there were globally 84 female scientists working alongside 1,466 male scientists in ENG and 396 female scientists working alongside 3,726 male scientists in PHYS.

However, the context of changing times is important: for the same three disciplines of ENG, MATH, and PHYS, the isolation of young female scientists powerfully decreased, from a 10-times difference for older cohorts to a 5-times difference for young cohorts (i.e., to 16.47% for ENG, 22.04% for MATH, and 20.23% for PHYS). In these three male-dominated disciplines in 2021, female scientists in young cohorts were at least twice as present as female scientists in older cohorts (on the role of gender team composition in science, see Fox & Mohapatra, 2007).

The change in gender participation in science has been gradual and the pattern unambiguous: across all disciplines, both those heavily male dominated and those closest to gender parity, the younger generations have generally always more female scientists and their higher percentages than older generations. Female scientists were more present in numbers and more present in percentages going down the 10 age groups and when moving from the cohort of old scientists to that of young scientists. From a longitudinal perspective, for all disciplines, the share of scientists in the youngest age group in 2000 was higher than in 2021 for both male and female scientists. There was a shrinking base of young scientists, both men and women, and there was an expanding base of old scientists, both men and women.

A broader takeaway message is that there are no other data sources than bibliometric sources to assess the global (rather than merely national) gender, disciplinary, and age group distribution, either cross-sectionally or from a longitudinal perspective. Changing the focus from publications to individual scientists opens a new perspective to study scientific careers, albeit with new limitations. New knowledge comes at a price that needs to be assessed. Specifically, it is heuristically useful to treat global bibliometric datasets as "structured" (as opposed to "unstructured" and "semistructured") Big Data, large in size and complexity, with which new algorithmic techniques are required to extract useful information (Holmes, 2017).

Most limitations of bibliometric datasets have been discussed for years (English language and STEMM focus, Anglo-Saxon bias, articles only, etc.; see Sugimoto & Larivière, 2018, pp. 38–44 on "cultural biases of data sources"). However, our use of a bibliometric dataset to define the individual attributes of the global scientific workforce requires a brief discussion of new limitations:

(1) Gender determination: A binary approach was used with different coverage for different countries as algorithms used by Scopus (and other gender-determining tools such as, e.g., Genderize.io or Gender Guesser, see Halevi, 2019, p. 566; Mihaljević & Santamaría, 2020: 1477-1478) work much better for some rather than for other countries; all gender-unknown cases were removed from our analysis.

(2) Discipline determination: A commercial academic journal classification was used as a proxy for the richness of nationally defined academic disciplines and lifetime Scopus-indexed publication history, with lifetime cited references being used to determine a single attribute of discipline (a single dominant value, possibly suppressing the changes between disciplines over time).

(3) Determining the country of affiliation: A single dominant value, possibly suppressing individual lifetime migration histories.

(4) Determining scientists' nonoccasional status: The threshold of three articles as an entry condition for inclusion in the population was arbitrary, underplaying the role of scientists in very early stages of academic careers; a higher threshold would decrease the population, and a lower threshold would increase it.

(5) Determining academic age: Although the correlation between biological age and academic age in STEMM disciplines was high (and possibly higher than 0.9, as we have shown for a sample of 20,000 Polish scientists with doctorates; Kwiek & Roszka, 2022b), the first publications in individual lifetime publication histories may appear in different moments of academic lives in different disciplines; additionally, publishing patterns clearly change over time; that is, scientists tend to start publishing earlier in their careers today than before.

Another takeaway is that there were clear differences between national-level studies, especially when bibliometric data were merged with administrative and biographical data, and a global study of the academic workforce and careers. In short, national studies can use commercial and noncommercial datasets available for a few countries only (e.g., the USA, Norway, Poland, and Italy: see Savage & Olejniczak, 2021; Abramo et al., 2022; Abramo et al., 2016), which may include globally directly unavailable biographical information such as gender, date of birth, dates of PhD and other degrees and ranks, national discipline classifications, and full employment history. In our longitudinal study of changing productivity classes of 2,343 full professors over 20–40 years of their careers (Kwiek & Roszka, 2022c) and in our study of the impact of early and late, as well as fast and slow promotions on productivity on a sample of 16,000 STEMM university professors (Kwiek & Roszka, 2022d), our dataset of about a million Polish Scopus-indexed publications from the past 50 years was enriched with full biographical and administrative data of 100,000 Polish scientists. In global studies, as opposed to national studies, biological age needs to be examined through a proxy of academic or professional age, gender needs to be inferred with probability thresholds, academic ranks should be used through a proxy of career length from the first publication, and national prestige ranks should be used through a proxy of global rankings. All scientists registered nationally must be replaced in global studies with publishing-only scientists, with Scopus- (or WoS-) indexed publications. Real scientists with national identification numbers available in national databases need to be replaced with Scopus Author IDs, and near-perfect administrative and biographical data need to be replaced with either inferred data or proxies. However, global exploratory research, provisionally mapping the terrain and testing the best tools and methodologies, is interesting in its generality before more sophisticated analyses arise.

The scholarly and policy implications of the present research are manifold. In scholarly terms, we make the first attempt to define the scientific community globally through attributes so far understudied on a large scale. The mapping of changing gender and age distribution of scientists globally over time, as well as a glimpse of the global scientific workforce today, opens science (and academic) profession studies to more detailed questions. The scientific workforce is often discussed in two policy contexts: the aging and accompanying problems for higher education and innovation systems and access to the science profession of young scientists. Our methodological approach and findings can be useful in examining the complex policy issue of entering and leaving the science profession, with the accompanying questions about changing productivity over scientists' life cycles, aging and changing publishing and collaboration patterns, changing academic time and work effort distribution, and so forth (especially in the academic sector).

Our research can be useful for policymakers, administrators, and large grant-making organizations, showing where the scientific workforce has been focusing their research efforts, how large segments of academics are involved in studies in particular disciplines, and where male and female scientists are disciplinary located. Our mapping of substantial gender differences between the various STEMM disciplines (and especially between ENG, COMP, MATH, and PHYS versus all others) may provide new empirical grounds useful in discussing women's participation in science and its discipline-based social, institutional, and political impediments.

Acknowledgments

Marek Kwiek is grateful for the comments from the host and audience of the RISIS (Research Infrastructure for Science and Innovation Policy Studies) Research Seminar held at the University of Lugano, Switzerland, on May 11, 2022: "The Changing Demographics of the Global Academic Workforce in OECD Countries, 1990–2020," which improved our methodological approach. We gratefully acknowledge the assistance of the International Center for the Studies of Research (ICSR) Lab and Kristy James, Senior Data Scientist. We also want to thank Dr. Wojciech Roszka from the CPPS Poznan Team for fruitful discussions.

Author contributions

Marek Kwiek: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing—original draft, Writing—review & editing. Lukasz Szymula: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing.

Competing interests

The authors have no competing interests.

Funding information

We gratefully acknowledge the support provided by NDS grant no. NdS/529032/2021/2021.

Data availability

We used data from Scopus, a proprietary scientometric database. For legal reasons, data from Scopus received through collaboration with the International Center for the Studies of Research (ICSR) Lab cannot be made openly available.

Electronic Data Appendices (EDA):
Tables 8–18, are available from GitHub:
https://github.com/lukaszszy/Young_Male_and_Female_Scientists

References

Abramo, G., D'Angelo, C. A., & Murgia, G. (2016). The combined effects of age and seniority on research performance of full professors. *Science and Public Policy, 43*(3), 301–319.

Abramo, G., Aksnes, D. W., & D'Angelo, C. A. (2021). Gender differences in research performance within and between countries: Italy vs Norway. *Journal of Informetrics,* 15(2), 101144.

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386. 10.1162/qss_a_00019

Boekhout, H., van der Weijden, I., & Waltman, L. (2021). Gender differences in scientific careers: A large-scale bibliometric analysis. Preprint: https://arxiv.org/abs/2106.12624

Boothby, C., Milojevic, S., Larivière, V., Radicchi, F., & Sugimoto, C. (2022). Consistent churn of early career researchers: An analysis of turnover and replacement in the scientific workforce. Preprint: https://doi.org/10.31219/osf.io/hdny6

Börner, K. (2010). *Atlas of science: Visualizing what we know*. MIT Press.

Elsevier. (2020). *The researcher journey through a gender lens.* Elsevier.

Fox, M. F. (2020). Gender, science, and academic rank: Key issues and approaches. *Quantitative Science Studies, 1*(3), 1001–1006.

Fox, M. F., & Mohapatra, S. (2007). Social-organizational characteristics of work and publication productivity among academic scientists in doctoral-granting departments. *Journal of Higher Education*, *78*(5), 542–571.

Fox, M. F., & Nikivincze, I. (2021). Being highly prolific in academic science: Characteristics of individuals and their departments. *Higher Education, 81,* 1237–1255.

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). Springer.

Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, *117*(9), 4609–4616.

Ioannidis, J. P. A., Boyack, K. W., & Klavans, R. (2014). Estimates of the continuously publishing core in the scientific workforce. *PLOS One.* 9(7): e101698.

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, *82*(5), 965–990.

King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., & West, J. D. (2017). Men set their own cites high: Gender and self-citation across fields and over time. *Socius*, *3*.

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. (2018). High research productivity in vertically undifferentiated higher education systems: Who are the top performers? *Scientometrics*, *115*(1), 415–462.

Kwiek, M. (2019). *Changing European academics. A comparative study of social stratification, work patterns and research productivity.* London and New York: Routledge.
Kwiek, M. (2020). Internationalists and locals: International research collaboration in a resource-poor system. *Scientometrics*, *124*, 57–105.
Kwiek, M., & Roszka, W. (2021a). Gender disparities in international research collaboration: A large-scale bibliometric study of 25,000 university professors. *Journal of Economic Surveys*, *35*(5), 1344–1388.
Kwiek, M., & Roszka, W. (2021b). Gender-based homophily in research: A large-scale study of man-woman collaboration. *Journal of Informetrics*, *15*(3), 1–38.
Kwiek, M., & Roszka, W. (2022a). Are female scientists less inclined to publish alone? The gender solo research gap. *Scientometrics, 127,* 1697–1735.
Kwiek, M., & Roszka, W. (2022b). Academic vs. biological age in research on academic careers: A large-scale study with implications for scientifically developing systems. *Scientometrics*, *127*, 3543–3575.
Kwiek, M., & Roszka, W. (2022c). Once highly productive, forever highly productive? Full professors' research productivity from a longitudinal perspective. *ArXiv,* preprint. https://arxiv.org/abs/2206.05814
Kwiek, M., & Roszka, W. (2022d). The young and the old, the fast and the slow: Age, productivity, and rank advancement of 16,000 STEMM university professors. *ArXiv,* preprint. https://arxiv.org/abs/2211.06319
Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C.R. (2013). Global gender disparities in science. *Nature, 504*, 211–213.
Menard, S. (2002). *Longitudinal research*. Sage.
Mihaljević, H., & Santamaría, L. (2020). Authorship in top-ranked mathematical and physical journals: Role of gender on self-perceptions and bibliographic evidence. *Quantitative Science Studies*, 1(4), 1468–1492.
Nielsen, M. W., & Andersen, J. P. (2021). Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences, 118*(7), e2012208118.
Nygaard, L. P., Piro, F., & Aksnes, D. (2022). Gendering excellence through research productivity indicators. *Gender and Education, 34*(6), 690–704.
Robinson-Garcia, N., Costas. R., Sugimoto, C. R., Larivière, V., & Nane, G. F. (2020). Task specialization across research careers. *eLife*, *9*, e60586. 10.7554/eLife.60586
Rowland, D. T. (2014). *Demographic methods and concepts*. Oxford University Press.
Salganik, M. J. (2018). *Bit by bit. Social research in a digital age.* Princeton University Press.
Savage, W. E., & Olejniczak, A. J. (2021). Do senior faculty members produce fewer research publications than their younger colleagues? Evidence from Ph.D. granting institutions in the United States. *Scientometrics, 126*, 4659–4686.
Selwyn, N. (2019). *What is digital sociology?* Polity Press.
Sugimoto, C., & Larivière, V. (2018). *Measuring research: What everyone needs to know.* Oxford University Press.
Wachter, K. W. (2014). *Essential demographic methods.* Harvard University Press.
Way, S. F., Morgan, A. C., Clauset, A., & Larremore, D. B. (2017). The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences, 114*(44), E9216–E9223. 10.1073/pnas.1702121114
Zhang, S., Wapman, K. H., Larremore, D. B., & Clauset, A. (2022). Labor advantages drive the greater productivity of faculty at elite universities. *Science Advances*, 8(46).