Quantifying Attrition in Science: A Cohort-Based, Longitudinal Study of Scientists in 38 OECD Countries

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

Lukasz Szymula
(1) Faculty of Mathematics and Computer Science, Adam Mickiewicz University of Poznan, Poznan, Poland, and

(2) Department of Computer Science, University of Colorado Boulder, USA
ORCID: orcid.org/0000-0001-8714-096X

Abstract

In the present research, we explore how members of the global scientific community leave academic science and how attrition differs across genders, academic disciplines, and over time. Our approach is global, cohort-based, and longitudinal: we track individual male and female scientists over time and quantify the phenomenon traditionally referred to as “leaving science.” Using publication metadata from Scopus—a global bibliometric database of publications and citations—we follow the details of the publishing careers of scientists who started publishing in 2000 ($N = 142{,}776$) and 2010 ($N = 232{,}843$). Our study is restricted to 16 STEMM disciplines (science, technology, engineering, mathematics, and medicine), and we track the individual scholarly output of the two cohorts until 2022. Survival analyses show that attrition becomes ever less gendered, while regression models show that publication quantity is more consequential than publication quality for careers. Behind the aggregated changes at the level of all STEMM disciplines combined, widely nuanced changes were found to occur at the level of disciplines and over time. Attrition in science means different things for men versus women depending on the discipline; moreover, it means different things for scientists from different cohorts entering the scientific workforce. Finally, global bibliometric datasets were tested in this study, opening new opportunities to explore gender and disciplinary differences in attrition.

Keywords:

Academic careers; Leaving science; Big data; Kaplan-Meier curves; Survival analysis; Attrition and retention; Regression models; Science of science

1. Introduction

In the present research, we explore how members of the global scientific community leave academic science (in our conceptualization, they cease publishing in academic journals) and how attrition differs across genders, academic disciplines, and over time. Our approach is global, cohort-based, and longitudinal: we track male and female scientists over time and quantify the phenomenon traditionally referred to as "leaving science" (Geuna & Shibayama, 2015; Preston, 2004; White-Lewis et al., 2023; Zhou & Volkwein, 2004). Our focus is on leaving science viewed as ceasing scholarly publishing, as large-scale longitudinal data on leaving academia are not currently available at a global level.

Using publication metadata from Scopus—a global bibliometric database of publications and citations—we follow the details of the publishing careers of 142,776 scientists from 38 OECD countries who began publishing in 2000 and 232,843 scientists who started publishing in 2010 (termed the 2000 and 2010 cohorts). Our study is restricted to 16 STEMM disciplines (science, technology, engineering, mathematics, and medicine), and we track the individual scholarly output of the two cohorts until 2022.

Bibliometric metadata are a perfect example of the digital traces left in publications used to study scientists who have traditionally been explored through surveys, interviews, and administrative and census data. Digitalized scholarly databases with bibliometric information are a new source for studying scientists as a population, although they first need to be repurposed to focus on individual scientists rather than individual publications. This allows for the exploration of questions about science and scientists at an unprecedented level of detail (Kashyap et al., 2022; Liu et al., 2023; Wang & Barabási, 2021). Individuals can be studied according to age, seniority, gender, discipline, and institutional type—and most importantly for the present study, scientists can be tracked over time. Moreover, cross-sectional studies can be complemented by longitudinal studies in which individual academic careers are tracked for years and decades. Using massive datasets, academic careers have been examined globally and nationally (e.g., Nielsen & Andersen, 2021; King et al., 2017; Boekhout et al., 2021; Nygaard et al., 2022; Spoon et al., 2023).

Traditionally, women tend to leave academia earlier than men. Moreover, they tend to do so in higher proportions than men, as the traditional narrative on gender differences in attrition in higher education suggests (Alper, 1993; Blickenstaff, 2005; Deutsch & Yao, 2014; Goulden et al., 2011; Preston, 2004; Shaw & Stanton, 2012). In our examination of attrition, we move beyond traditional cross-sectional approaches (single or repeated snapshots in time) and instead use global longitudinal data at the level of individual scientists. Furthermore, we test new possibilities opened up by global bibliometric datasets for the large-scale quantification of attrition in scientific careers.

2. Theoretical Contexts

Male scientists, female scientists, and attrition in science

Both young male and female scientists face various barriers to entering and pursuing academic careers (Preston, 2004; Wohrer, 2014); however, attrition has traditionally been examined as a specifically female-dominated phenomenon: women have been reported as facing "chilly" workplace cultures, difficulties in maintaining work–life balance, and hardship surviving motherhood periods while working in academia (Cornelius et al., 1988; Goulden et al., 2011; Levine et al., 2011; Maranto & Griffin, 2011; White-Lewis et al., 2023; Wolfinger et al., 2008). The major dimensions of these barriers are academic promotions, research productivity, scholarly impact, access to research grants, awards, as well as recognition for scholarly achievements: women are under-represented in senior academic positions; have a lower likelihood of collaborating internationally in research, publishing in high-impact journals, and being highly cited; as well as a higher likelihood of longer career breaks and grant rejections (Fochler et al., 2016; Hammarfelt, 2017; Kwiek & Roszka, 2021a, 2021b; Lindahl, 2018; Shibayama & Baba, 2015; Sugimoto & Larivière, 2023; Tang & Horta, 2023). In the majority of STEMM disciplines, women enter predominantly male-dominated environments in which "old boy" networks operate and where a "chilly climate" dominates (Santos et al., 2020).

Although both men and women leave academic science in high proportions—about one-third of all scientists leave science within the first five years and about a half within a decade (as we later show)—attrition is believed to be higher for women than for men (Preston, 2004; Kaminski & Geisler, 2012). The "leaky pipeline" and "chilly climate" hypotheses explain this difference for STEM disciplines: there is a loss of talent at every stage of the academic career pipeline because of systemic barriers for women (Blickenstaff, 2005; Goulden et al., 2011; Shaw & Stanton, 2012; Wolfinger et al., 2008) and a hostile or unwelcoming work environment can discourage women from pursuing their careers (Cornelius et al., 1988; Spoon et al., 2023).

In the "leaky pipeline" model, individuals either progress through a series of academic stages or leave academia altogether. The "chilly climate" for female faculty in STEM disciplines—the "perception of exclusion" and a sense of "not belonging"—is grounded in relational demography (low percentages of women within selected disciplines), but there are also other individual antecedents. These include the percentage of women in a particular department, procedural fairness, and gender equity (Maranto & Griffin, 2011: 143-146). Importantly, from a policy perspective, the concept of the "chilly climate" has informed prolonged significant efforts to promote gender equity on campuses in the US and across Europe (Britton, 2017). As it was succinctly put three decades ago in *Science*: "the culture of science fails to attract women who might otherwise become gifted scientists" (Alper, 1993: 409).

Global datasets and big data approaches to science and scientists

Attrition in science can at long last be quantified beyond individual institutions and countries, and large-scale datasets can be used for this purpose. Global and longitudinal approaches to academic careers have been made possible only recently by increasing access to digital databases with comprehensive information about scientists, their research outputs, and their citation-based impact on global scholarly conversations (Kashyap et al., 2022; Wang & Barabási, 2021). The advent of new digital datasets, access to immense computing power, and a more general turn toward structured big data in social science research have led to a recent explosion of studies about the various aspects of academic careers, with an impressive line of research focused on the differences between men and women in science from various perspectives (e.g., King et al., 2017; Nielsen & Andersen, 2021; Sugimoto & Larivière, 2023).

Large datasets provide the unique capacity to test traditional beliefs and conceptual frameworks about science and scientists (Liu et al., 2023). Digital data that trace the entirety of the scientific enterprise can be used to capture "its inner workings at a remarkable level of detail and scale" (Wang & Barabási, 2021). It is now possible to systematically explore the career histories of hundreds of thousands of individual scientists and the details of their careers. Today, massive datasets are available at researchers' fingertips, although not without new limitations (e.g., Boekhout et al., 2021; Liu et al., 2023; Sugimoto & Larivière, 2023).

Leaving science as a scholarly theme

The theme of leaving science has not been comprehensively examined globally, however. It has traditionally been explored either through small-scale case study research (mostly survey- and interview-based research) or through multi-year US studies of postsecondary faculty (e.g., Rosser, 2004; Xu, 2008; Zhou & Volkwein, 2004). Most recently, White-Lewis et al. (2023) examined actual departure decisions of 2,289 US faculty who left their institutions in 2015–2019. Women were found to leave academia at higher rates overall than men at every career age. Importantly, for women, workplace climate matters more than work–life balance in leaving their academic positions (Spoon et al., 2023).

Leaving science has been previously studied through such concepts as "faculty departure intentions" (Zhou & Volkwein, 2004), "intentions to leave" (Rosser, 2004), "faculty turnover" (Ehrenberg et al., 1991), and "faculty turnover intentions" (Smart, 1990; Xu, 2008). The majority of attrition studies focused on a single institution, and its geographical scope has been limited to the United States (e.g., Minotte & Pedersen, 2021; Levine et al., 2011).

Previous studies have shown that women's stronger turnover intentions are highly correlated with an academic culture that provides women with fewer advancement opportunities and limited research support relative to men and women's roles vis-à-vis family responsibilities (Xu, 2008); however, disciplinary variations in faculty turnover matter, as academics in different disciplines exhibit varying attitudinal and behavioral patterns and there are distinct opportunities offered inside and outside academic environments (Zhou & Volkwein, 2004).

What is important for us in this paper is what Kanter (1977) termed the "proportional scarcity" of women in those disciplines in which women scientists have traditionally occupied token status—where they were either alone or nearly alone in a peer group of men scientists. As we have shown elsewhere in great detail (Kwiek & Szymula, 2023), the disciplines with highly skewed gender ratios are mathematics, physics, and astronomy, computer sciences, and engineering, according to the Scopus classifications (MATH, PHYS, COMP, and ENG). "Tokens" (i.e., minorities among "dominants," with a ratio of about

15:85 – which is generally the current proportion of women in these disciplines) are often treated as representatives of their categories and "as symbols rather than individuals" (Kanter, 1977: 54). All their actions are public; they are visible as category members, and their acts (here under performance pressure to publish) tend to have added symbolic consequences.

Perceptions that academics have of their work lives have been reported to exert a direct impact on their satisfaction and, subsequently, their intention to leave (Rosser, 2004). Academics are pushed to stay/leave their current institutions by a number of internal forces, which can be classified into three major clusters of factors: organizational characteristics, individual characteristics, and work experiences. These factors influence faculty job satisfaction, which in turn influences intentions to leave. They are also pulled by a number of external factors to leave their institutions.

Internal factors include individual and family characteristics (e.g., gender and family/marital status), organizational characteristics (e.g., wealth or unionization), and work experiences (e.g., workload, productivity, and compensation). External factors include the following: external job market, extrinsic rewards, research opportunities, teaching opportunities, and family considerations. Internal factors directly influence job satisfaction and perceptions of the organizational environment, which influence intentions to leave; external factors have also been shown to either strengthen or weaken intentions to leave (Zhou & Volkwein, 2004: 144-147). Smart (1990: 406-409) proposed a causal model to assess the predictors of the intention to leave a current institution for another position in either an academic or nonacademic setting: the three major sets of determinants were individual characteristics (e.g., age and working time distribution), contextual variables (e.g., salary, influence, and career satisfaction), and external conditions (e.g., economic and societal conditions).

The decision to make a career move involves a comparison of the expected present values of the pecuniary and non-pecuniary conditions of employment at the current institution and its alternative (Ehrenberg et al., 1991). From an economic perspective, depending on salary structures, different decisions to stay/leave can be arrived at: in institutions with low salary dispersion, the most productive academics may tend to leave (because they feel undercompensated), whereas in institutions with high salary dispersion, less productive academics may tend to leave (because they feel underpaid relative to their colleagues; Ehrenberg et al., 1991: 107-108). Smart (1990) has shown that being male, spending more time on research, and being more productive positively influence the intentions to leave of tenured faculty, whereas salary satisfaction is an influential variable only for nontenured faculty.

Conceptual frameworks employed to study faculty rationales for leaving higher education include "push" and "pull" factors (i.e., features of the current academic workplace and external features that attract faculty outside their institutions; White-Lewis et al., 2023). However, if push factors are minimal, then pull factors are expected to weigh less in departure decisions (White-Lewis et al., 2023). Explanations of quitting science include problems of work–life balance (Rosser, 2004; Smart, 1990), low job security, and low salaries (O'Meara et al., 2016; Zhou & Volkwein, 2004), colleagues, and workload concerns

(Wohrer, 2014), as well as various types of discrimination in the workplace (Preston, 2004; Smart, 1990) and hostile workplace climates (Cornelius et al., 1988; Spoon et al., 2023).

There are important conceptual differences between “leaving the institution” (for academic leavers) and “leaving academia” (for non-academic leavers), between “intentions to leave” and “actual departure decisions” (White-Lewis et al., 2023), and our conceptualization, in which “leaving science” is examined through the notion of “not publishing” in academic journals any longer. Our focus on publishing over the years until the event in survival analysis of finally “not publishing” occurs at some point goes beyond institutions and sectors to a more general level (publishers vs. non-publishers).

Scholarly publishing events and survival analysis

Scientific life can be conceptualized as a sequence of scholarly publishing events, from the first publication event to subsequent publications and, in many cases, to the last publication ever, when scientists simply cease publishing. In our study, scientists who published for the first time in 2000 form the 2000 cohort, from which, gradually, year by year, both men and women scientists attrit at varying rates of intensity.

Leaving science is conceptualized as an event and analyzed within what is termed *survival analysis* (Allison, 2014; Mills, 2011). Although the general theme of “leaving science” has been widely explored (Geuna & Shibayama, 2015; Preston, 2004), leaving science as an event marked by ceasing publishing has not been extensively studied using survival analysis (and, to the best of our knowledge, has not been studied before from a global quantitative perspective).

In survival analysis, what is explored are questions related to the timing of and the time span leading up to the occurrence of an event (Mills, 2011). An event of interest is the final publication marking the final year of remaining in science; the time span leading up to the event of leaving science is referred to as the *survival time*. A classic statistical technique for survival analysis is the Kaplan–Meier estimate of survival (Mills, 2011). A plot of the Kaplan–Meier estimator is a series of characteristically declining horizontal steps of various heights.

When some of the subjects of a study do not experience the event before the end of the study (as in our case, in which some scientists continued publishing during the study period and did not “leave science” according to our definition of the term), they are termed right-censored observations. For right-censored observations, we have partial information: the event may have occurred sometime after or just before the last year of our study (or will occur), but we do not know exactly in which year. In our case, authors whose last publication was dated 2019 or afterward have been marked as censored observations (i.e., as leaving science in 2020 or some time afterward). To classify

an author as leaving science or staying in science, the final publication must be dated 2018 (marked as leaving science in 2019) or earlier. Uncensored cases represent observations for which we know both the starting year (2000 for the 2000 cohort, 2010 for the 2010 cohort) and the ending year of being in science, which must be 2019 or earlier (determined by the date of the last publication and leaving science the following year; see “Publishing breaks and publishing frequency” in the Online Supplementary Material).

For each year, the initial number of scientists entering the time interval consists of scientists who will leave science during this interval and those who will stay in science into the following interval. The total probability of survival until a given time interval is calculated by multiplying all the probabilities of survival across all time intervals preceding that time (Mills, 2011). The two survival curves—for men and women—can be compared statistically to test whether the difference between survival times for the two groups is statistically significant.

Compared with previous studies, the current research presents a different geographical scale (38 OECD countries combined), thus moving away from single-country institutionally focused case study research designs based on surveys and interviews and toward multi-country research focused on academic disciplines. Moreover, the current study employs a different methodology (i.e., survival analysis and logistic regression analysis) than previous research and examines cross-disciplinary and gender differences in attrition over two decades via a longitudinal approach to large, nonoverlapping cohorts of scientists. We test the power of structured, reliable, and curated big data (of the bibliometric type: the Scopus dataset).

This research is also longitudinal in the strict sense of the term: rather than using a cross-sectional (one snapshot) or repeated cross-sectional (several snapshots) approach, cohorts of scientists are tracked over time for up to two decades (2000–2022) on a yearly basis. Consistent variables and stable classifications are used across national science systems and across time, a wealth of individual micro-level data is leveraged, and time is regarded as a critical variable.

Limitations

The present study is not without limitations. Our research is clearly focused on publishing scientists of any sector, rather than on higher education personnel of all ranks, which is found in traditional accounts of gendered attrition in science (White-Lewis et al., 2023; Zhou & Volkwein, 2004; Deutsch & Yao, 2014; Goulden et al., 2011; Kaminski & Geisler, 2012).

The conceptualization of leaving science as stopping publishing does not entail any other academic roles, such as teaching or administration, or any nonacademic roles, such as work in individual firms, corporations, and governments, even if prior research experience is deemed vital for these career paths. Here, “exit from science is a slippery concept since ‘in science’ and ‘out of science’ are not easily defined terms,” with porous boundaries separating science from nonscience jobs (Preston, 2004).

We discuss the entirety of academic life cycles, from entering to leaving science, through a proxy of publishing the first and last scholarly publications indexed in the Scopus database. As a result, a sequence of publications replaces a sequence of much wider cognitive and social processes encompassing the various dimensions of doing science (Sugimoto &

Larivière, 2023), with the assumption that not publishing in scholarly journals anymore means not doing science any more. Therefore, our representations of scientific careers are necessarily simplified, and our representation of scholarly output is necessarily reduced to globally indexed publications. Collecting individual-level lifetime publication portfolios and calculating lifetime publication-related rates (e.g., international collaboration rate, median team size, and total scholarly output) for every scientist in our dataset is an arduous task; in these portfolios, nonindexed publications and most publications in languages other than English are not counted.

Consequently, in our research, the breadth of scientists' activities in academia (such as mentoring students, refereeing papers, reviewing grant proposals, and editing journals) is ignored (Liu et al., 2023). Generally, higher-level epistemic questions need to be addressed each time the data are repurposed (Kashyap et al., 2022): commercial and governmental datasets were not originally created by researchers for research, and their repurposing needs to be accompanied by a vision of an ideal dataset for examining attrition globally.

In a word, in this research, active participation in science is defined through publishing; consequently, not publishing is defined as leaving science, in accordance with the Mertonian tradition in the sociology of science. Non-publishers in STEMM (see Kwiek 2019) can continue their work in the academic sector in other academic roles; however, it is not possible to verify their intra-sectorial or extra-sectorial employment at the global level using our datasets.

3. Data and Methods

Dataset and dataflow: What we can know about individual scientists from publication metadata?

The present study uses publication and citation bibliometric metadata on researchers starting to publish in the Scopus database for the first time in 2000 and in 2010 (as well as in the years in between). The full population of researchers from the 2000 to 2010 cohorts is N=2,127,803 (Supplementary Table 1, panel 1). Scopus is the largest global abstract and citation database of peer-reviewed literature and it is particularly suitable for global analyses at the micro-level of individual scientists because it is well organized around Scopus Authors IDs, apart from the focus on publications and their metadata (Baas et al., 2020). A list of STEMM disciplines is under Table 1 and a description of ascribing variables to individual scientists is presented under Figure 1).

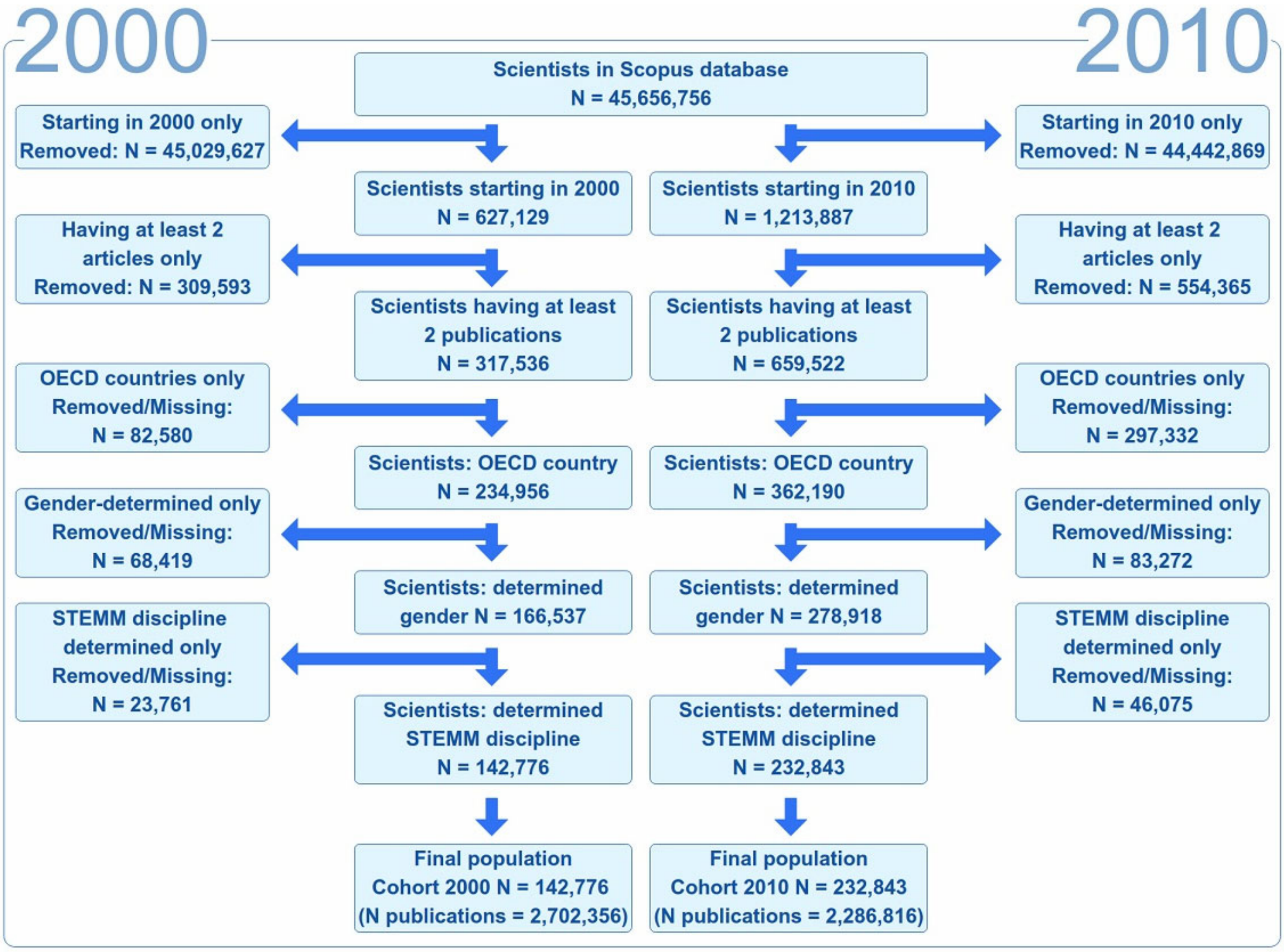


Figure 1: Dataflow: subsequent steps to define the 2000 (left) and the 2010 (right) cohorts of scientists. The dataflow shows two major cohorts: scientists starting to publish in 2000 (left) and scientists starting to publish in 2010 (right). The steps for both cohorts were as follows: to define scientists with at least two publications, to define their country affiliation as an OECD country, to define their gender (binary: male or female), and to define their discipline as STEMM (see the Methods section of the Supplementary Materials).

List of STEMM disciplines: We used all 16 STEMM disciplines, as defined by the journal classification system of 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.

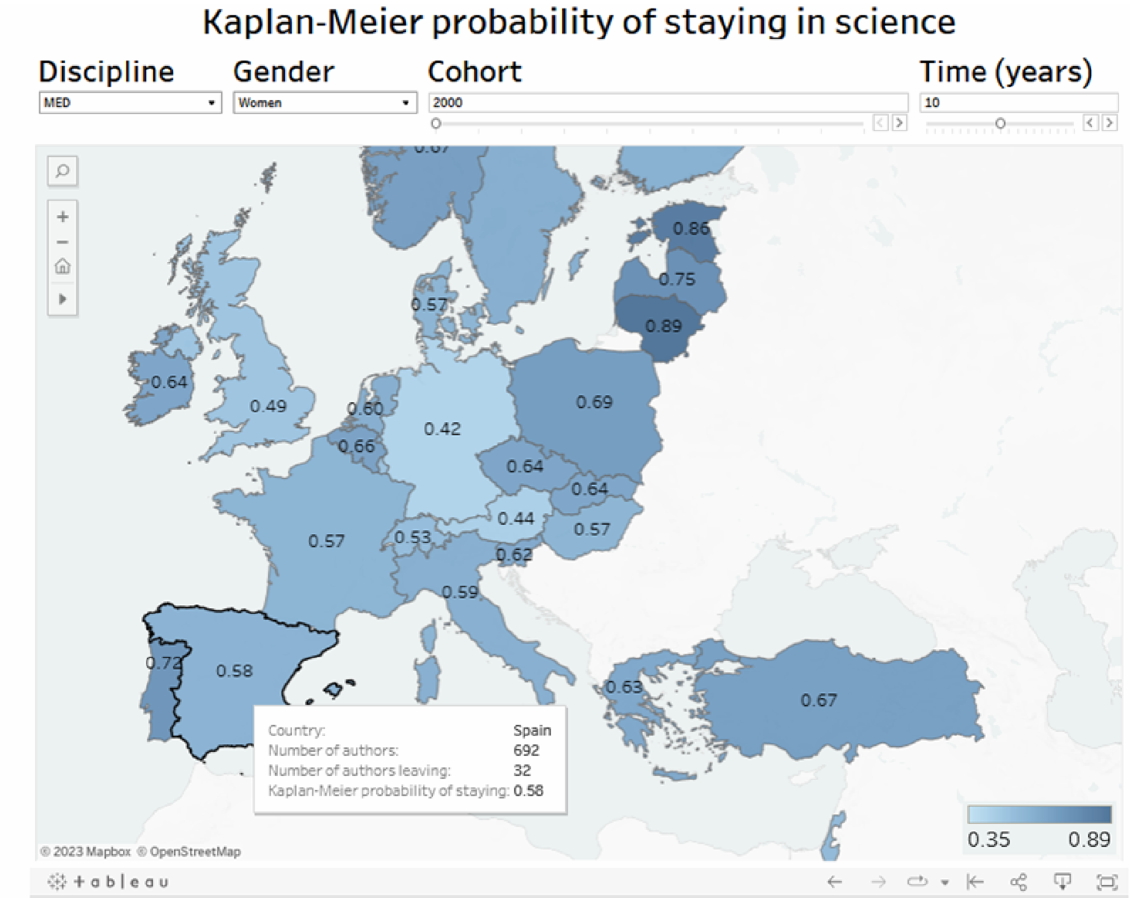
Table 1: Example of the two-cohort database: micro-level data (demographic, institutional, and publishing pattern data) for selected scientists from Cohort 2000 (Panel 1) and Cohort 2010 (Panel 2), N=375,619. For every scientist in the basic population of the study, we have a number of demographic, institutional, and publishing pattern micro-level data that are computed using the data from a global bibliometric dataset. For the 375,619 scientists from the two cohorts, we use the dataset with micro-level data at an individual level.

Scientist ID, the two-cohort database	**Gender**	**Discipline**	**Country affiliation**	**Institutional type**	**Year entering science (year of first publication)**	**Year leaving science (year of last publication plus 1)**	**International collaboration rate, lifetime (%)**	**Average publication journal percentile, lifetime (1-99)**	**Median team size, lifetime**	**FWCI 4y - Field-Weighted Citation Impact, 4 years**	**Scholarly output, lifetime**
Panel 1: Scientists – Cohort 2000 (N=142,776)											
ID 1	Female	MED	Spain	Rest	2000	2020	60.26	31.24	6.5	0.81	78
ID 2	Male	COMP	United States	TOP200	2000	2004	40.00	99.00	4	4.95	10
ID 3	Female	AGRI	France	Rest	2000	2008	21.43	68.15	4	0.88	14
ID 4	Male	PHYS	Japan	TOP200	2000	2013	0.00	90.00	5	1.37	3
ID 5	Female	CHEM	Denmark	Rest	2000	2001	75.00	1.00	3	1.19	4
…											
ID 142776	Male	MED	Germany	Rest	2000	2017	26.67	72.60	3	2.05	30
Panel 2: Scientists – Cohort 2010 (N=232,843)											
ID 142777	Male	ENER	United Kingdom	TOP200	2010	2012	33.33	98.00	5	1.15	6
ID 142778	Female	IMMU	Switzerland	TOP200	2010	2020	27.27	82.10	5	0.78	11
ID 142779	Female	BIO	Belgium	Rest	2010	2017	100.00	29.50	4	0.10	2
ID 142780	Male	ENG	Canada	Rest	2010	2014	14.29	31.43	2.5	2.04	7
ID 142781	Male	MED	Italy	Rest	2010	2012	100.00	14.00	10	0.13	3
…											
ID 375619	Female	AGRI	Australia	TOP200	2010	2015	0.00	91.08	5	1.93	9

Table 2: Variables used in the analysis.

No.	Variable	Description
1.	Gender	Gender (binary: female/male) provided by ICSR Lab. Variable classified based on the first name, last name and dominant country from the first year of publishing using Namsor tool. Gender accepted with probability score >= 0.85 only.
2.	Discipline	Dominant discipline based on the modal value from all disciplines assigned to the journals of all cited references in all papers in scientists' lifetime publication portfolios.
3.	Country affiliation	Dominant country based on the modal value from all countries indicated in scientists' lifetime publication portfolios.
4.	TOP200 institutional affiliation	Binary value indicating belonging (true/false) to one of the 200 top institutions. The list of top institutions was ranked based on the institutions' total scholarly output between 2019 and 2022. Each author has been assigned to one institution as the dominant one based on the modal value from institutions indicated in author's lifetime publication portfolio.
5.	Year of the start of publishing career	First (earliest) publishing year from author's lifetime publication portfolio.
6.	Publication minimum	Binary value indicating having (true/false) at least two publications in author's publication portfolio. Variable delivered from "Scholarly Output" variable value.
7.	Average journal percentile rank (lifetime)	Average of the journals' percentiles assigned to each publication in author's lifetime publishing portfolio. The percentile value has been taken from the 2022 Journal CiteScore metric for discipline with the highest percentile value.
8.	International Collaboration Rate (lifetime)	Share of author's international collaborative publications among all collaborative publications (solo publications excluded). For a publication to be considered collaborative, the number of all authors in the paper had to be greater than or equal to two. For a publication to be considered international, the number of affiliation countries in the paper had to be greater than or equal to two.
9.	Field-weighted four-year citation impact (FWCI 4y)	Average of the FWCI 4y metric values assigned to each publication in author's lifetime publication portfolio. The FWCI 4y metric value of a publication means the ratio of the number of citations of that publication (obtained in the publication year and three consecutive years) to the average number of citations for a similar publication (publication from the same discipline group in 4-digit ASJC discipline classification) in the same time frame.
10.	Median Team Size (lifetime)	Median of the number of authors for each publication (author + number of collaborators) in author's lifetime publication portfolio. For publications with the number of authors greater than 10, the number of authors is 10.
11.	Exit year from publishing (and classifying censored observations)	Next year after last publishing year from author's lifetime publication portfolio. If the exit year occurred after 2019 then author has been classified as right-censored to assure having exit year value.
12.	Scholarly Output (lifetime)	Number of publications (no type specified) in author's lifetime publication portfolio

Our study has a global focus: we analyze the global science profession through the global cohorts of scientists (in 38 OECD countries combined) across time and STEMM disciplines. However, it is also possible, outside of our scope of interest, to compare the differences by country in an interactive format where we provide the results of the Kaplan–Meier probability by country, discipline, and gender for all 11 cohorts (the 2000–2010 cohorts, see Figure 2 and https://public.tableau.com/app/profile/marek.kwiek/viz/Attrition-in-science-OECD/Dashboard).



Kaplan-Meier probability of staying in science

Discipline: MED
Gender: Women
Cohort: 2000
Time (years): 10

Country: Spain
Number of authors: 692
Number of authors leaving: 32
Kaplan-Meier probability of staying: 0.58

Figure 2: A snapshot of an interactive map (available from https://public.tableau.com/app/profile/marek.kwiek/viz/Attrition-in-science-OECD/Dashboard) in which full Kaplan-Meier probabilities are provided by country, discipline, gender and for eleven cohorts 2000-2010 (N=2,127,803, of which 1,289,756 identified as men and 838,047 as women). Here: the probability of staying in science for women from the 2000 cohort of scientists after 10 years for major European OECD member states in MED, reaching 42% in Germany as opposed to as much as 69% in Poland. Highlighted data for Spain.

4. Results

4.1. Descriptive results – Empirical regularities and patterns

Leaving science: the 2000 cohort

First, we tracked 142,776 STEMM scientists from the cohort 2000 until the time they stopped publishing (or until 2022). Figure 3 presents the Kaplan–Meier survival curve for all disciplines combined, with the proportion of scientists staying in science (and probability of staying in science) shown on the Y axis and years spent in publishing since 2000 on the X axis (with small crosses for right-censored cases for the three most recent years of 2020–2022).

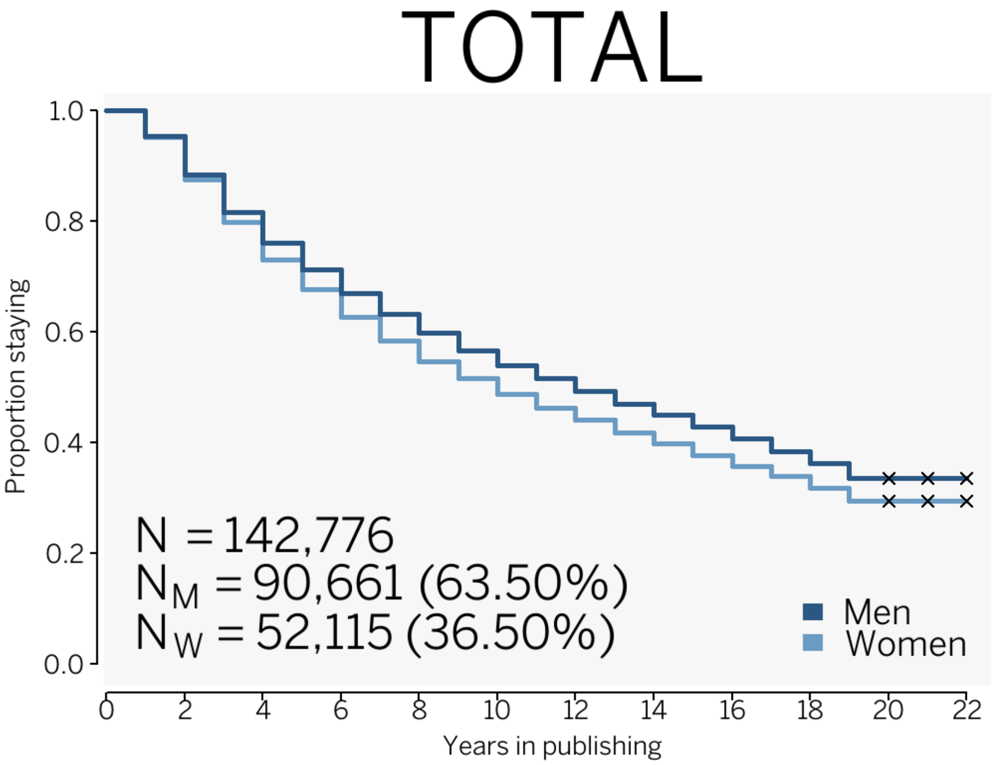


Figure 3: Kaplan–Meier survival curve by gender, all disciplines combined, for the 2000 cohort of scientists (N=142,776). The largest declines in the curve for both men and women appear at years 1, 2, 3, and 4. From year 4 onwards, the difference between men and women becomes visible: every year starting with year 4, the proportion of surviving men is higher than the proportion of surviving women—which generally follows the patterns known from past research.

Simplifying the results, about one-third of the 2000 cohort of scientists leaves science after 5 years, about a half after 10 years, and about two-thirds by the end of the period examined (after 19 years), with the share of the leavers being consistently lower for men and higher for women. From year 4 onwards, for every year examined, higher percentages of women than men leave science. Thus, women are about one-tenth more likely to drop out of science than men after both 5 and 10 years (12.54% and 11.52%, respectively) and are 6.33% more likely to drop out at the end of the studied period. After 19 years (only uncensored observations), the Kaplan–Meier probability of staying for women is 0.294 (29.4% of women from the original cohort continue publishing); for men, in contrast, it is considerably higher and reaches 0.336 (33.6%).

However, this aggregated general picture of attrition in science for all disciplines combined obscures the different disaggregated pictures for particular disciplines, with substantial cross-disciplinary variation (Figure 4).

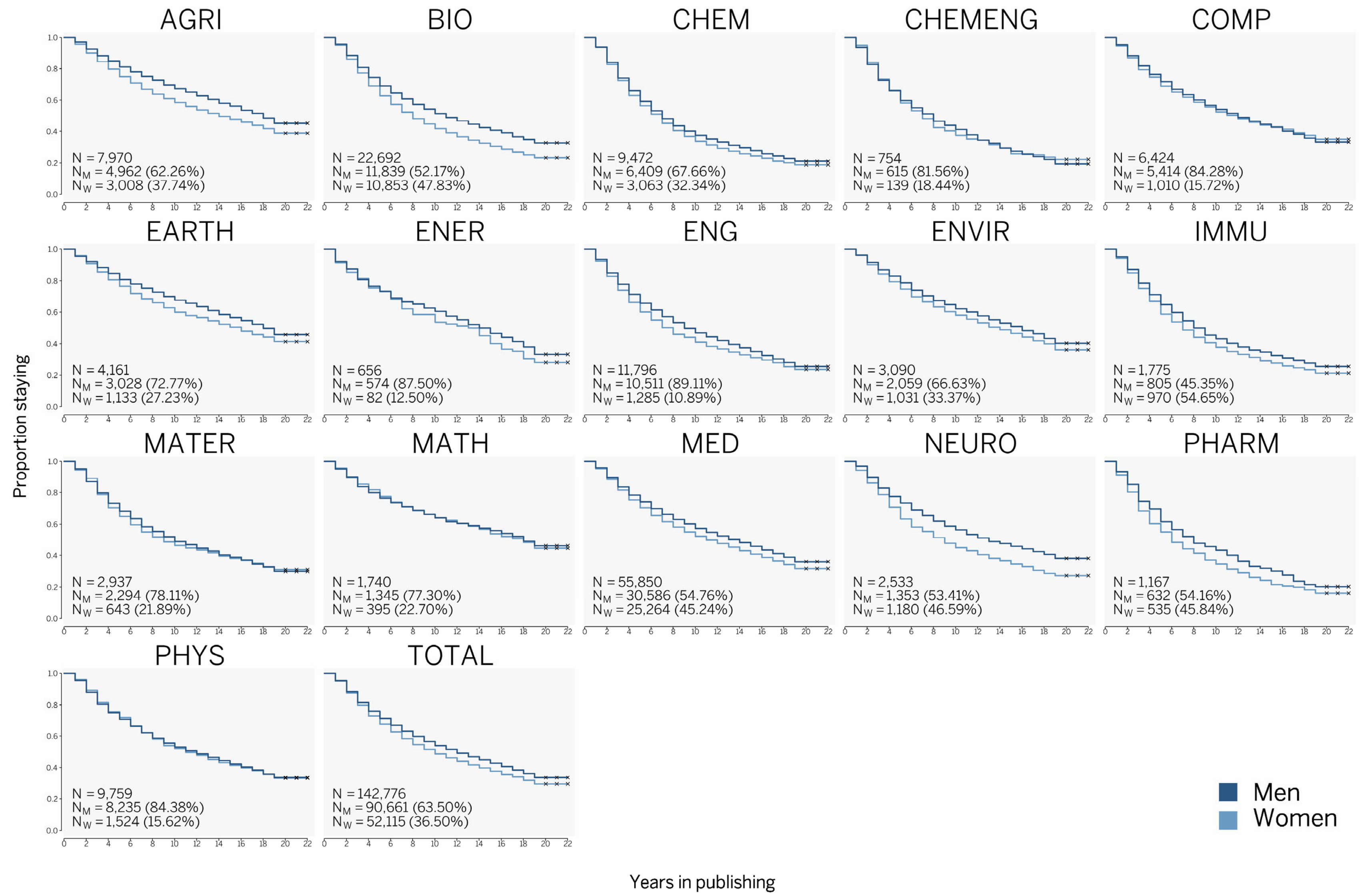


Figure 4: Kaplan–Meier curves by discipline and gender for the 2000 cohort of scientists (N=142,776)

Table 3: Kaplan–Meier estimate for the 2000 cohort population by gender (all disciplines combined) with total counts for men and women, time (in years), number of observations of scientists leaving science, and Kaplan–Meier probability of staying with 95% confidence interval. For women, the probability of leaving science in the first year is 4.9% (1.000 minus 0.951), in the second year 12.5%, in the third year 20.1%, and in the fourth year 26.9%; thus, the cumulative probability after four years is 26.9%. For women, the probability of leaving science after five years is 32.3%, after 10 years is 51.3%, and at the end of the study period (i.e., after 19 years) is 70.6%. In contrast, for men, the corresponding results are 28.7%, 46%, and 66.4%, which is significantly lower for each period studied. The year-by-year data indicate that both the numbers and percentages of men and women leaving science are the biggest in the first 6 to 8 years, and Kaplan–Meier probability of staying from year to year is generally higher in the second decade of publishing than in the first decade of publishing. The Kaplan–Meier probability of staying is lower than 50% for women in year 10 and for men in year 12. The estimated probability that a woman will survive in science 15 years or more is 37.7%, as opposed to 42.9% for men (see Supplementary Tables 3-6 for median publication breaks and median exit time from science).

	Women			**Men**		
Time (years)	**n**	**n leaving science**	**KM probability (staying) with 95% CI and SE**	**n**	**n leaving science**	**KM probability (staying) with 95% CI and SE**
1	52,115	2,530	0.951 (0.950–0.953)[1]	90,661	4,151	0.954 (0.953–0.956)[1]
2	49,585	3,985	0.875 (0.872–0.878)[1]	86,510	6,302	0.885 (0.883–0.887)[1]
3	45,600	3,948	0.799 (0.796–0.803)[2]	80,208	6,114	0.817 (0.815–0.820)[1]
4	41,652	3,553	0.731 (0.727–0.735)[2]	74,094	5,062	0.761 (0.759–0.764)[1]
5	38,099	2,838	0.677 (0.673–0.681)[2]	69,032	4,356	0.713 (0.710–0.716)[2]
6	35,261	2,602	0.627 (0.623–0.631)[2]	64,676	3,934	0.670 (0.667–0.673)[2]
7	32,659	2,183	0.585 (0.581–0.589)[2]	60,742	3,458	0.632 (0.629–0.635)[2]
8	30,476	1,961	0.547 (0.543–0.551)[2]	57,284	3,110	0.598 (0.594–0.601)[2]
9	28,515	1,665	0.515 (0.511–0.520)[2]	54,174	2,774	0.567 (0.564–0.570)[2]
10	26,850	1,472	0.487 (0.483–0.491)[2]	51,400	2,465	0.540 (0.537–0.543)[2]
11	25,378	1,264	0.463 (0.458–0.467)[2]	48,935	2,225	0.515 (0.512–0.518)[2]
12	24,114	1,158	0.440 (0.436–0.445)[2]	46,710	2,055	0.493 (0.489–0.496)[2]
13	22,956	1,151	0.418 (0.414–0.423)[2]	44,655	2,032	0.470 (0.467–0.473)[2]
14	21,805	1,089	0.398 (0.393–0.402)[2]	42,623	1,889	0.449 (0.446–0.453)[2]
15	20,716	1,048	0.377 (0.373–0.382)[2]	40,734	1,884	0.429 (0.425–0.432)[2]
16	19,668	1,033	0.358 (0.353–0.362)[2]	38,850	1,959	0.407 (0.404–0.410)[2]
17	18,635	1,002	0.338 (0.334–0.342)[2]	36,891	2,020	0.385 (0.381–0.388)[2]
18	17,633	1,064	0.318 (0.314–0.322)[2]	34,871	2,070	0.362 (0.359–0.365)[2]
19	16,569	1,228	0.294 (0.290–0.298)[2]	32,801	2,350	0.336 (0.333–0.339)[2]

Note: (1) Standard Error 0.001, (2) Standard Error 0.002.

Our special interest in scientific careers is in disciplines with the biggest and smallest (or none at all) differences in survival curves between men and women. In the two largest disciplines and two disciplines with the largest number of women, medicine (MED) and biochemistry, genetics, and molecular biology (BIO), the curves for men and women are clearly divergent.

Focusing on BIO (Figure 5, left panel, contrasted with physics and astronomy (PHYS), right panel), with 47.83% of women and 22,692 scientists in the cohort examined, the largest declines are in years 2, 3, and 4; and from year 3, there is an ever-increasing men–women difference between the two survival curves, here showing smoother declines.

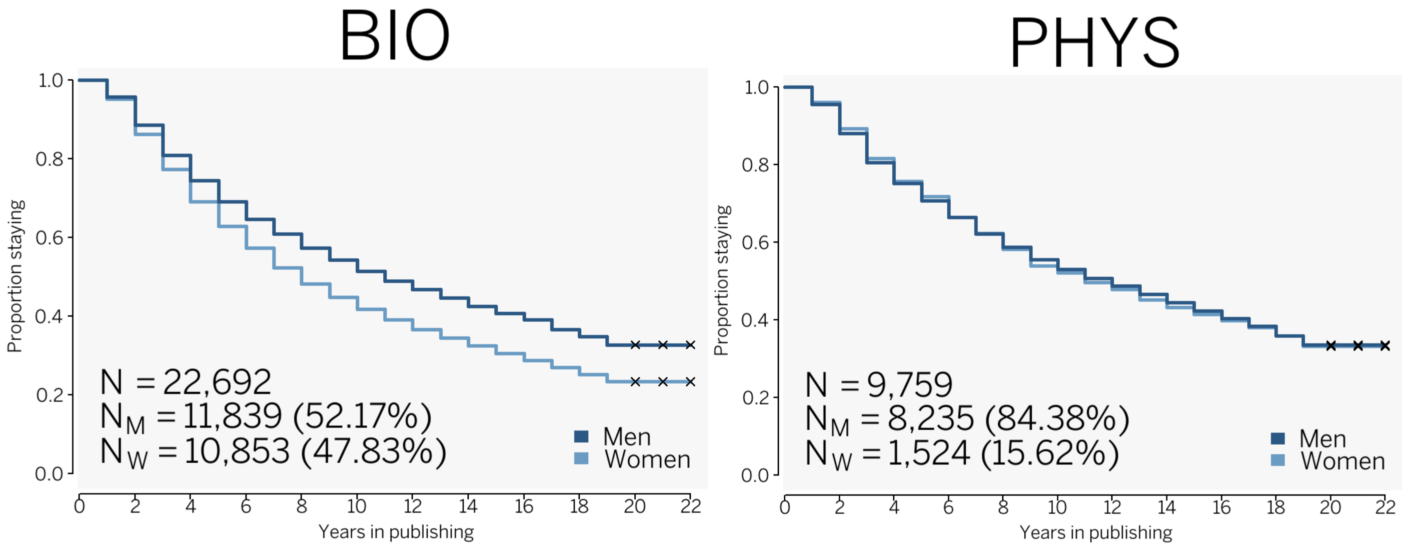


Figure 5: Kaplan–Meier survival curve by gender, BIO (N=22,692, left panel) vs. PHYS (N=9,759, right panel), the 2000 cohort of scientists. For women in BIO, the probability of leaving science after 5 years is 37.2%, after 10 years is 58.3%, and at the end of the study period (i.e., after 19 years) is 76.6%. For men, the corresponding figures are strikingly lower: they are 30.8%, 48.6%, and 67.3%, respectively. Thus, women are as much as one-fifth more likely to drop out of science than men after both 5 and 10 years (20.78% and 19.96%, respectively), and 13.82% more likely to drop out at the end of the studied period. PHYS is a perfect example of the lack of gender differences in attrition. For women in PHYS, the probability of leaving science after 5 years is 28.1% (and 29.2% for men), after 10 years is 47.9% (46.9% for men), and at the end of the study period (i.e., after 19 years) is 66.9% (66.5% for men).

Strikingly, in the three math-intensive disciplines, MATH, COMP, and PHYS, which have very low numbers and percentages of women, the survival curves for men and women are nearly identical (see PHYS in Figure 5, right panel), with the two curves almost overlapping. As confirmed for 38 OECD countries by a large sample of physicists and astronomers all starting publishing in 2000, gender differences in attrition in PHYS do not exist. Overall, for half of the disciplines, the differences are statistically significant, and for the other half, they are not (see Supplementary Table 7).

To have a more comprehensive picture of attrition and retention in science, we have correlated the data on men and women to produce survival regression curves (panels A in the following Figure 6), hazard rate curves (panels B), and kernel density curves (panels C) to tell the story of attrition and retention from different angles.

The survival regression curves for all disciplines combined (Total, Figure 6, panel A) indicate a much steeper decline for men and women in the early years of publishing career and a much smoother decline in later years of publishing career, with the increasing divergence between the curves for men and women. The surviving men stay in science at a higher rate than the surviving women, and both men and women later in their careers stay in science at a higher rate than earlier in their careers (the regression function shows a higher decline at the beginning of the career than in its later years). The hazard rate curves (Figure 6, Panel B) provide a similar picture: the rate of attrition for both men and women is higher in the early publishing years, and it is considerably lower in later years; and for the first 15 years, women have higher chances of leaving science than men. The peak for men is in year 4 and for women year 6.

Finally, a different story is told by the kernel density curves (Figure 6, panel C), which show how all men and women who actually left science from cohort 2000 are distributed over time. The shares of scientists who left science are the highest for the first 8 years, with years 3–4 being critical. Both men and women are more likely to leave science early on after starting publishing careers than later on. Once they have survived the first 10 years, their likelihood of leaving is much smaller.

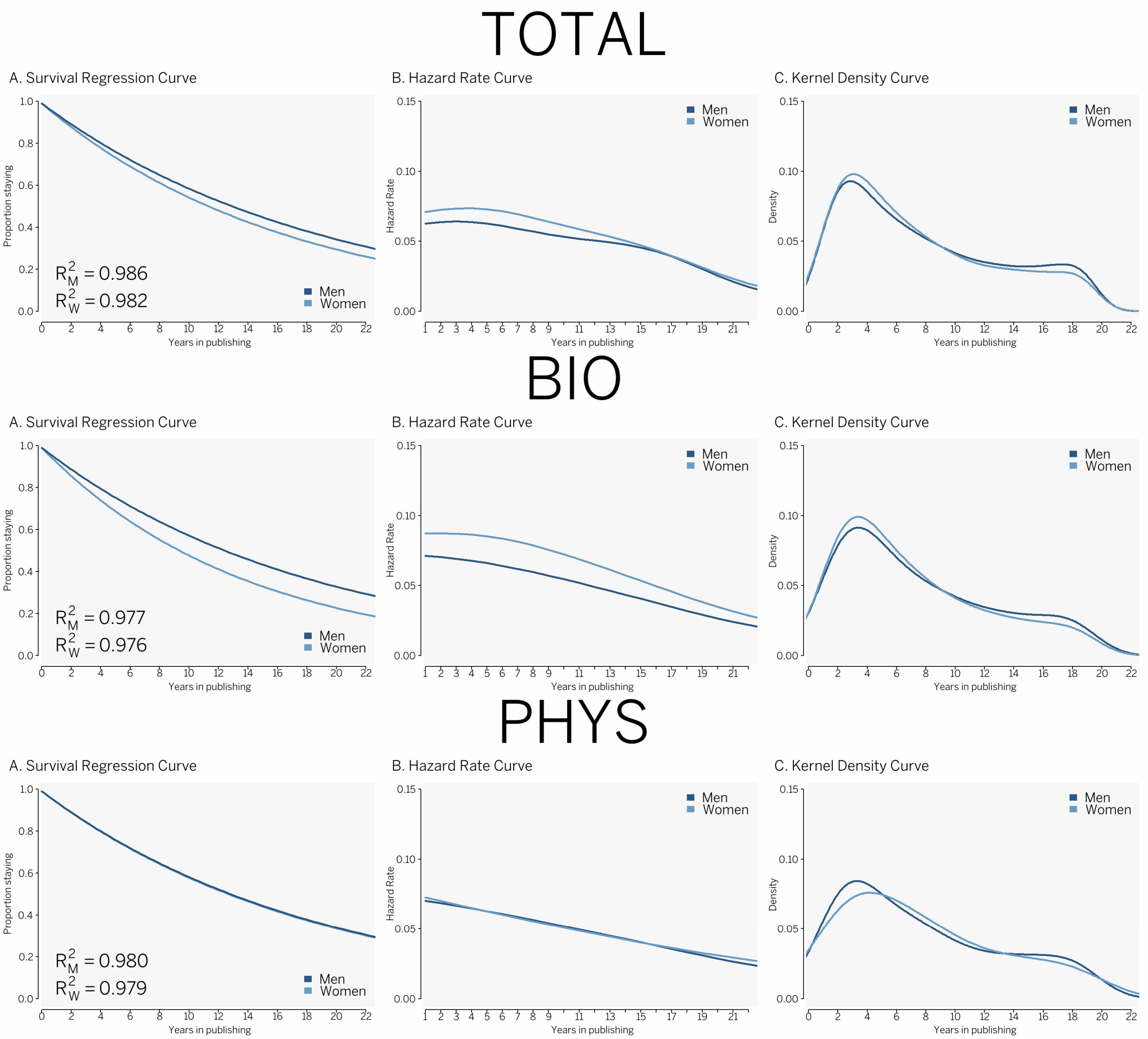


Figure 6: Survival regression curve, hazard rate curve, and kernel density curve, all disciplines combined (Total, N=142,776)), BIO (N=22,692), and PHYS (N=9,750), the 2000 cohort

Again, this generalized story for all disciplines combined obscures a number of much more nuanced discipline-specific stories. The Kaplan–Meier curves for BIO and PHYS analyzed above tell fundamentally different stories. These findings are confirmed when the survival regression curves are compared (BIO and PHYS, Figure 6, Panel A), with similar higher probabilities to leave in early years and lower probabilities in later years—but with stark differences between men and women in the two disciplines. Our data show that women in BIO disappear from science with the passage of time in ever-larger proportions compared with men; in contrast, women in PHYS disappear from science in almost exactly the same proportions as men in the whole period examined.

Also, hazard rate curves (Figure 6, Panels B) tell a similar story in which, in BIO, the attrition rate for women is substantially higher than the attrition rate for men across all the years examined; in PHYS, in contrast, the hazard curve rates for men and women are almost identical.

Kernel density curves for BIO and PHYS (Figure 6, Panel C) clearly show similar intradisciplinary patterns for men and women (higher attrition in early years) and different cross-disciplinary patterns for men and women. Actually, while in BIO in the early years, attrition for women is higher than for men, and in PHYS in years 2 through 6, it is higher for men. This is a surprising finding, which is confirmed also for COMP and MATH.

Leaving science: the 2010 cohort

Finally, we move on to the 2010 cohort (Figure 7). The Kaplan–Meier survival curves for men and women are drastically different for the two cohorts. Most importantly, the Kaplan–Meier curves for all disciplines combined for the 2010 cohort are nearly identical for men and women (see Supplementary Table 12). They are also nearly identical for math-intensive COMP, PHYS, and ENG (Figure 8).

After 9 years from entering science, the probability of staying for women is 0.414 (41.4% of women from the original cohort continue publishing); for men, the probability of staying is only slightly higher at 0.424 (42.4%), which is a dramatic lack of difference compared with the results for the 2000 cohort, where the results were substantially gender sensitive. In exactly the same eight disciplines, including the big four of math-intensive COMP, ENG, MATH, and PHYS, the statistical tests show that there are no statistically significant differences in the survival curves for men and women; however, for all disciplines combined, as well as for the largest disciplines of MED, BIO, and AGRI, with women representing about 50%, the differences are statistically significant (see Supplementary Table 13).

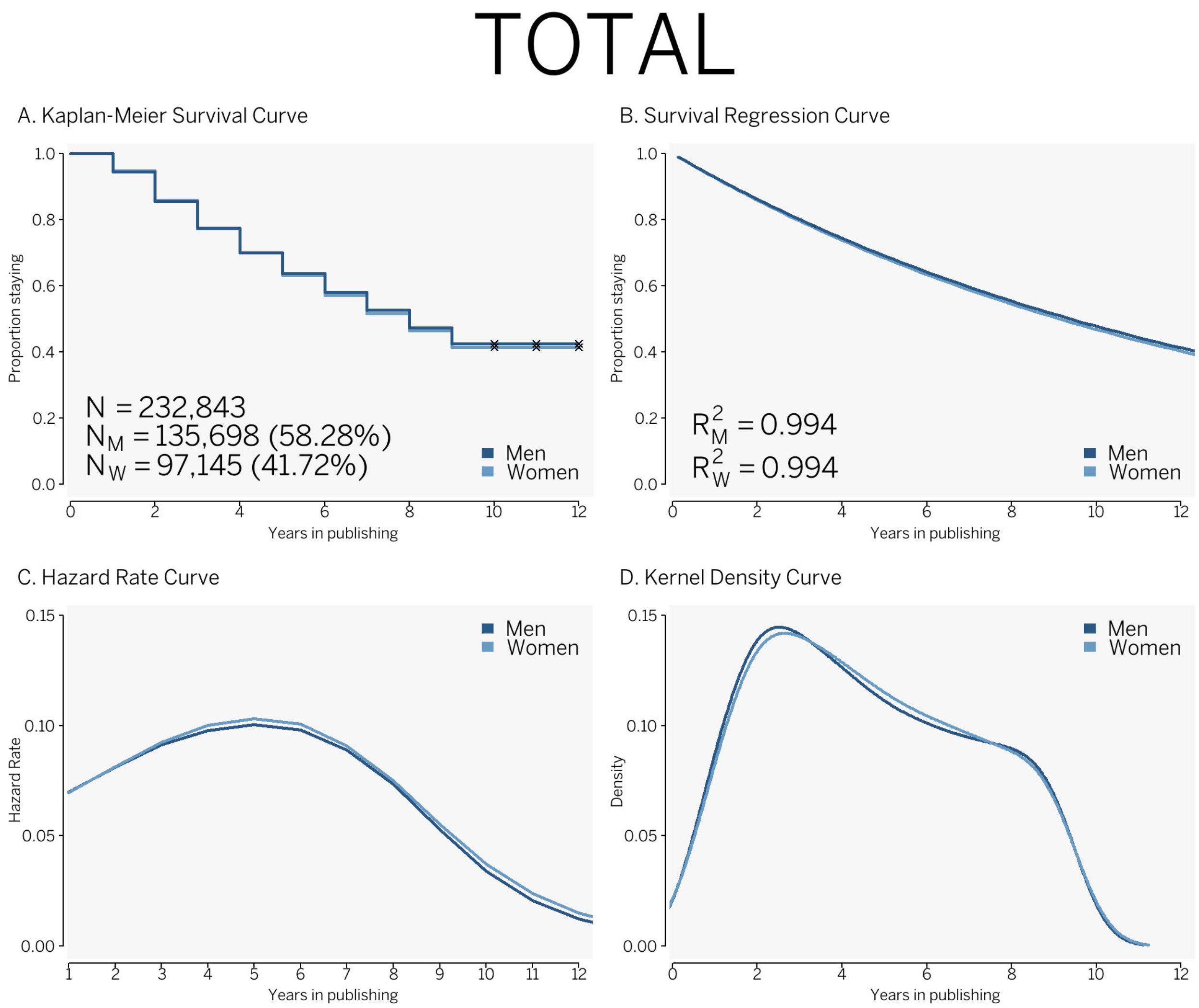


Figure 7: Kaplan–Meier curve, survival regre0ssion curve, hazard rate curve, and kernel density curve, the 2010 cohort (N=232,843)

From a cross-cohort comparative perspective, our analyses indicate a fundamental transformation in attrition patterns across disciplines (and in all disciplines combined). The differences between men and women, which are so starkly visible for the 2000 cohort, almost disappear for the 2010 cohort.

This is a finding with important potential research and policy implications in view of the vast previous literature on attrition and retention in academic science. The science sector globally seems to be undergoing powerful transformations, and the findings valid for older cohorts of scientists (here: the 2000 cohort) may not be applicable to younger cohorts (here: the 2010 cohort). Time in science—in this case, the difference of about a decade—matters.

The Kaplan–Meier survival curves for all disciplines combined take vastly different forms for the two cohorts: for the 2010 cohort (Figure 7), first, attrition rates are much higher and declines much more dramatic (50% or more of both men and women disappear in year 8), and there are no gender differences at all. This finding is confirmed by the shapes of the survival regression curves for the two cohorts, again with no gender difference; the confirmation comes from the hazard rate curves, which are nearly identical for men and women, and from the kernel density curve, which testifies to the generally similar distribution of scientists who left science over time. Specifically, the kernel density distribution shows dramatically high attrition rates for both men and women in the first 4 years of staying in science (Figure 7, panels B, C, and D).

Again, as for the 2000 cohort, a big picture of all disciplines combined hides behind smaller and complicated discipline-specific pictures. An analysis of disciplinary variations in the three survival perspectives for the two cohorts is followed by detailed figures: Supplementary Figures 7 through 12).

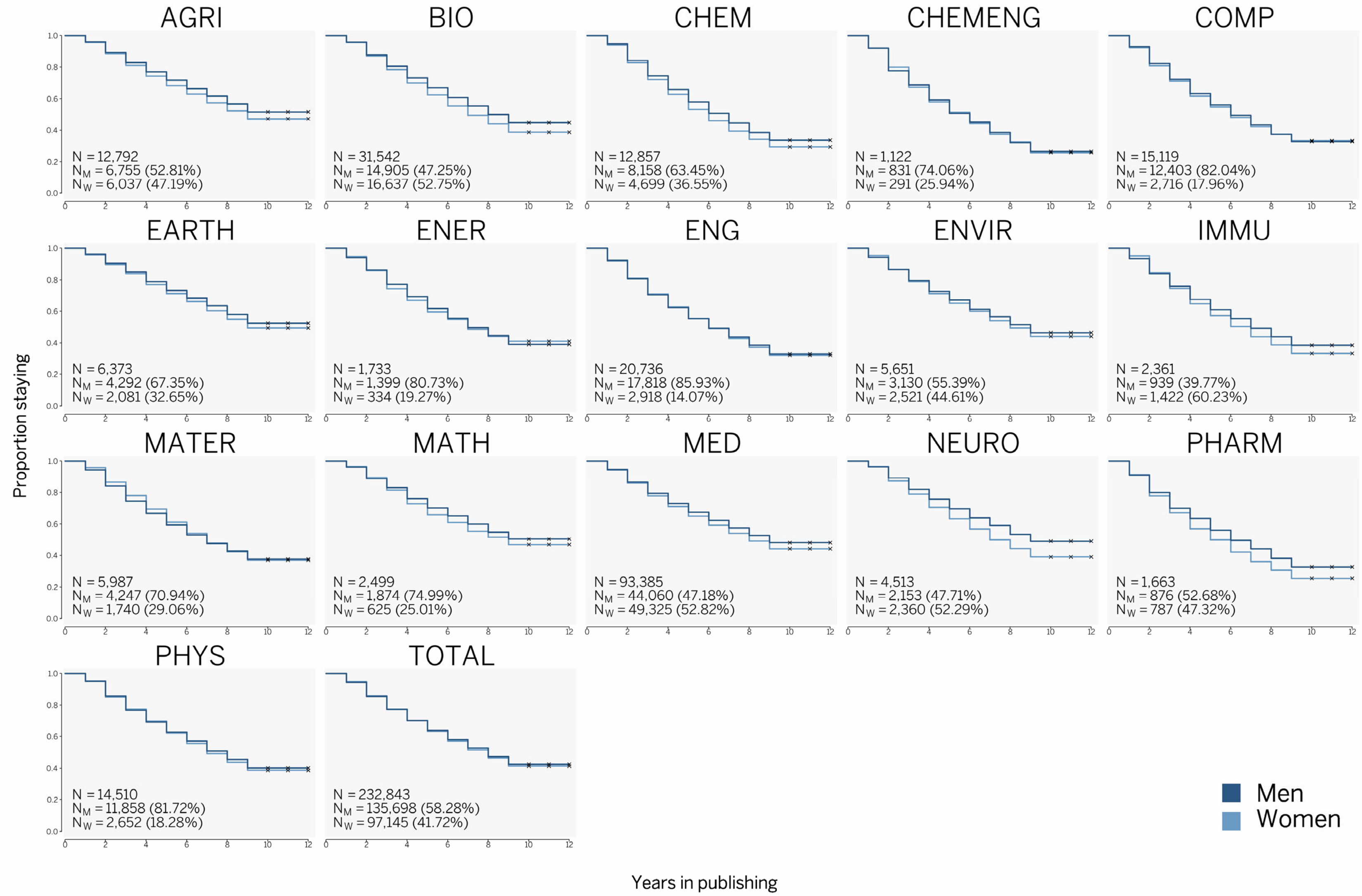


Figure 8: Kaplan–Meier curve by discipline and gender in the 2010 cohort (N=232,843)

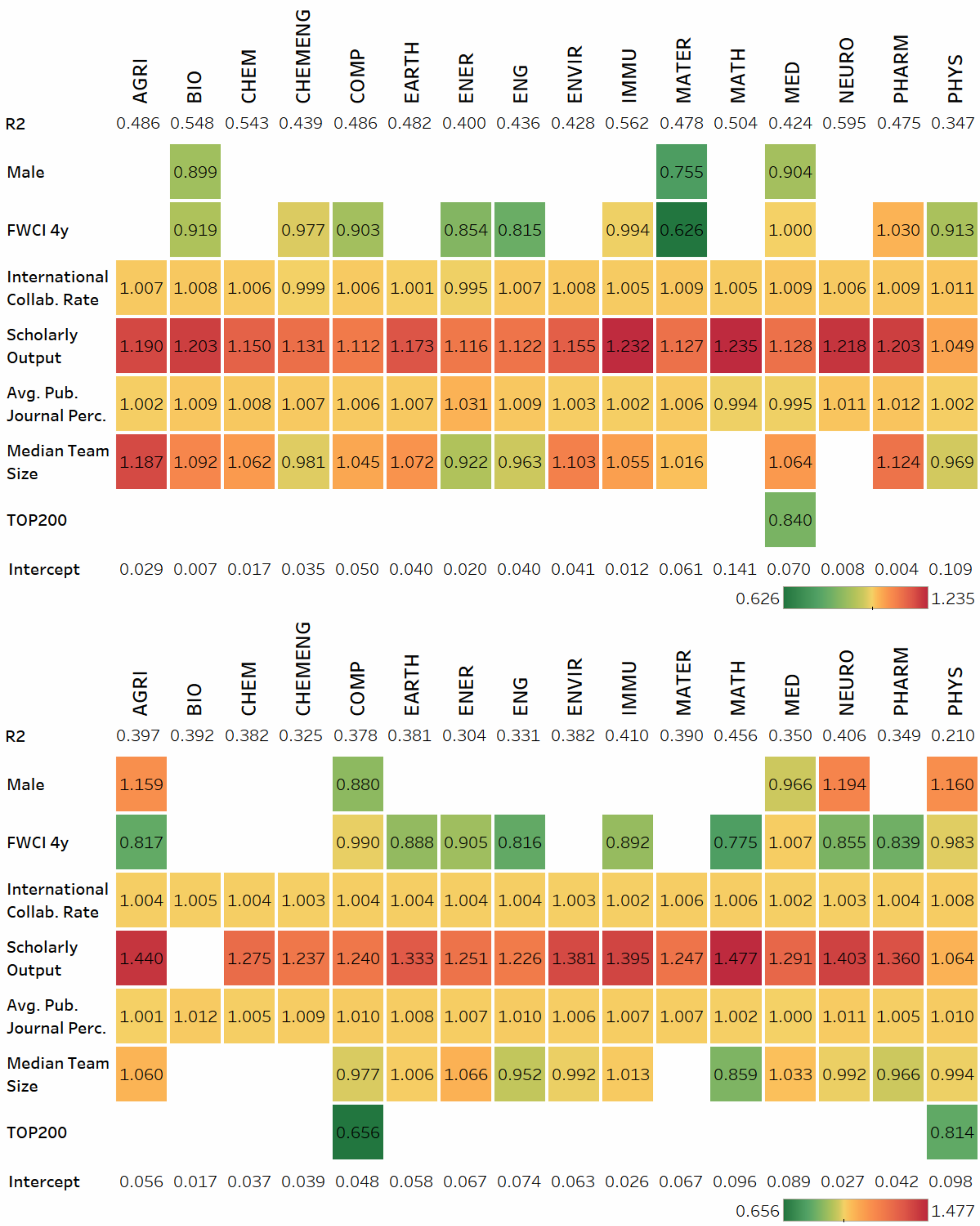
4.2. Logistic regression analysis

Finally, we use logistic regression models to estimate the odds ratios of staying in science for the various cohorts of scientists within different disciplines from a multidimensional perspective. Specifically, we compare the regression results for the two cohorts separated in time by a decade (the 2000 cohort vs. the 2010 cohort) and, thus, scientists starting publishing in different periods and under different conditions. We have also constructed models for all cohorts of scientists between 2000 and 2010 (N=2,127,803 scientists). The success in our models is understood as entering about 30% of scientists who stayed in science after 19 years in the case of the 2000 cohort (women 29.5%, CI 29.0-29.8%; men 33.6%, CI 33.3-33.9%, Table 3) and entering about 40% of scientists who stayed in science after 9 years in the case of the 2010 cohort (women 41.4%, CI 41.1-41.7%; men 42.4%, CI 42.2-42.7%, Supplementary Table 12).

The literature on attrition in science (e.g., Preston, 2004; Deutsch & Yao, 2014; Kaminski & Geisler, 2012) and on academic careers more generally (e.g., Fochler et al., 2016; Hamermesh & Pfann, 2011) suggests that a career may prove to be relatively predictable and based on similar milestones, with increasingly similar, mostly research-related, requirements, regardless of the cohort examined. We hypothesized that the factors traditionally associated with successful academic careers, such as high research productivity (Stephan, 2012), publishing in high-impact journals (Shibayama & Baba, 2015; Heckman & Moktan, 2018), and gender (Sugimoto & Larivière, 2023), should play important roles for all cohorts examined.

Based on previous research, our major interest is in the role of publication quantity and quality (publication numbers vs. their relative location in a highly stratified system of academic journals) and the role of gender in retention in science. We hypothesize that the combination of larger publication numbers and more prestigious academic venues should be important predictors increasing the chances of survival in global science and that gender, in the multidimensional context of all predictors combined, should play a marginal role.

What do the regression results for the two cohorts (Figure 9) suggest? There is a powerful message: publication number (variable: Scholarly Output) is the single most influential predictor, increasing the odds of staying in science for both cohorts. For the 2000 cohort, the lifetime (cumulative) number of publications of all types (not only research articles) is statistically significant for every discipline. An increase in the total number of publications by one increases the odds of continuing publishing in the future i.e. of staying in science on average (all other things being equal) by about 20% in AGRI, BIO, MATH, NEURO, and PHARM and by about 15% in CHEM, EARTH, and ENVIR. This predictor is the most powerful in MATH (23.5%) and the least powerful in PHYS (4.9%), which is in line with general cross-disciplinary differences in publishing patterns in which mathematicians tend to publish less and in much smaller teams (and often solo) compared with physicists and astronomers, who tend to publish more and in much larger teams (Mihaljević & Santamaria, 2020); consequently, an increase by one publication in lifetime scholarly output seems to be more consequential in terms of attrition/retention in science for the former than for the latter.



	AGRI	BIO	CHEM	CHEMENG	COMP	EARTH	ENER	ENG	ENVIR	IMMU	MATER	MATH	MED	NEURO	PHARM	PHYS
R2	0.486	0.548	0.543	0.439	0.486	0.482	0.400	0.436	0.428	0.562	0.478	0.504	0.424	0.595	0.475	0.347
Male		0.899									0.755		0.904			
FWCI 4y		0.919		0.977	0.903		0.854	0.815		0.994	0.626		1.000		1.030	0.913
International Collab. Rate	1.007	1.008	1.006	0.999	1.006	1.001	0.995	1.007	1.008	1.005	1.009	1.005	1.009	1.006	1.009	1.011
Scholarly Output	1.190	1.203	1.150	1.131	1.112	1.173	1.116	1.122	1.155	1.232	1.127	1.235	1.128	1.218	1.203	1.049
Avg. Pub. Journal Perc.	1.002	1.009	1.008	1.007	1.006	1.007	1.031	1.009	1.003	1.002	1.006	0.994	0.995	1.011	1.012	1.002
Median Team Size	1.187	1.092	1.062	0.981	1.045	1.072	0.922	0.963	1.103	1.055	1.016		1.064		1.124	0.969
TOP200													0.840			
Intercept	0.029	0.007	0.017	0.035	0.050	0.040	0.020	0.040	0.041	0.012	0.061	0.141	0.070	0.008	0.004	0.109

	AGRI	BIO	CHEM	CHEMENG	COMP	EARTH	ENER	ENG	ENVIR	IMMU	MATER	MATH	MED	NEURO	PHARM	PHYS
R2	0.397	0.392	0.382	0.325	0.378	0.381	0.304	0.331	0.382	0.410	0.390	0.456	0.350	0.406	0.349	0.210
Male	1.159				0.880								0.966	1.194		1.160
FWCI 4y	0.817				0.990	0.888	0.905	0.816		0.892		0.775	1.007	0.855	0.839	0.983
International Collab. Rate	1.004	1.005	1.004	1.003	1.004	1.004	1.004	1.004	1.003	1.002	1.006	1.006	1.002	1.003	1.004	1.008
Scholarly Output	1.440		1.275	1.237	1.240	1.333	1.251	1.226	1.381	1.395	1.247	1.477	1.291	1.403	1.360	1.064
Avg. Pub. Journal Perc.	1.001	1.012	1.005	1.009	1.010	1.008	1.007	1.010	1.006	1.007	1.007	1.002	1.000	1.011	1.005	1.010
Median Team Size	1.060				0.977	1.006	1.066	0.952	0.992	1.013		0.859	1.033	0.992	0.966	0.994
TOP200					0.656											0.814
Intercept	0.056	0.017	0.037	0.039	0.048	0.058	0.067	0.074	0.063	0.026	0.067	0.096	0.089	0.027	0.042	0.098

Figure 9: Logistic regression model and odds ratio estimates for staying in science. Upper panel: cohort 2000 (N=142,776), publication period 2000–2019 (20 years). Bottom panel: cohort 2010 (N=232,843), publication period 2010–2019 (10 years). Only statistically significant results (p-value

below 0.05). The models were run on individual-level microdata separately for 16 STEMM disciplines. The individual variables included gender (binary), four variables related to individual lifetime publishing patterns (International Collaboration Rate; Scholarly Output; Average Publication Journal Percentile; and Median Team Size), and the variable related to the individual impact on global science through citations—here in the form of individual field-weighted citation impact averaged for every research output in the 4-year period following its publication (FWCI 4 y). The institutional variable (Top200) is the affiliation with one of the 200 institutions with the highest number of publications (of any type) over a 5-year period (2018–2022). We have computed the variables separately for each individual (see the Methods section). See also inverse correlation matrices and main diagonals in Supplementary Figure 18 (the 2000 cohort) and Supplementary Figure 19 the 2010 cohort).

Interestingly, for the 2010 cohort, the role of publication numbers is much bigger than for the 2000 cohort, which is not surprising in the context of increasing competition for research funding, academic prestige, and tenure-track and tenured academic positions in universities. Scientists from the 2010 cohort started scholarly publishing under different conditions, which include more competitive access to research funding, channeling young scientists more toward externally funded research projects than permanent employment, and increasing the numbers of postdocs (e.g., Fochler et al., 2016; Stephan, 2012).

Our models show that, for the survival of scientists from the 2010 cohort, publication numbers are much more important than for scientists from the earlier cohort: an increase in lifetime publication number by one increases the odds of staying in science by about 40% in several disciplines: AGRI, ENVIR, IMMU, and NEURO. One additional publication increases the chances of staying in science by about one-third in EARTH, MED, PHARM. In the regression models, we examine the role of both quantity and quality of publications in the presence of all other variables.

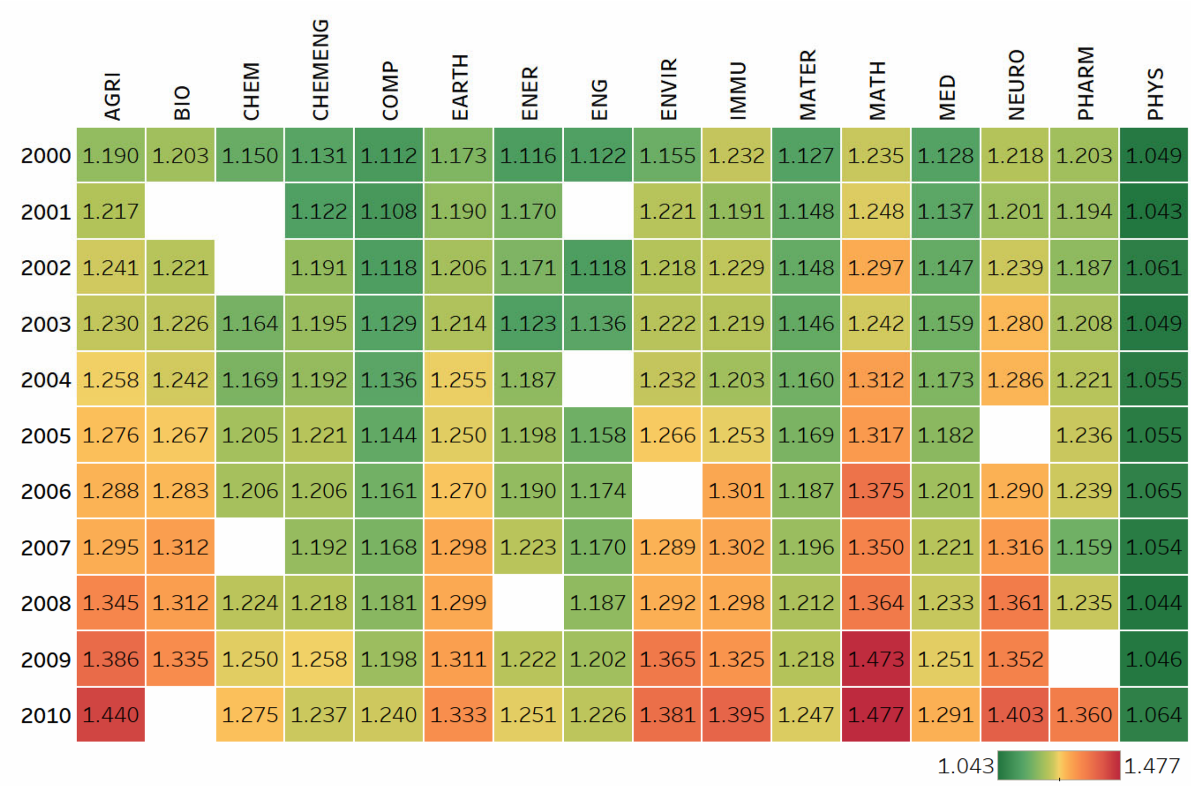
The lesson from our models which combine quantity and quality for the 2000 and 2010 cohorts, is that the quantity of publications is substantially more important as a predictor, increasing the odds of staying in science than the quality of publications (all other things being equal), with significant cross-disciplinary variation. Survival in science seems to require high publication productivity.

In contrast, increasing the lifetime field-weighted citation impact within 4 years after each publication by one unit (e.g. FWCI 4 y from 0.9 to 1.9) is either statistically insignificant, may actually decrease the odds, or may slightly increase the odds. Citations collected from individual publications do not serve as an influential predictor of staying in science (see the Methods section in Supplementary Material on how variables have been constructed).

Also, the other predictor based on journal quality we constructed (variable: Average Publication Journal Percentile) functions as a less important variable in the model than expected. Although for both cohorts the predictor is statistically significant for all disciplines, its power is smaller than expected, here based on the literature about the role of top journals in academic careers (Hamermesh & Pfann, 2011; Heckman & Moktan, 2018). In our models, for the 2000 cohort, an increase by one journal percentile rank lifetime on average increases the odds of staying in science in most disciplines by less than 1%. In other words, lifetime publishing in journals ranked 30 percentiles higher in the Scopus database, on average, increases the odds in the majority of disciplines by about 30% or less. A similar power and direction of influence is noted for this predictor for the 2010 cohort. The direction of influence of this variable follows general expectations from literature—more publications in top journals have a positive impact on persistence of academic careers—but its power is much smaller than expected from literature (e.g., Fochler et al., 2016; Hammarfelt, 2017; Lindahl, 2018; Shibayama & Baba, 2015). Thus, our model shows that, for survival in science, quality needs to be combined with quantity. These results demonstrate the crucial role of skillful navigation in the globally stratified system of journals for career advancement in science, in which both numbers and journal prestige matter.

Finally, the role of gender as a predictor of staying in science for both cohorts is smaller than expected from the literature: in a multidimensional regression analysis, in the presence of other variables, being male is not as important a predictor of staying in science in the vast majority of disciplines as might be expected. The general model of staying in science we have constructed using mostly micro-level variables based on bibliometric data fits the phenomenon under exploration relatively well (see Supplementary Figure 14).

A summary of the role of lifetime scholarly output for the 11 cohorts (the 2000 through 2010) sends a clear message (Figure 10): in every discipline (except PHYS physics and astronomy), the impact of one additional publication has been steadily increasing for each successive cohort of scientists. For instance, in medicine (MED), one additional publication increases the chances of staying in science by 12.8% for the 2000 cohort, 18.2% for the 2005 cohort, and by 29.1% for the 2010 cohort.



	AGRI	BIO	CHEM	CHEMENG	COMP	EARTH	ENER	ENG	ENVIR	IMMU	MATER	MATH	MED	NEURO	PHARM	PHYS
2000	1.190	1.203	1.150	1.131	1.112	1.173	1.116	1.122	1.155	1.232	1.127	1.235	1.128	1.218	1.203	1.049
2001	1.217			1.122	1.108	1.190	1.170		1.221	1.191	1.148	1.248	1.137	1.201	1.194	1.043
2002	1.241	1.221		1.191	1.118	1.206	1.171	1.118	1.218	1.229	1.148	1.297	1.147	1.239	1.187	1.061
2003	1.230	1.226	1.164	1.195	1.129	1.214	1.123	1.136	1.222	1.219	1.146	1.242	1.159	1.280	1.208	1.049
2004	1.258	1.242	1.169	1.192	1.136	1.255	1.187		1.232	1.203	1.160	1.312	1.173	1.286	1.221	1.055
2005	1.276	1.267	1.205	1.221	1.144	1.250	1.198	1.158	1.266	1.253	1.169	1.317	1.182		1.236	1.055
2006	1.288	1.283	1.206	1.206	1.161	1.270	1.190	1.174		1.301	1.187	1.375	1.201	1.290	1.239	1.065
2007	1.295	1.312		1.192	1.168	1.298	1.223	1.170	1.289	1.302	1.196	1.350	1.221	1.316	1.159	1.054
2008	1.345	1.312	1.224	1.218	1.181	1.299		1.187	1.292	1.298	1.212	1.364	1.233	1.361	1.235	1.044
2009	1.386	1.335	1.250	1.258	1.198	1.311	1.222	1.202	1.365	1.325	1.218	1.473	1.251	1.352		1.046
2010	1.440		1.275	1.237	1.240	1.333	1.251	1.226	1.381	1.395	1.247	1.477	1.291	1.403	1.360	1.064

Figure 10: Comparison of scholarly output (lifetime) variable for 11 cohorts (2000 to 2010) for the model of odds ratio estimates of staying in science from the cohort starting year up to and including 2019, by discipline (only cohorts and disciplines with statistically significant results, p-value<0.05)

5. Discussion and Conclusions

Attrition and retention in science are long-term processes and require comprehensive, longitudinal, and global datasets to study if we want to move beyond single countries and analyze the phenomena by discipline and over time. "Leaving science" is undergoing significant transformations as new cohorts of scientists enter science every year under new (working, professional, and other) conditions (Milojevic et al., 2018).

Our cohort-based and longitudinal study shows that behind aggregated changes at the level of all STEMM disciplines combined, there are widely nuanced changes that are occurring, with various intensities at the level of disciplines and over time. Attrition in science means different things for men and different things for women in different disciplines, and it means different things for scientists from different cohorts entering the scientific workforce.

About one-third of the 2000 cohort of scientists leaves science after 5 years, about a half after 10 years, and about two-thirds by the end of the period examined (after 19 years), with the share of the leavers being consistently lower for men and higher for women. Women are about one-tenth more likely to drop out of science than men after both 5 and 10 years (12.54% and 11.52%, respectively) and are 6.33% more likely to drop out at the end of the studied period.

Gender differences in attrition are becoming smaller with each successive cohort of scientists entering science between 2000 and 2010: with ever more women in science and ever more women within cohorts, attrition is becoming ever less gendered. At the level of 38 OECD countries, gender differences clearly visible for the 2000 cohort for all disciplines combined disappear for the 2010 cohort. However, the aggregated pictures hide behind them nuanced disaggregated pictures for both cohorts, with different ongoing processes for different disciplines.

The Kaplan–Meier curves for the two contrasted disciplines of biochemistry, genetics, and molecular biology (BIO, large, 47.83% of women) and physics and astronomy (PHYS, small, only 15.62% of women) for the 2000 cohort tell fundamentally different stories: in BIO, women are characterized by markedly lower survival rates than men, with the divergence increasing with each successive time interval, and in PHYS, the survival rates for men and women for two decades are nearly identical. Women in BIO disappear from science in ever-larger proportions over two decades compared with men; in contrast, women in PHYS disappear in almost exactly the same proportions as men for the whole period examined.

However, for the 2010 cohort, the dramatic lack of differences in attrition for men and women for all disciplines combined has much less pronounced but still existing gender differences in attrition in disciplines. In both cases, in such math-intensive disciplines as MATH, COMP, and PHYS, gender differences are nonexistent. In the highly mathematized disciplines, the share of women entering science is small or very small, but women stay on in science in exactly the same proportions as men—which may suggest that they are extremely professionally successful, despite possible chilly or unwelcoming workplace climates. In disciplines with very low representation of women, where women are very visible minorities (constituting 20% and less of all publishing scientists and often acting the role of "tokens," or exemplary figures in university departments (Kanter, 1977), the newcoming and surviving women stay in the system of science as powerfully as men do (see also Supplementary Figure 3 for COMP and Supplementary Figure 5 for MATH).

Our regression analysis goes beyond two-dimensional analyses of attrition from various angles (men, women, and disciplines) and estimates odds ratios of staying in science within a universe of many bibliometric-type factors. In the models, we have been especially interested in the role of publication quantity and quality, as well as gender. There are several key messages from our regressions.

First, in the presence of numerous independent variables, publication numbers (lifetime scholarly output) emerge as the single most important predictor, increasing the odds of staying in science. The role of publication quantity is lower for the 2000 cohort compared with the most recent 2010 one. Our models show that the quantity of publications is substantially more important as a predictor of increasing the odds of staying in science than the quality of publications (all other things being equal). On average, higher productivity is correlated with retention in science, and lower productivity is correlated with attrition in science, with different impacts of publishing one more publication across different disciplines.

Second, in contrast, the two quality-related independent variables we computed are less important than expected: neither citations to individual publication portfolios computed for 4-year citation windows nor lifetime average journal percentile ranks can serve as influential predictors of staying in science. However, for both cohorts, the percentile rank predictor is statistically significant for all disciplines, and its power is smaller than expected based on the literature linking top journals and academic careers (Hamermesh & Pfann, 2011; Heckman & Moktan, 2018). Lifetime publishing in journals ranked 30 percentiles higher in the Scopus database, on average, increases the odds in the majority of disciplines by about 30%. The direction of influence of this variable follows general expectations from the literature—more publications in top journals have a positive impact on the persistence of academic careers—but its power is unexpectedly low (Fochler et al., 2016; Hammarfelt, 2017; Lindahl, 2018; Shibayama & Baba, 2015). One explanation could be that our models have been constructed for disciplines in 38 countries, with different emphases on journal quality (as proxied by Scopus percentile ranks) in different countries.

Third, the role of gender is smaller than expected from previous literature (Preston, 2004, Deutsch & Yao, 2014; Levine et al., 2011): in a multidimensional regression analysis, in the presence of other variables, being male is not statistically significant predictor of staying in science in the vast majority of disciplines.

Finally, bibliometric independent variables tend to explain retention in science relatively well. The model we have constructed for hundreds of thousands of scientists using micro-level variables fits the phenomenon well, with pseudo R^2 for all disciplines in the 0.40–0.60 range for the oldest cohort and with pseudo R^2 in the range of 0.30–045 for the youngest cohort.

Although women have traditionally been believed (Preston, 2004; 2014; Goulden et al., 2011; Shaw & Stanton, 2012; Wolfinger et al., 2008) to be leaving science earlier and in higher proportions than men (which is generally confirmed in our analyses of the 2000 cohort), leading to much higher attrition rates across many disciplines, for more recent cohorts, in contrast, the gender differences in attrition rates may be no longer present, especially for math-intensive disciplines with low numbers and percentages of women. For new generations of scientists, generally, attrition in science has been on the rise and very high (58.6% of women and 57.6% of men from the cohort 2010 disappear from science within 9 years)—but it seems to be much less gendered than traditionally assumed.

Changes in the participation of women in science over the past three decades have been tectonic, and bibliometric-based longitudinal studies of men and women scientists open new opportunities for global cross-gender and cross-cohort analyses. In a fast-changing science environment (Fochler et al., 2016; Wang & Barabási, 2021; Stephan, 2012), with hundreds of thousands of newcomers to science every year, our traditional assumptions about how men and women disappear from science may need careful revisions, and our intention was to sketch some tentative general answers and possible directions for further, more detailed studies.

This paper is accompanied by Supplementary Material *available online.*

Acknowledgments

MK gratefully acknowledges the support provided by the MEiN NDS grant no. NdS/529032/2021/2021. LS is grateful for the support of his doctoral studies provided by the NCN grant 2019/35/0/HS6/02591. 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 many fruitful discussions.

References

Allis on, P.D. (2014). *Event History and Survival Analysis*. Sage.

Alper, J. (1993). The Pipeline Is Leaking Women All the Way Along. *Science*, Vol. 260, 16 April,

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

Blickenstaff, J.C. (2005). Women and science careers: leaky pipeline or gender filter? *Gender and Education,* 17(4), 369–386.

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

Britton, D. M. (2017). Beyond the Chilly Climate: The Salience of Gender in Women's Academic Careers. *Gender & Society*, *31*(1), 5-27.

Cornelius, R., Constantinople, A., & Gray, J. (1988). The Chilly Climate: Fact Or Artifact? *The Journal of Higher Education*, 59(5), 527–55.

Deutsch, F.M., & Yao, B. (2014). Gender Differences in Faculty Attrition in the USA. *Community, Work &. Family*, 17, 392–408.

Ehrenberg, R.G., Kasper, H., & Rees, D.I. (1991). Faculty Turnover in American Colleges and Universities. *Economics of Education Review*, 10(2), 99–110.

Fochler, M., Felt, U., & Müller, R. (2016). Unsustainable Growth, Hyper-Competition, and Worth in Life Science Research: Narrowing Evaluative Repertoires in Doctoral and Postdoctoral Scientists' Work and Lives. *Minerva*, 54(2), 175–200.

Geuna, A., & Shibayama, S. (2015). Moving out of Academic Research: Why Do Scientists Stop Doing Research? In: A. Geuna (Ed.), *Global Mobility of Research Scientists* (pp. 271–297). Amsterdam: Elsevier.

Goulden, M., Mason, M.A., & Frasch, K. (2011). Keeping women in the science pipeline. *The ANNALS of the American Academy of Political and Social Science*, 638, 141–162 (2011).

Hamermesh, D.S., & Pfann, G.A. (2011). Reputation and Earnings: The Roles of Quality and Quantity in Academe. Economic Inquiry, 50(1), 1–16.

Hammarfelt, B. (2017). Recognition and Reward in the Academy: Valuing Publication Oeuvres in Biomedicine, Economics and History. *Aslib Journal of Information Management*, 69(5), 607–623.

Heckman, J.J., & Moktan, S. (2018). Publishing and Promotion in Economics. The Tyranny of the Top Five. *NBER Working Paper* 25093.

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.

Kaminski, D., & Geisler, C. (2012). Survival Analysis of Faculty Retention in Science and Engineering by Gender. *Science*, 335, 864–866.

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.

Kashyap, R., Rinderknecht, R.G., Akbaritabar, A., Alburez-Gutierrez, D., Gil-Clavel, S., Grow, A., Zhao, X., et al. (2022). *Digital and Computational Demography*. https://doi.org/10.31235/osf.io/7bvpt

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., & 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. (2022). Are Female Scientists Less Inclined to Publish Alone? The Gender Solo Research Gap. *Scientometrics*, 127, 1697–1735.

Kwiek, M., & Roszka, W. (2023a). Once Highly Productive, Forever Highly Productive? Full Professors' Research Productivity from a Longitudinal Perspective. *Higher Education*. Online first https://doi.org/10.1007/s10734-023-01022-y

Kwiek, M., Roszka, W. (2023b). The Young and the Old, the Fast and the Slow: A Large-Scale Study of Productivity Classes and Rank Advancement. *Studies in Higher Education* (accepted, in print). Online-first, soon available from: https://doi.org/10.1080/03075079.2023.2288172

Kwiek, M., & Szymula, L. (2023). Young Male and Female Scientists: A Quantitative Exploratory Study of the Changing Demographics of the Global Scientific Workforce. *Quantitative Science Studies*. Online first: https://doi.org/10.1162/qss_e_00276

Levine, R.B., Lin, F., Kern, D.E., Wright, S.M., & Carrese, J. (2011). Stories from Early-Career Women Physicians Who Have Left Academic Medicine: a Qualitative Study At a Single Institution. *Academic Medicine*, 86, 752–758.

Lindahl, J. (2018). Predicting Research Excellence at the Individual Level: The Importance of Publication Rate, Top Journal Publications, and Top 10% Publications in the Case of Early Career Mathematicians. *Journal of Informetrics*, 12(2), 518–533.

Liu, L., Jones, B.F., Uzzi, B., et al. (2023). Data, Measurement and Empirical Methods in the Science of Science. *Nature Human Behaviour*, 7, 1046–1058.

Maranto, C. L., & Griffin, A. E. (2010). The antecedents of a "chilly climate" for women faculty in higher education. *Human Relations,* 64(2), 139–159.

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.

Mills, M. (2011). *Introducing Survival and Event History Analysis*. Sage.

Milojevic, S., Radicchi, F., & Walsh, J.P. (2018). Changing Demographics of Scientific Careers: the Rise of the Temporary Workforce. *Proceedings of the National Academy of Sciences*, 115. 12616–12623.

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.

O'Meara, K., Bennett, J.C., & Niehaus, E. (2016). Left Unsaid: The Role of Work Expectations and Psychological Contracts in Faculty Careers and Departure. *The Review of Higher Education*, 39(2), 269–297

Preston, A.E. (2004). *Leaving Science. Occupational Exit from Scientific Careers*. Russell Sage Foundation.

Rosser, V.J. (2004). Faculty Members' Intentions to Leave: A National Study on their Worklife and Satisfaction. *Research in Higher Education*, 45(3), 285–309.

Santos, J. M., Horta, H., & Amâncio, L. (2020). Research agendas of female and male academics: a new perspective on gender disparities in academia. *Gender and Education,* 1–19.

Shaw, A.K., & Stanton, D.E. (2012). Leaks in the Pipeline: Separating Demographic Inertia from Ongoing Gender Differences in Academia. *Proceedings of the Royal Society B. Biological Sciences*, 279(1743), 3736–41.

Shibayama, S., & Baba, Y. (2015). Impact-Oriented Science Policies and Scientific Publication Practices: The Case of Life Sciences in Japan. *Research Policy*, 44(4), 936–950.

Smart, J.C. (1990). A Casual Model of Faculty Turnover Intentions. Research in Higher Education, 31(5), 405–424.

Spoon, K. et al. (2023). Gender and Retention Patterns Among U.S. Faculty. *Science Advances*, 9, eadi2205. DOI:10.1126/sciadv.adi2205

Stephan, P.E. (2012). *How Economics Shapes Science*. Harvard University Press.

Sugimoto, C., & Larivière, V. (2023). *Equity for Women in Science. Dismantling Systemic Barriers to Advancement*. Harvard University Press.

Tang, L. & Horta, H. (2023). Supporting Academic Women's Careers: Male and Female Academics' Perspectives at a Chinese Research University. *Minerva*. https://doi.org/10.1007/s11024-023-09506-y

Wang, D., & Barabási, A.-L. (2021). *The Science of Science*. Cambridge University Press.

White-Lewis, D.K., O'Meara, K., Mathews, K., et al. (2023). Leaving the Institution or Leaving the Academy? Analyzing the Factors that Faculty Weigh in Actual Departure Decisions. *Research in Higher Education*, 64, 473–494.

Wohrer, V. (2014). To Stay or to Go? Narratives of Early-Stage Sociologists About Persisting in Academia. *Higher Education Policy*, 27, 469–487.

Wolfinger, N.H., Mason, M.A., & Goulden, M. (2008). Problems in the Pipeline: Gender, Marriage, and Fertility in the Ivory Tower. *Journal of Higher Education*, 79(4), 388–405.

Xu, Y.J. (2008). Gender Disparity in STEM Disciplines: A Study of Faculty Attrition and Turnover Intentions. *Research in Higher Education* 49, 607–624.

Zhou, Y., & Volkwein, J. F. (2004). Examining the influence on faculty departure intentions: A comparison of tenured versus nontenured faculty at research universities using NSOPF-99. *Research in Higher Education*, 45(2), 139–176.

Supplementary Material to:

Quantifying Attrition in Science: A Cohort-Based and Longitudinal Study of Scientists in 38 OECD Countries

This supplementary material contains:

S1. Data

S2. Methods

S3. Supplementary Tables

S4. Supplementary Figures

S1. Data

The data show that the analyzed cohorts have a quicker increase in female scientists than male scientists, with the share of women publishing for the first time in a given year rising from 36.50% in 2000 to 41.72% in 2010 (Panel 1 of Supplementary Table 1). Panels 2 and 3 of Table ST1 show the distribution of the 2000 cohort and 2010 cohort by discipline. For both cohorts, the largest percentage of new publishers was assigned to MED Medicine (39.12% and 40.11%, respectively), with the next two largest disciplines being BIO and ENG (see column percentages). Overall, the three largest disciplines comprise nearly two-thirds of STEMM scientists, and the structure of scientists by largest discipline is relatively unchanged over time: the three largest disciplines comprise 63.27% of scientists for the 2000 cohort and 62.57% of scientists for the 2010 cohort. For both cohorts, nearly 7 out of every 10 women scientists publishing for the first time are in two disciplines (MED and BIO, 69.31% and 67.90%, respectively). The next disciplines with a high number of women are CHEM and AGRI (5.88% and 5.77%, respectively) for the 2000 cohort and AGRI and CHEM (6.21% and 4.84%, respectively) for the 2010 cohort. In contrast, the smallest share of women scientists in both cohorts by discipline is observed for ENG, ENER, PHYS, and COMP (see line percentages, 10.89%, 12.50%, 15.62%, and 15.725%, respectively) for the first cohort and ENG, COMP, PHYS, and ENER (14.07%, 17.96%, 18.28% and 19.27%, respectively).

For both cohorts, the largest percentage of scientists was assigned to the US (more than 30%, as shown in Table ST 2), with Japan being the second largest country (13.78% for the 2000 cohort and 10.06% for the 2010 cohort, respectively). Scientists from the 10 countries with the highest number of observations comprise 82.48% for the 2000 cohort and 79.324% for the 2010 cohort (the remaining 28 countries are included in the Other category).

Supplementary Table 1: Panels 1–3. The population of all 11 cohorts (2000 to 2010). Numbers and row and column percentages for the 2000 and 2010 cohorts by discipline and gender

		Men			**Women**			**Total**		
		N	**% row**	**% col**	**N**	**% row**	**% col**	**N**	**% row**	**% col**
	All cohorts	**1,289,756**	**60.61**	**100.00**	**838,047**	**39.39**	**100.00**	**2,127,803**	**100.00**	**100.00**
Panel 1: All cohorts by year	2000	90,661	63.50	7.03	52,115	36.50	6.22	142,776	100.00	6.71
	2001	93,887	63.03	7.28	55,079	36.97	6.57	148,966	100.00	7.00
	2002	103,310	62.30	8.01	62,514	37.70	7.46	165,824	100.00	7.79
	2003	108,564	61.96	8.42	66,664	38.04	7.95	175,228	100.00	8.24
	2004	117,442	61.47	9.11	73,624	38.53	8.79	191,066	100.00	8.98
	2005	128,452	61.58	9.96	80,129	38.42	9.56	208,581	100.00	9.80
	2006	126,171	60.51	9.78	82,339	39.49	9.83	208,510	100.00	9.80
	2007	126,128	59.58	9.78	85,568	40.42	10.21	211,696	100.00	9.95
	2008	127,243	58.72	9.87	89,437	41.28	10.67	216,680	100.00	10.18
	2009	132,200	58.59	10.25	93,433	41.41	11.15	225,633	100.00	10.60
	2010	135,698	58.28	10.52	97,145	41.72	11.59	232,843	100.00	10.94
Panel 2: the 2000 Cohort by discipline	AGRI	4,962	62.26	5.47	3,008	37.74	5.77	7,970	100.00	5.58
	BIO	11,839	52.17	13.06	10,853	47.83	20.83	22,692	100.00	15.89
	CHEM	6,409	67.66	7.07	3,063	32.34	5.88	9,472	100.00	6.63
	CHEMENG	615	81.56	0.68	139	18.44	0.27	754	100.00	0.53
	COMP	5,414	84.28	5.97	1,010	15.72	1.94	6,424	100.00	4.50
	EARTH	3,028	72.77	3.34	1,133	27.23	2.17	4,161	100.00	2.91
	ENER	574	87.50	0.63	82	12.50	0.16	656	100.00	0.46
	ENG	10,511	89.11	11.59	1,285	10.89	2.47	11,796	100.00	8.26
	ENVIR	2,059	66.63	2.27	1,031	33.37	1.98	3,090	100.00	2.16
	IMMU	805	45.35	0.89	970	54.65	1.86	1,775	100.00	1.24
	MATER	2,294	78.11	2.53	643	21.89	1.23	2,937	100.00	2.06
	MATH	1,345	77.30	1.48	395	22.70	0.76	1,740	100.00	1.22
	MED	30,586	54.76	33.74	25,264	45.24	48.48	55,850	100.00	39.12
	NEURO	1,353	53.41	1.49	1,180	46.59	2.26	2,533	100.00	1.77
	PHARM	632	54.16	0.70	535	45.84	1.03	1,167	100.00	0.82
	PHYS	8,235	84.38	9.08	1,524	15.62	2.92	9,759	100.00	6.84

Panel 3: the 2010 Cohort by discipline	AGRI	6,755	52.81	4.98	6,037	47.19	6.21	12,792	100.00	5.49
	BIO	14,905	47.25	10.98	16,637	52.75	17.13	31,542	100.00	13.55
	CHEM	8,158	63.45	6.01	4,699	36.55	4.84	12,857	100.00	5.52
	CHEMENG	831	74.06	0.61	291	25.94	0.30	1,122	100.00	0.48
	COMP	12,403	82.04	9.14	2,716	17.96	2.80	15,119	100.00	6.49
	EARTH	4,292	67.35	3.16	2,081	32.65	2.14	6,373	100.00	2.74
	ENER	1,399	80.73	1.03	334	19.27	0.34	1,733	100.00	0.74
	ENG	17,818	85.93	13.13	2,918	14.07	3.00	20,736	100.00	8.91
	ENVIR	3,130	55.39	2.31	2,521	44.61	2.60	5,651	100.00	2.43
	IMMU	939	39.77	0.69	1,422	60.23	1.46	2,361	100.00	1.01
	MATER	4,247	70.94	3.13	1,740	29.06	1.79	5,987	100.00	2.57
	MATH	1,874	74.99	1.38	625	25.01	0.64	2,499	100.00	1.07
	MED	44,060	47.18	32.47	49,325	52.82	50.77	93,385	100.00	40.11
	NEURO	2,153	47.71	1.59	2,360	52.29	2.43	4,513	100.00	1.94
	PHARM	876	52.68	0.65	787	47.32	0.81	1,663	100.00	0.71
	PHYS	11,858	81.72	8.74	2,652	18.28	2.73	14,510	100.00	6.23

Supplementary Table 2: Numbers and column percentages for the 2000 and 2010 cohorts by country, the top 10 countries by number of observations for the 2000 and 2010 cohorts (the remaining 28 OECD countries clustered as Other)

	The 2000 Cohort		**The 2010 Cohort**	
	N	**%**	**N**	**%**
Australia	3,456	2.42	6,221	2.67
Canada	4,891	3.43	9,329	4.01
France	7,645	5.35	10,860	4.66
Germany	12,061	8.45	21,236	9.12
Italy	6,736	4.72	10,796	4.64
Japan	19,671	13.78	23,420	10.06
South Korea	3,152	2.21	7,703	3.31
Spain	4,834	3.39	9,041	3.88
United Kingdom	10,222	7.16	14,574	6.26
United States	45,093	31.58	71,565	30.74
Other	25,015	17.52	48,098	20.66
Total	142,776	100.00	232,843	100.00

S2. Methods

Determining gender

To determine the scientist's gender, a dataset from the approach used by Elsevier for gender reports and made available in the ICSR Lab platform was used. The dataset contained the author's identifier from the Scopus database and two variables determined using the Namsor tool: gender and probability score based on three input parameters: author's first name, author's last name, and author's dominant country from the first year of publishing (N_{author}=34,596,581). A gender probability score greater than or equal to 0.85 (N_{author}=21,508,029) was selected to ensure high-quality gender classification. This threshold has been suggested and described by Elsevier as a value that returns high evaluation metrics (15).

Determining discipline

To classify the author into a discipline, the dominant discipline from the disciplines assigned to the journals of the cited papers in the author's lifetime publishing portfolio was used. For this purpose, a table of publications was selected. All papers cited by the author ($N_{citedreference}$=2,092,766,869) were collected from the author's publication portfolio of any type. Each cited paper had assigned disciplines that came from sources (journals, etc.). Disciplines assigned to sources were based on 4-digit ASJC codes. To switch to 2-digit codes, the first two digits of the 4-digit value were selected. To avoid repetition of the same 2-digit discipline at the level of the cited paper, only unique values were selected. The next step was to count for each author how many times they cited a paper from a given discipline. Assigning the dominant discipline to an author involved selecting the discipline for which the number of cited references was the highest (modal value). Authors who failed to be assigned a discipline or had two or more dominant disciplines were removed from the sample ($N_{included}$=32,794,309, $N_{removed}$=12,862,447). The set of authors and their disciplines was then narrowed down to only those who had a discipline from the STEMM field ($N_{included}$=29,927,584, $N_{excluded}$=2,866,725).

Determining country affiliation

To classify an author into a country, the dominant country indicated in the author's publication portfolio was used. For this purpose, a publication table was selected. From the publication portfolio of an author of any type of publications, all the affiliation countries indicated by the author were collected. Then, for each author, the number of times they indicated a country was counted. The assignment of the dominant country was based on selecting the country for which the number of publications of any type was the highest (modal value). Authors who failed to be assigned to a country or had two or more dominant countries were removed from the sample ($N_{inclued}$=39,405,552, $N_{excluded}$=6,251,204). The set of authors and their countries was then narrowed down to OECD countries only ($N_{inclued}$=23,619,928, $N_{excluded}$=15,785,624).

Determining the TOP200 institutional affiliation

To determine the author's TOP200/Rest institutional affiliation, two steps were required. (1) At the very beginning, the author's institution was determined; (2) then, a ranking of institutions was created with an indication of the institution's affiliation to Top200 or Rest classes. (3) Finally, the value of the institutional affiliation to Top200 or Rest was assigned to the author based on the author's institution and class (Top200, Rest) according to the ranking.

To classify an author into an institutional affiliation, the dominant institution indicated in the author's lifetime publication portfolio was used. For this, the publication table was selected. From the publication portfolio of an author of any type of publications, all the affiliations indicated by the author were collected. Then, for each author, the number of times they indicated a particular affiliation was counted. Assignment of the dominant affiliation was based on selecting the affiliation for which the number of publications of any type was the highest (modal value). Authors who failed to be assigned an affiliation or had two or more dominant affiliations were removed from the sample ($N_{inclued}$=36,658,578, $N_{excluded}$=8,998,178).

The next step was to create a ranking of institutions. For this, a table of publications from 2018 to 2022 inclusive was selected. For each institution indicated by the author in the publication, the number of publications was counted. Then, the list of institutions was ranked (from the institution with the highest number of publications to the institution with the lowest number of publications; $N_{institutions}$=1,655,061).

For the resultant ranking, the institution's membership in one of two classes was determined: Top200 and Rest. For an institution to be classified as Top200, it had to be among the top 200 institutions in the ranking. If the institution was not among the top 200 institutions, it received a Rest value.

Finally, each author, according to their dominant institution, was assigned the value of the institution's membership in the Top200 or Rest according to the ranking or the value of Rest if the institution was not ranked.

Determining the year of the start of publishing career

Publication tables were selected to assign authors to the appropriate cohorts (to indicate the first year of their publishing career). For each author, their identifier from the Scopus database and year of their first publication of any type were determined. The set of authors was then narrowed down only to those whose year of first publication was between 2000 and 2010 inclusive. The 2000 ($N_{authors}$=142,776) and 2010 ($N_{authors}$=232,843) cohorts were selected, and all 11 cohorts from 2000 to 2010 ($N_{authors}$=2,127,803) were selected for further comparisons in the logistic regression models.

Determining publication minimum

To identify nonoccasional scientists, it was decided to include in the sample scientists with at least two publications of any type. For this, a publication table was used, based on which the number of publications was counted for each scientist. Then, the set of scientists was narrowed down only to those who had a minimum of two publications ($N_{included}$=21,411,651, $N_{exclude}$d=24,245,105). No filtering by source and publication type was applied when selecting the scientists' publications.

Determining journal percentile ranks (lifetime)

To assign journal percentile scores to authors, their lifetime publication portfolios were determined. For each publication, the value of the highest percentile of the CiteScore metric from the journal for 2022 was assigned. The selection of the highest percentile value was based on choosing the journal percentile for the discipline for which the percentile value was the highest. Then, to determine the percentile value at the author's level, its average value from all the given author's publications was determined. This variable was used in logistic regression models.

Determining international collaboration rate

To assign an international collaboration rate to authors, their lifetime publication portfolio was selected. Each publication was initially classified into two binary variables: (1) international publication and (2) collaborative publication. For a publication to be classified as an international publication, the number of unique countries from the authors' affiliations had to be greater than or equal to 2. For a publication to be classified as a collaborative publication, the number of authors also had to be greater than or equal to 2. The next step was to identify a publication as an international collaborative publication if it was labeled as both international and collaborative. To determine an author's international collaboration rate, the percentage of their number of international collaborative publications to their number of total collaborative publications was calculated. This variable was used in logistic regression models.

Determining field-weighted our-year citation impact

To assign FWCI_4y values to authors, their lifetime publication portfolio was selected. For each publication, the value of its FWCI 4y metric determined and provided by ICSR Lab was attached. The rule for determining a publication's FWCI 4y was based on ASJC's 334 disciplines, referring to the ratio of the number of citations obtained by a publication in the year of publication and three consecutive years to the average number of citations of publications from the same discipline in the same period. Then, to determine the value of the 4-year FWCI metric at the author level, the Elsevier approach of calculating its average value from the author's publications was used. This variable was used in logistic regression models.

Determining average team size

To assign team size values to authors, their lifetime publication portfolios were selected. For each publication, their number of authors, including the author, was counted. To avoid cases of excessively large teams, if the number of authors was greater than or equal to 10, a value of 10 was assigned. If the number of authors was less than 10, then the number of authors remained unchanged. Next, to determine the value of team size at the author level, the median value of the author's publication team size was used. This variable was used in the logistic regression models.

Determining the exit year from publishing and classifying censored observations

To assign authors to the year of exit from publishing, the tables of publications were selected. For each author, their identifier from the Scopus database and year of last publication of any type were indicated. Then, the year following the year of last publication was selected as the year of exit from publishing (year of last publication + 1). To classify an author as a censored observation, the year 2019 was selected as the last year in the study. Any author for whom the year of exiting publishing was after 2019 (>2019) was classified as a censored observation.

Determining Scholarly Output

To determine scholarly output for authors, their publication portfolio was selected. For each author, their number of publications from the entire publication period (from the start of the cohort up to and including the last publication, or October 2022 if the author was marked as remaining in publishing) was counted. This variable was determined for the logistic regression models.

Publishing breaks and publishing frequency

To estimate the magnitude of the error in the case of a potential return to publishing by scientists after 2022 (for which years we do not yet have the data), we conducted analyses of the median frequency of publishing by scientists on an annual basis for the two cohorts studied. For the 2000 cohort, the median publication frequency for almost 80% of newly publishing scientists (86.70% of men and 83.66% of women, see Table ST 3) is to publish annually: break 0 years column over a 12-year period (2010–2021); for 7.96% of men and 9.82% of women, the publication frequency is every second year.

In contrast, for the 2010 cohort, the median publication frequency for two-thirds of newly publishing scientists (69.55% of men and 65.10% of women) is publishing annually, and for more than 15% of scientists, it is publishing every 2 years (men 16.66%, women 18.70%). In both cases, there are large disciplinary variations.

Supplementary Table 3: Median publication breaks for the 2000 cohort for the subpopulation of scientists publishing in the last 12 years (2010–2021), percent of scientists, men (left panel) and women (right panel). Zero year break means publishing at least one article every year; one year break means publishing every second year; two year break means publishing at least one article every third year, etc. (percent)

	Men scientists					**Women scientists**				
	0 years	**1 year**	**2 years**	**3 years**	**4 and more years**	**0 years**	**1 year**	**2 years**	**3 years**	**4 and more years**
AGRI	85.75	8.34	2.65	0.90	2.36	84.29	9.75	2.29	1.26	2.41
BIO	87.52	8.23	2.27	0.74	1.24	81.78	11.28	3.07	1.94	1.93
CHEM	86.85	7.92	2.00	1.08	2.15	85.09	8.80	2.68	0.96	2.47
CHEMENG	85.05	7.48	3.74	1.87	1.86	82.14	14.29			3.57
COMP	88.51	7.20	1.94	0.80	1.55	89.58	6.55	2.38	0.30	1.19
EARTH	86.25	7.15	2.29	1.92	2.39	86.51	8.73	1.59	1.06	2.11
ENER	83.23	11.61	2.58	1.94	0.64	81.25	12.50		6.25	0.00
ENG	83.22	9.41	3.40	1.52	2.45	86.33	7.03	3.12	1.17	2.35
ENVIR	84.71	7.48	4.62	1.11	2.08	82.98	11.70	2.13	1.42	1.77
IMMU	86.44	8.47	3.39	1.13	0.57	79.26	10.11	3.72	5.32	1.59
MATER	89.09	7.44	1.65	0.83	0.99	88.16	7.89	2.63	0.66	0.66
MATH	82.43	13.18	2.72	1.05	0.62	75.00	15.32	6.45	1.61	1.62
MED	86.18	8.14	2.21	1.26	2.21	83.12	10.00	2.84	1.44	2.60
NEURO	89.58	7.71	1.25	0.83	0.63	85.53	9.54	2.30	0.66	1.97
PHARM	80.99	12.40	1.65	1.65	3.31	82.43	9.46	5.41	2.70	0.00
PHYS	90.67	5.28	2.01	0.63	1.41	90.62	4.79	2.40	0.80	1.39
TOTAL	86.70	7.96	2.36	1.10	1.88	83.66	9.82	2.79	1.47	2.26

Supplementary Table 4:, Median publication breaks for the 2010 cohort for the subpopulation of scientists publishing over the past 12 years (2010–2021), percent of scientists, men (left panel) and women (right panel).

	Men scientists					Women scientists				
	0 years	1 year	2 years	3 years	4 and more years	0 years	1 year	2 years	3 years	4 and more years
AGRI	62.30	18.97	7.44	3.59	7.70	62.45	19.69	7.68	3.58	6.60
BIO	68.78	17.94	6.23	2.85	4.20	64.36	20.96	6.92	3.24	4.52
CHEM	72.49	15.27	5.50	2.57	4.17	70.11	16.92	5.96	2.89	4.12
CHEMENG	71.67	15.80	5.35	2.35	4.83	65.30	21.27	5.60	1.12	6.71
COMP	74.07	15.80	4.83	2.01	3.29	74.20	15.01	5.77	2.03	2.99
EARTH	65.17	17.76	6.66	3.82	6.59	64.54	19.19	6.93	3.52	5.82
ENER	64.95	19.20	7.44	2.43	5.98	68.04	19.30	5.06	2.85	4.75
ENG	70.92	16.45	5.51	2.74	4.38	70.17	16.39	6.00	2.53	4.91
ENVIR	62.88	18.68	7.93	3.29	7.22	61.38	20.18	7.87	3.91	6.66
IMMU	65.95	17.88	7.97	2.85	5.35	66.77	19.65	6.94	3.25	3.39
MATER	74.41	15.26	5.20	2.27	2.86	74.24	16.18	4.67	2.04	2.87
MATH	66.87	18.39	7.26	3.38	4.10	61.59	23.34	7.78	3.48	3.81
MED	66.53	17.15	6.49	3.31	6.52	63.44	18.46	7.15	3.71	7.24
NEURO	68.91	18.86	5.29	2.94	4.00	65.77	20.23	6.27	2.98	4.75
PHARM	61.90	20.30	6.39	3.88	7.53	59.58	20.56	7.78	3.47	8.61
PHYS	80.65	12.28	3.25	1.53	2.29	81.83	11.68	3.25	1.19	2.05
TOTAL	69.55	16.66	5.89	2.85	5.05	65.10	18.70	6.85	3.36	5.99

S3. Supplementary Tables

Examining the differences between men and women in attrition, we analyze the median time to leave for the whole cohort (also the probability of staying for the value 0.5 for proportion of staying in Kaplan–Meier survival curves for each discipline separately) and the mean time to leave for the subpopulation of the cohort with scientists who left publishing, both by discipline and by gender and discipline.

The median time (in years) to leave science differs substantially by gender and discipline. For all disciplines except three, half of the women in the cohort starting in 2000 leave science sooner than half the men in the same cohort (while in CHEM, COMP, and MATH, the median time to leave is the same). The difference is especially stark in the disciplines with the highest numbers of women, such as MED and BIO, in which half of the women population leaves science sooner than half the men population (after 11 vs. 14 years and after 8 vs. 11 years, respectively; see Supplementary Table ST 5).

Supplementary Table 5: Median exit time for the 2000 cohort (calculated in years, from the beginning of the study period) by discipline and gender and the difference in time between men and women.

Discipline	Men (years)	Women (years)	Difference in years	Total
AGRI	18	14	4	17
BIO	11	8	3	9
CHEM	7	7	0	7
CHEMENG	8	7	1	8
COMP	12	12	0	12
EARTH	18	16	2	18
ENER	14.5	13	1.5	14
ENG	9	8	1	9
ENVIR	16	14	2	15
IMMU	9	7	2	8
MATER	10	9	1	10
MATH	18	18	0	18
MED	14	11	3	13
NEURO	13	9	4	11
PHARM	8	6	2	7
PHYS	12	11	1	12
TOTAL	12	10	2	11

Supplementary Table 6: Mean exit time from science (calculated in years) for the 2000 cohort for the subpopulation of scientists who exited science (not including scientists remaining in science; only observations labeled as uncensored remain), without gender breakdown, mean value, 95% confidence interval, and p-value by Student's t-test

Discipline	Mean (years)	Lower 95 CI	Upper 95 CI	P value
AGRI	8.796	8.632	8.959	<0.0001
BIO	7.341	7.262	7.419	<0.0001
CHEMENG	7.046	6.636	7.457	<0.0001
CHEM	6.612	6.504	6.720	<0.0001
COMP	7.908	7.745	8.070	<0.0001
EARTH	8.908	8.674	9.143	<0.0001
ENER	8.751	8.184	9.318	<0.0001
ENG	7.542	7.430	7.655	<0.0001
ENVIR	8.657	8.405	8.909	<0.0001
IMMU	6.951	6.691	7.211	<0.0001
MATER	7.442	7.213	7.670	<0.0001
MATH	8.256	7.883	8.629	<0.0001
MED	8.077	8.021	8.133	<0.0001
NEURO	7.439	7.193	7.686	<0.0001
PHARM	6.975	6.649	7.300	<0.0001
PHYS	7.778	7.647	7.909	<0.0001
TOTAL	7.775	7.742	7.809	<0.0001

We also tested differences between the survival curves for the two groups (men and women) within all disciplines to see whether they are statistically significant. From a wide variety of statistical tests and by following (56), we have selected six tests: the log-rank test, the Wilcoxon test, the Gehan test, the Peto and Peto test, the Peto and Peto test (by Andersen), and the Tarone–Ware test. A comparison of the test results is given in Supplementary Table 7. Although it is not useful to discuss the test results in detail, some conclusions need more attention. First, for all disciplines combined and for the two largest disciplines of MED and BIO with large percentages of women, the differences between the survival curves for men and women are statistically significant for all tests conducted. In contrast, for the three math-intensive disciplines of COMP, MATH, and PHYS, which have small percentages of women, the differences are not statistically significant for any of the tests used.

Supplementary Table 7: Statistical significance tests for the difference between groups (men, women), 6 methods, by discipline, for the 2000 cohort, Chisq and p-value given for each test

	log-rank		**Wilcoxon**		**Fleming-Harrington**		**Peto-Peto**		**Peto-Peto (by Andersen)**		**Tarone-Ware**	
Discipline	**Chisq**	**p-value**	**Chisq**	**p-value**	**Chisq**	**p-value**	**Chisq**	**p-value**	**Chisq**	**p-value**	**Chisq**	**p-value**
AGRI	48.194	0.000	59.842	0.000	21.831	0.000	59.961	0.000	59.962	0.000	54.913	0.000
BIO	269.074	0.000	228.836	0.000	257.291	0.000	229.686	0.000	229.682	0.000	254.455	0.000
CHEM	10.470	0.001	9.756	0.002	9.929	0.002	9.665	0.002	9.664	0.002	10.575	0.001
CHEMENG	0.046	0.831	0.002	0.961	0.001	0.977	0.001	0.971	0.001	0.970	0.001	0.969
COMP	0.109	0.742	0.175	0.676	1.628	0.202	0.163	0.686	0.163	0.686	0.004	0.947
EARTH	10.909	0.001	14.171	0.000	4.812	0.028	14.236	0.000	14.237	0.000	12.775	0.000
ENER	1.066	0.302	0.817	0.366	1.296	0.255	0.847	0.357	0.846	0.358	0.959	0.327
ENG	9.631	0.002	15.579	0.000	4.733	0.030	15.497	0.000	15.497	0.000	13.216	0.000
ENVIR	6.309	0.012	7.182	0.007	3.357	0.067	7.140	0.008	7.140	0.008	6.849	0.009
IMMU	6.389	0.011	6.823	0.009	4.529	0.033	6.799	0.009	6.799	0.009	6.839	0.009
MATER	0.023	0.878	0.413	0.520	0.011	0.917	0.384	0.536	0.384	0.535	0.192	0.661
MATH	0.199	0.655	0.060	0.807	0.646	0.421	0.068	0.794	0.068	0.794	0.119	0.730
MED	151.919	0.000	156.673	0.000	114.527	0.000	156.511	0.000	156.511	0.000	157.995	0.000
NEURO	39.434	0.000	38.425	0.000	30.168	0.000	38.699	0.000	38.696	0.000	39.736	0.000
PHARM	7.630	0.006	9.563	0.002	4.807	0.028	9.711	0.002	9.711	0.002	9.049	0.003
PHYS	0.056	0.812	0.002	0.965	0.657	0.418	0.002	0.966	0.002	0.966	0.024	0.876
TOTAL	347.634	0.000	351.413	0.000	293.414	0.000	350.815	0.000	350.814	0.000	359.962	0.000

Supplementary Table 8: Exit time for the 2010 cohort (in years, from the beginning of the study period), 50 percent of the population by discipline and gender, and the difference in time between men and women

Discipline	**Men (years)**	**Women (years)**	**Total (years)**	**Men / Women Difference (in years)**
AGRI	10	9	9	1
BIO	9	7	8	2
CHEM	7	6	6	1
CHEMENG	6	6	6	0
COMP	6	6	6	0
EARTH	10	9	10	1
ENER	7	7	7	0
ENG	6	6	6	0
ENVIR	9	8	9	1
IMMU	7	7	7	0
MATER	7	7	7	0
MATH	10	9	9	1
MED	9	8	9	1
NEURO	9	7	8	2
PHARM	6	5	6	1
PHYS	8	7	8	1
TOTAL	8	8	8	0

Supplementary Table 9: Mean time (in years) of exit from science for the 2010 cohort for the subpopulation of scientists who exited science (not including scientists remaining in science; only observations labeled as uncensored), without gender breakdown, mean value, 95% confidence interval, and p-value by Student's t-test

Discipline	**Mean (years)**	**Lower 95 CI**	**Upper 95 CI**	**P-value**
AGRI	4.876	4.815	4.937	<0.0001
BIO	4.832	4.797	4.867	<0.0001
CHEMENG	4.443	4.273	4.613	<0.0001
CHEM	4.606	4.556	4.656	<0.0001
COMP	4.473	4.426	4.520	<0.0001
EARTH	5.078	4.989	5.167	<0.0001
ENER	4.659	4.511	4.806	<0.0001
ENG	4.421	4.379	4.462	<0.0001
ENVIR	4.723	4.634	4.811	<0.0001
IMMU	4.633	4.511	4.755	<0.0001
MATER	4.625	4.547	4.702	<0.0001
MATH	4.842	4.707	4.976	<0.0001
MED	4.648	4.627	4.670	<0.0001
NEURO	4.868	4.774	4.962	<0.0001
PHARM	4.383	4.240	4.525	<0.0001
PHYS	4.705	4.654	4.757	<0.0001
TOTAL	4.662	4.649	4.675	<0.0001

Supplementary Table 10: Mean time (in years) of exit from science for the 2010 cohort for the subpopulation of scientists who exited science (not including scientists remaining in science; only observations marked as uncensored), by gender, mean value, 95% confidence interval, and p-value by Student's t-test

	Men				**Women**			
Discipline	**Mean (years**	**Lower 95 CI**	**Upper 95 CI**	**P_value**	**Mean (years)**	**Lower 95 CI**	**Upper 95 CI**	**P_value**
AGRI	4.901	4.815	4.988	<0.0001	4.850	4.764	4.936	<0.0001
BIO	4.861	4.808	4.914	<0.0001	4.808	4.762	4.855	<0.0001
CHEMENG	4.434	4.236	4.633	<0.0001	4.468	4.132	4.803	<0.0001
CHEM	4.650	4.586	4.714	<0.0001	4.535	4.455	4.615	<0.0001
COMP	4.503	4.450	4.555	<0.0001	4.337	4.227	4.447	<0.0001
EARTH	5.100	4.991	5.210	<0.0001	5.033	4.881	5.185	<0.0001
ENER	4.716	4.551	4.881	<0.0001	4.411	4.085	4.737	<0.0001
ENG	4.414	4.369	4.458	<0.0001	4.461	4.352	4.571	<0.0001
ENVIR	4.712	4.590	4.833	<0.0001	4.736	4.607	4.866	<0.0001
IMMU	4.620	4.415	4.825	<0.0001	4.641	4.489	4.793	<0.0001
MATER	4.549	4.456	4.641	<0.0001	4.808	4.667	4.949	<0.0001
MATH	4.868	4.710	5.026	<0.0001	4.769	4.514	5.023	<0.0001
MED	4.648	4.616	4.681	<0.0001	4.649	4.619	4.678	<0.0001
NEURO	4.881	4.737	5.025	<0.0001	4.858	4.735	4.982	<0.0001
PHARM	4.437	4.231	4.644	<0.0001	4.328	4.130	4.525	<0.0001
PHYS	4.699	4.642	4.756	<0.0001	4.733	4.614	4.851	<0.0001
TOTAL	4.648	4.630	4.665	<0.0001	4.681	4.661	4.701	<0.0001

The 2000 cohort: mean time to leave science, disciplinary variation

The mean time to leave science for scientist from the 2000 cohort who actually left science is generally between 6 and 9 years, depending on the discipline. For all disciplines combined by gender, the mean time to leave is 7.927 (7.884-7.970) years for men vs. 7.527 (7.474-7.581) years for women, with no overlapping 95% confidence intervals and p<0.0001 indicating statistically significant difference. For all disciplines except two, the mean time to leave for men is higher than the mean time to leave for women, that is, men on average stay longer in science than women. The only odd discipline is MATH where the opposite is the case.

Supplementary Table 11: Mean time (in years) to exit science for the 2000 cohort for the subpopulation of scientists who exited science (not including scientists remaining in science; only observations labeled as uncensored remain), by gender, mean value, 95% confidence interval, and p-value (Student's t-test)

	Men				**Women**			
Discipline	**Mean (years)**	**Lower 95% CI**	**Upper 95% CI**	**P-value**	**Mean (years)**	**Lower 95% CI**	**Upper 95% CI**	**P-value**
AGRI	9.221	9.007	9.436	<0.0001	8.166	7.917	8.416	<0.0001
BIO	7.578	7.463	7.693	<0.0001	7.113	7.006	7.220	<0.0001
CHEMENG	7.162	6.700	7.623	<0.0001	6.519	5.624	7.413	<0.0001
CHEM	6.707	6.574	6.841	<0.0001	6.419	6.237	6.600	<0.0001
COMP	7.991	7.814	8.168	<0.0001	7.449	7.040	7.858	<0.0001
EARTH	9.185	8.904	9.467	<0.0001	8.223	7.805	8.641	<0.0001
ENER	8.777	8.165	9.390	<0.0001	8.576	7.039	10.114	<0.0001
ENG	7.631	7.511	7.751	<0.0001	6.837	6.515	7.159	<0.0001
ENVIR	8.798	8.486	9.110	<0.0001	8.393	7.966	8.819	<0.0001
IMMU	7.195	6.797	7.593	<0.0001	6.759	6.416	7.102	<0.0001
MATER	7.568	7.307	7.828	<0.0001	6.986	6.513	7.460	<0.0001
MATH	8.184	7.758	8.610	<0.0001	8.493	7.718	9.269	<0.0001
MED	8.257	8.180	8.334	<0.0001	7.873	7.793	7.954	<0.0001
NEURO	7.782	7.425	8.140	<0.0001	7.105	6.766	7.444	<0.0001
PHARM	7.401	6.941	7.861	<0.0001	6.496	6.039	6.952	<0.0001
PHYS	7.781	7.638	7.925	<0.0001	7.762	7.437	8.086	<0.0001
TOTAL	7.927	7.884	7.970	<0.0001	7.527	7.474	7.581	<0.0001

Supplementary Table 12: Kaplan-Meier estimates for the 2010 cohort by gender with total numbers of men and women and by gender with percentages by gender. The values: time (in years), number of observations at the beginning of the study period, number of observations exiting the science, Kaplan-Meier probability of staying, standard error, and 95% confidence interval for Kaplan-Meier probability of staying, total

gender	Time (years)	n	n_leaving	km_probability_staying	standard_error	lower	upper
Women	1	97,145	5,030	0.948	0.001	0.947	0.950
	2	92,115	8,686	0.859	0.001	0.857	0.861
	3	83,429	8,090	0.776	0.001	0.773	0.778
	4	75,339	7,369	0.700	0.001	0.697	0.703
	5	67,970	6,470	0.633	0.002	0.630	0.636
	6	61,500	5,904	0.572	0.002	0.569	0.575
	7	55,596	5,499	0.516	0.002	0.513	0.519
	8	50,097	4,984	0.464	0.002	0.461	0.468
	9	45,113	4,929	0.414	0.002	0.411	0.417
Men	1	135,698	7,375	0.946	0.001	0.944	0.947
	2	128,323	12,183	0.856	0.001	0.854	0.858
	3	116,140	11,164	0.774	0.001	0.771	0.776
	4	104,976	9,869	0.701	0.001	0.698	0.703
	5	95,107	8,552	0.638	0.001	0.635	0.640
	6	86,555	7,851	0.580	0.001	0.577	0.583
	7	78,704	7,322	0.526	0.001	0.523	0.529
	8	71,382	7,078	0.474	0.001	0.471	0.477
	9	64,304	6,745	0.424	0.001	0.422	0.427

Supplementary Table 13: Statistical significance tests for the difference between groups (men, women), 6 methods, by discipline, for the 2010 cohort, Chisq and p-value given for each test

	log-rank		**Wilcoxon**		**Fleming-Harrington**		**Peto-Peto**		**Peto-Peto (by Andersen)**		**Tarone-Ware**	
Discipline	**Chisq**	**p_value**	**Chisq**	**p_value**	**Chisq**	**p_value**	**Chisq**	**p_value**	**Chisq**	**p_value**	**Chisq**	**p_value**
AGRI	25.276	0.000	23.786	0.000	22.011	0.000	23.606	0.000	23.606	0.000	24.763	0.000
BIO	116.265	0.000	102.331	0.000	114.957	0.000	100.782	0.000	100.781	0.000	110.694	0.000
CHEM	27.263	0.000	25.519	0.000	23.756	0.000	25.261	0.000	25.261	0.000	26.995	0.000
CHEMENG	0.035	0.852	0.007	0.933	0.127	0.721	0.010	0.919	0.010	0.919	0.018	0.892
COMP	0.096	0.757	0.927	0.336	0.354	0.552	0.963	0.326	0.963	0.326	0.448	0.503
EARTH	5.212	0.022	5.209	0.022	3.816	0.051	5.206	0.023	5.206	0.023	5.257	0.022
ENER	0.039	0.843	0.027	0.869	0.454	0.500	0.025	0.875	0.025	0.875	0.000	0.999
ENG	0.674	0.412	0.228	0.633	1.283	0.257	0.241	0.624	0.241	0.624	0.414	0.520
ENVIR	2.643	0.104	2.056	0.152	3.690	0.055	1.949	0.163	1.948	0.163	2.373	0.123
IMMU	5.444	0.020	3.800	0.051	8.365	0.004	3.540	0.060	3.539	0.060	4.687	0.030
MATER	0.092	0.761	1.225	0.268	1.806	0.179	1.323	0.250	1.323	0.250	0.518	0.472
MATH	2.713	0.100	2.712	0.100	2.358	0.125	2.653	0.103	2.653	0.103	2.755	0.097
MED	133.217	0.000	115.084	0.000	133.183	0.000	114.122	0.000	114.121	0.000	125.146	0.000
NEURO	40.606	0.000	35.183	0.000	40.236	0.000	34.564	0.000	34.562	0.000	38.245	0.000
PHARM	9.925	0.002	7.769	0.005	11.471	0.001	7.590	0.006	7.587	0.006	9.029	0.003
PHYS	1.741	0.187	1.058	0.304	2.949	0.086	1.062	0.303	1.062	0.303	1.394	0.238
TOTAL	17.687	0.000	9.812	0.002	37.827	0.000	9.362	0.002	9.362	0.002	13.654	0.000

Supplementary Table 14: Survival regression model (exponential method) for the 2010 cohort, women, abbreviated data including exp(B) for the time variable, R^2 metrics and p-value for the model in terms of disciplines

	Men				**Women**			
Discipline	**Mean (years)**	**Lower 95 (**	**Upper 9**	**P_value**	**Mean (years)**	**Lower 95 (**	**Upper 95 (**	**P_value**
AGRI	4.901	4.815	4.988	<0.0001	4.850	4.764	4.936	<0.0001
BIO	4.861	4.808	4.914	<0.0001	4.808	4.762	4.855	<0.0001
CHEMENG	4.434	4.236	4.633	<0.0001	4.468	4.132	4.803	<0.0001
CHEM	4.650	4.586	4.714	<0.0001	4.535	4.455	4.615	<0.0001
COMP	4.503	4.450	4.555	<0.0001	4.337	4.227	4.447	<0.0001
EARTH	5.100	4.991	5.210	<0.0001	5.033	4.881	5.185	<0.0001
ENER	4.716	4.551	4.881	<0.0001	4.411	4.085	4.737	<0.0001
ENG	4.414	4.369	4.458	<0.0001	4.461	4.352	4.571	<0.0001
ENVIR	4.712	4.590	4.833	<0.0001	4.736	4.607	4.866	<0.0001
IMMU	4.620	4.415	4.825	<0.0001	4.641	4.489	4.793	<0.0001
MATER	4.549	4.456	4.641	<0.0001	4.808	4.667	4.949	<0.0001
MATH	4.868	4.710	5.026	<0.0001	4.769	4.514	5.023	<0.0001
MED	4.648	4.616	4.681	<0.0001	4.649	4.619	4.678	<0.0001
NEURO	4.881	4.737	5.025	<0.0001	4.858	4.735	4.982	<0.0001
PHARM	4.437	4.231	4.644	<0.0001	4.328	4.130	4.525	<0.0001
PHYS	4.699	4.642	4.756	<0.0001	4.733	4.614	4.851	<0.0001
TOTAL	4.648	4.630	4.665	<0.0001	4.681	4.661	4.701	<0.0001

Supplementary Table 15: Survival regression model (exponential method) for the 2010 cohort, women, full data

Discipline	**Intercept - expB**	**Intercept - std_err**	**Intercept - tvalue**	**Intercept - pvalue**	**time - expB**	**time – std_err**	**time – tvalue**	**time - pvalue**	**Measures - R^2**	**Measures - pval**
AGRI	0.047	0.050	-60.597	<0.0001	3.596	0.032	40.001	<0.0001	0.996	<0.0001
BIO	0.050	0.059	-51.106	<0.0001	3.956	0.037	36.848	<0.0001	0.995	<0.0001
CHEMENG	0.091	0.054	-44.345	<0.0001	3.444	0.034	36.025	<0.0001	0.995	<0.0001
CHEM	0.068	0.061	-43.922	<0.0001	3.864	0.039	34.785	<0.0001	0.994	<0.0001
COMP	0.087	0.052	-47.228	<0.0001	3.275	0.033	36.110	<0.0001	0.995	<0.0001
EARTH	0.045	0.014	-220.404	<0.0001	3.484	0.009	139.632	<0.0001	1.000	<0.0001
ENER	0.063	0.080	-34.689	<0.0001	3.528	0.051	24.937	<0.0001	0.989	<0.0001
ENG	0.089	0.034	-71.318	<0.0001	3.215	0.022	54.295	<0.0001	0.998	<0.0001
ENVIR	0.055	0.063	-46.202	<0.0001	3.511	0.040	31.461	<0.0001	0.993	<0.0001
IMMU	0.058	0.081	-35.353	<0.0001	3.972	0.051	26.962	<0.0001	0.990	<0.0001
MATER	0.048	0.069	-44.115	<0.0001	4.114	0.044	32.415	<0.0001	0.993	<0.0001
MATH	0.040	0.088	-36.655	<0.0001	4.033	0.056	25.034	<0.0001	0.989	<0.0001
MED	0.062	0.052	-53.879	<0.0001	3.298	0.033	36.440	<0.0001	0.995	<0.0001
NEURO	0.043	0.093	-34.036	<0.0001	4.268	0.059	24.658	<0.0001	0.989	<0.0001
PHARM	0.099	0.050	-46.023	<0.0001	3.357	0.032	37.980	<0.0001	0.995	<0.0001
PHYS	0.056	0.066	-43.525	<0.0001	3.721	0.042	31.295	<0.0001	0.993	<0.0001
TOTAL	0.059	0.053	-52.889	<0.0001	3.495	0.034	36.948	<0.0001	0.995	<0.0001

Supplementary Table 16: Survival regression model (exponential method) for the 2010 cohort, men, abbreviated data including exp(B) for the time variable, R^2 metrics and p-value for the model in terms of disciplines

Discipline	**time – expB**	**Measures - R^2**	**Measures - pval**
AGRI	3.474	0.996	<0.0001
BIO	3.734	0.994	<0.0001
CHEMENG	3.374	0.992	<0.0001
CHEM	3.838	0.993	<0.0001
COMP	3.407	0.997	<0.0001
EARTH	3.549	0.999	<0.0001
ENER	3.489	0.999	<0.0001
ENG	3.208	0.995	<0.0001
ENVIR	3.155	0.999	<0.0001
IMMU	3.261	0.997	<0.0001
MATER	3.551	0.990	<0.0001
MATH	3.629	0.993	<0.0001
MED	3.190	0.995	<0.0001
NEURO	3.796	0.992	<0.0001
PHARM	3.005	0.997	<0.0001
PHYS	3.647	0.990	<0.0001
TOTAL	3.364	0.995	<0.0001

Supplementary Table 17: Survival regression model (exponential method) for the 2010 cohort, men, full data

Discipline	**Intercept - expB**	**Intercept - std_err**	**Intercept - tvalue**	**Intercept - pvalue**	**time - expB**	**time - std_err**	**time - tvalue**	**time - pvalue**	**Measures - R^2**	**Measures - pval**
AGRI	0.044	0.048	-65.679	<0.0001	3.474	0.030	41.290	<0.0001	0.996	<0.0001
BIO	0.047	0.061	-49.987	<0.0001	3.734	0.039	33.890	<0.0001	0.994	<0.0001
CHEMENG	0.093	0.065	-36.302	<0.0001	3.374	0.041	29.321	<0.0001	0.992	<0.0001
CHEM	0.061	0.067	-41.612	<0.0001	3.838	0.043	31.494	<0.0001	0.993	<0.0001
COMP	0.079	0.043	-58.808	<0.0001	3.407	0.027	44.694	<0.0001	0.997	<0.0001
EARTH	0.040	0.023	-139.032	<0.0001	3.549	0.015	85.970	<0.0001	0.999	<0.0001
ENER	0.062	0.027	-102.258	<0.0001	3.489	0.017	72.486	<0.0001	0.999	<0.0001
ENG	0.089	0.051	-47.764	<0.0001	3.208	0.032	36.183	<0.0001	0.995	<0.0001
ENVIR	0.062	0.025	-112.577	<0.0001	3.155	0.016	73.412	<0.0001	0.999	<0.0001
IMMU	0.072	0.036	-72.619	<0.0001	3.261	0.023	51.506	<0.0001	0.997	<0.0001
MATER	0.065	0.075	-36.465	<0.0001	3.551	0.048	26.657	<0.0001	0.990	<0.0001
MATH	0.043	0.062	-50.498	<0.0001	3.629	0.040	32.558	<0.0001	0.993	<0.0001
MED	0.060	0.049	-57.247	<0.0001	3.190	0.031	37.111	<0.0001	0.995	<0.0001
NEURO	0.041	0.069	-46.051	<0.0001	3.796	0.044	30.297	<0.0001	0.992	<0.0001
PHARM	0.099	0.033	-69.779	<0.0001	3.005	0.021	52.325	<0.0001	0.997	<0.0001
PHYS	0.057	0.076	-37.956	<0.0001	3.647	0.048	26.938	<0.0001	0.990	<0.0001
TOTAL	0.062	0.053	-52.496	<0.0001	3.364	0.034	36.180	<0.0001	0.995	<0.0001

In order to check the combined multidimensional influence of dependent variables in the model, an analysis of inverse correlation matrices was performed and main diagonals were analyzed: none of the variables in any of the models is characterized by significantly larger values than the others; no collinearity is reported (see Table ST 18 for cohort 2000 and Table ST 19 for cohort 2010).

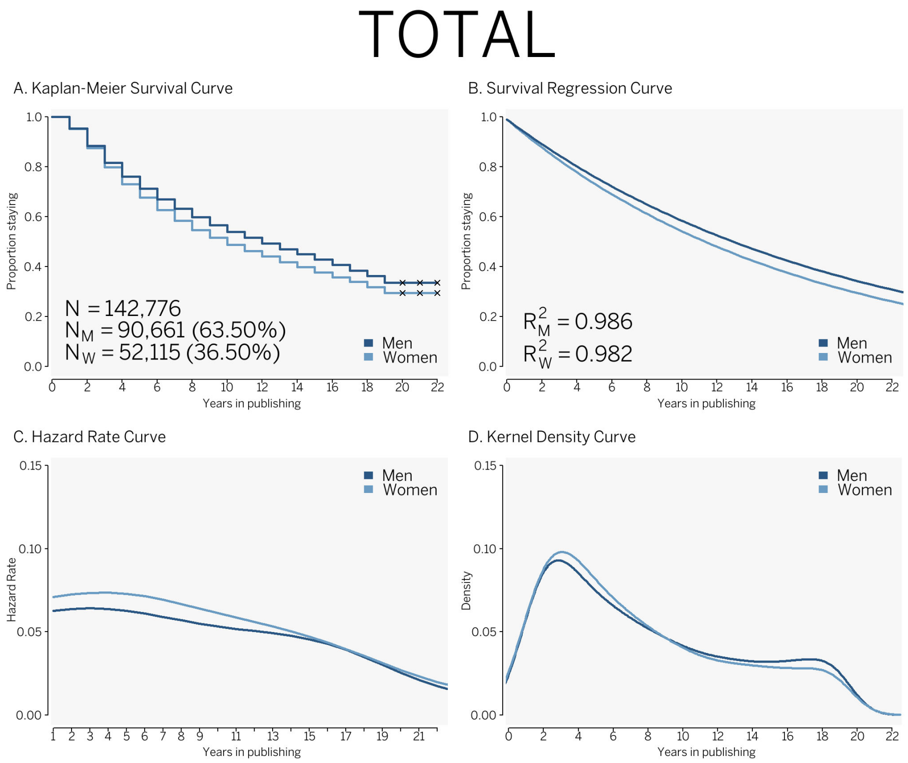
Supplementary Table 18. Inverse correlation matrix main diagonals in the models' independent variables. Models for cohort 2000

Variable / Discipline	**Gender**	**FWCI 4y**	**International Collaboration Rate (Lifetime)**	**Scholarly Output (Lifetime)**	**Average Publication Journal Percentile (Lifetime)**	**Median Team Size (Lifetime)**	**TOP200**	**Country**
AGRI	1.021	1.254	1.211	1.132	1.249	1.131	1.032	1.039
BIO	1.025	1.143	1.175	1.085	1.152	1.171	1.019	1.047
CHEM	1.006	1.084	1.128	1.072	1.117	1.081	1.026	1.016
CHEMENG	1.011	1.025	1.049	1.057	1.071	1.048	1.042	1.020
COMP	1.002	1.129	1.141	1.119	1.099	1.033	1.026	1.051
EARTH	1.020	1.075	1.335	1.130	1.178	1.197	1.018	1.034
ENER	1.036	1.103	1.188	1.136	1.326	1.174	1.047	1.157
ENG	1.008	1.049	1.064	1.073	1.082	1.037	1.037	1.032
ENVIR	1.015	1.131	1.195	1.108	1.112	1.112	1.026	1.038
IMMU	1.016	1.207	1.171	1.077	1.150	1.189	1.026	1.053
MATER	1.012	1.124	1.121	1.111	1.148	1.109	1.047	1.032
MATH	1.012	1.178	1.103	1.159	1.077	1.021	1.045	1.022
MED	1.018	1.084	1.161	1.082	1.125	1.121	1.038	1.043
NEURO	1.025	1.152	1.147	1.093	1.154	1.144	1.020	1.070
PHARM	1.007	1.078	1.101	1.025	1.084	1.088	1.037	1.031
PHYS	1.006	1.204	1.418	1.216	1.162	1.334	1.027	1.018

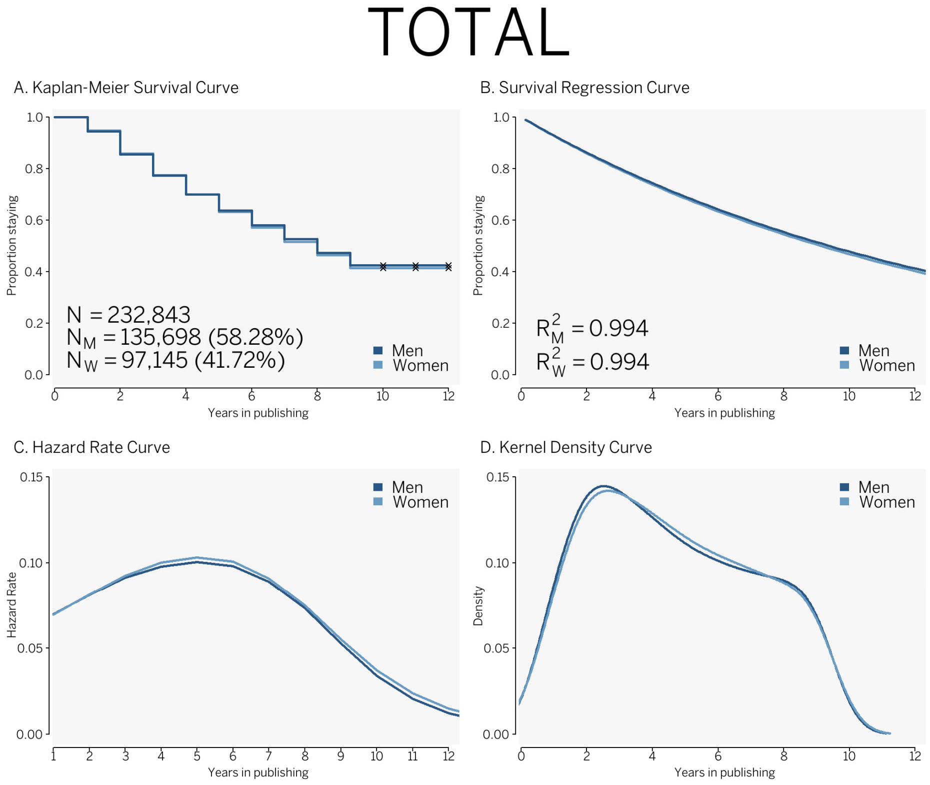
Supplementary Table 19. Inverse correlation matrix main diagonals in the models' independent variables. Models for cohort 2010

Variable / Discipline	**Gender**	**FWCI 4y**	**International Collaboration Rate (Lifetime)**	**Scholarly Output (Lifetime)**	**Average Publication Journal Percentile (Lifetime)**	**Median Team Size (Lifetime)**	**TOP200**	**Country**
AGRI	1.021	1.312	1.173	1.069	1.328	1.139	1.042	1.045
BIO	1.017	1.114	1.204	1.051	1.178	1.206	1.026	1.047
CHEM	1.010	1.133	1.141	1.077	1.178	1.109	1.050	1.023
CHEMENG	1.020	1.018	1.059	1.054	1.114	1.062	1.054	1.037
COMP	1.003	1.056	1.104	1.095	1.105	1.058	1.039	1.035
EARTH	1.027	1.143	1.366	1.108	1.210	1.267	1.029	1.038
ENER	1.007	1.049	1.073	1.096	1.201	1.090	1.053	1.115
ENG	1.004	1.114	1.066	1.090	1.149	1.080	1.056	1.026
ENVIR	1.009	1.166	1.196	1.101	1.175	1.101	1.031	1.031
IMMU	1.013	1.176	1.189	1.076	1.184	1.218	1.031	1.046
MATER	1.008	1.217	1.142	1.100	1.201	1.169	1.059	1.031
MATH	1.018	1.054	1.095	1.086	1.083	1.024	1.031	1.019
MED	1.016	1.060	1.163	1.055	1.201	1.164	1.041	1.053
NEURO	1.027	1.247	1.165	1.069	1.218	1.159	1.024	1.070
PHARM	1.013	1.095	1.122	1.024	1.142	1.126	1.030	1.021
PHYS	1.009	1.226	1.387	1.185	1.156	1.294	1.028	1.017

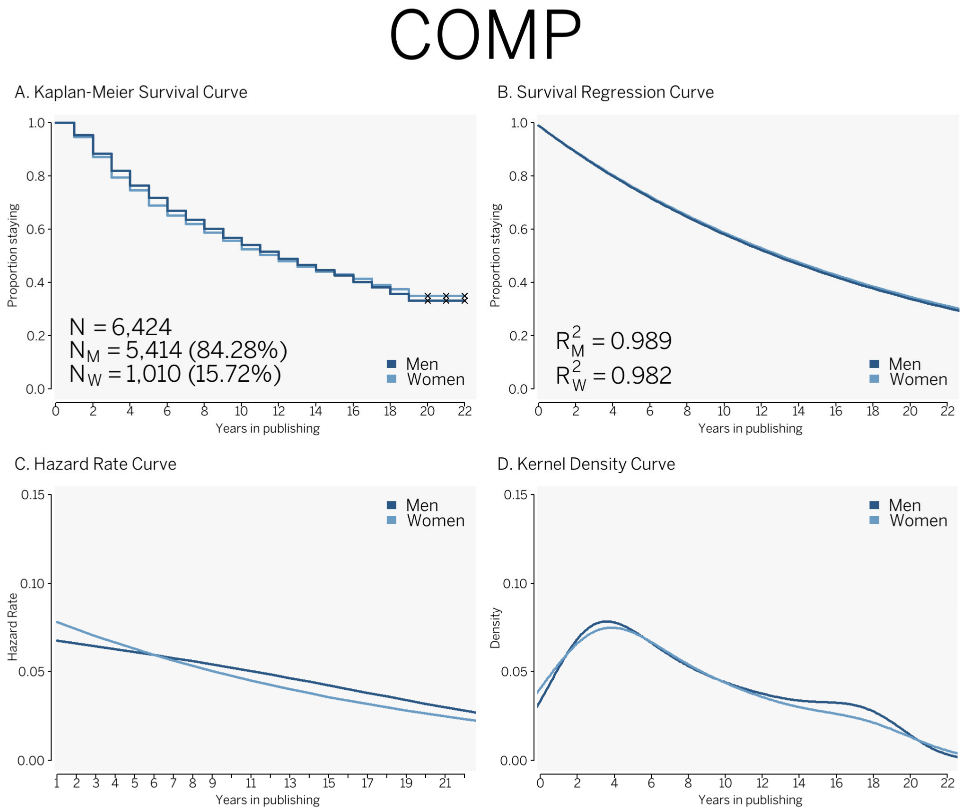
S4. Supplementary Figures
The 2000 and 2010 cohorts compared



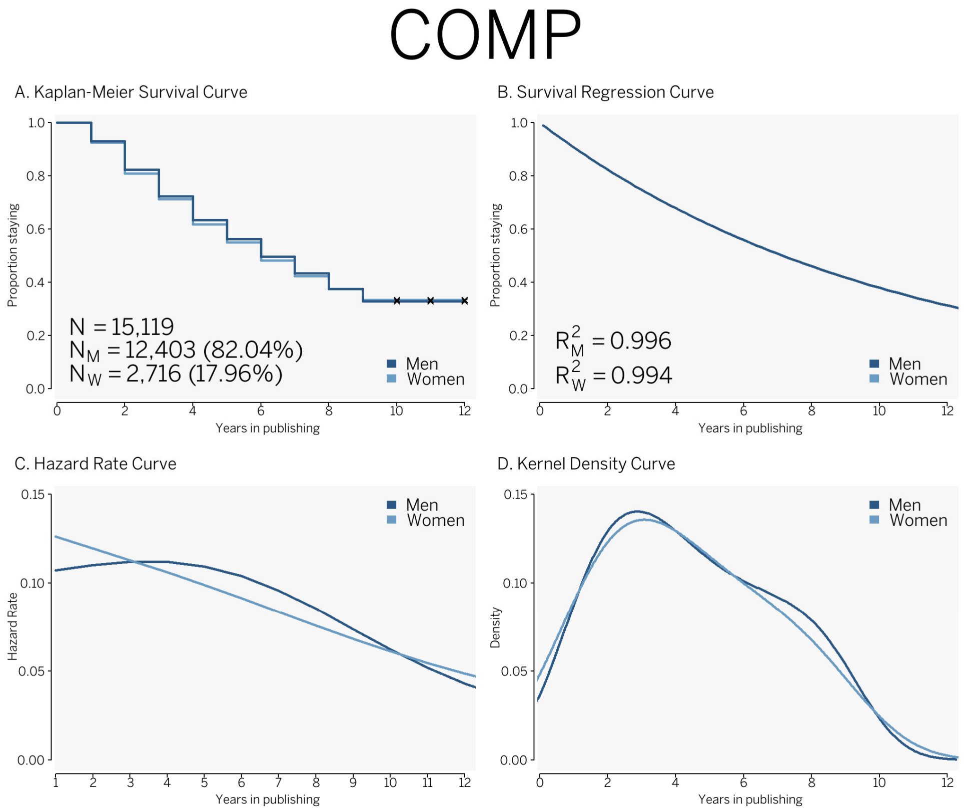
Supplementary Figure 1: Kaplan-Meier Curve, Survival Regression Curve, Hazard Rate Curve and Kernel Density Curve, all disciplines combined, scientists from the 2000 cohort



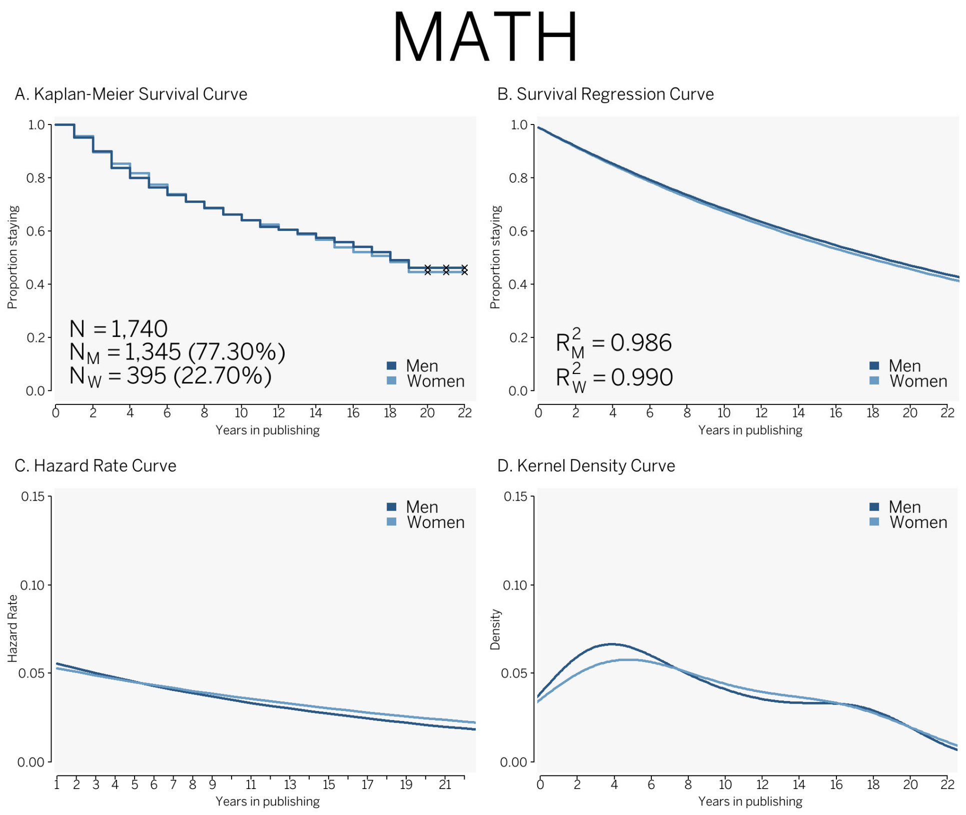
Supplementary Figure 2: Kaplan-Meier Curve, Survival Regression Curve, Hazard Rate Curve and Kernel Density Curve, all disciplines combined, scientists from the 2010 cohort (N=232,843)



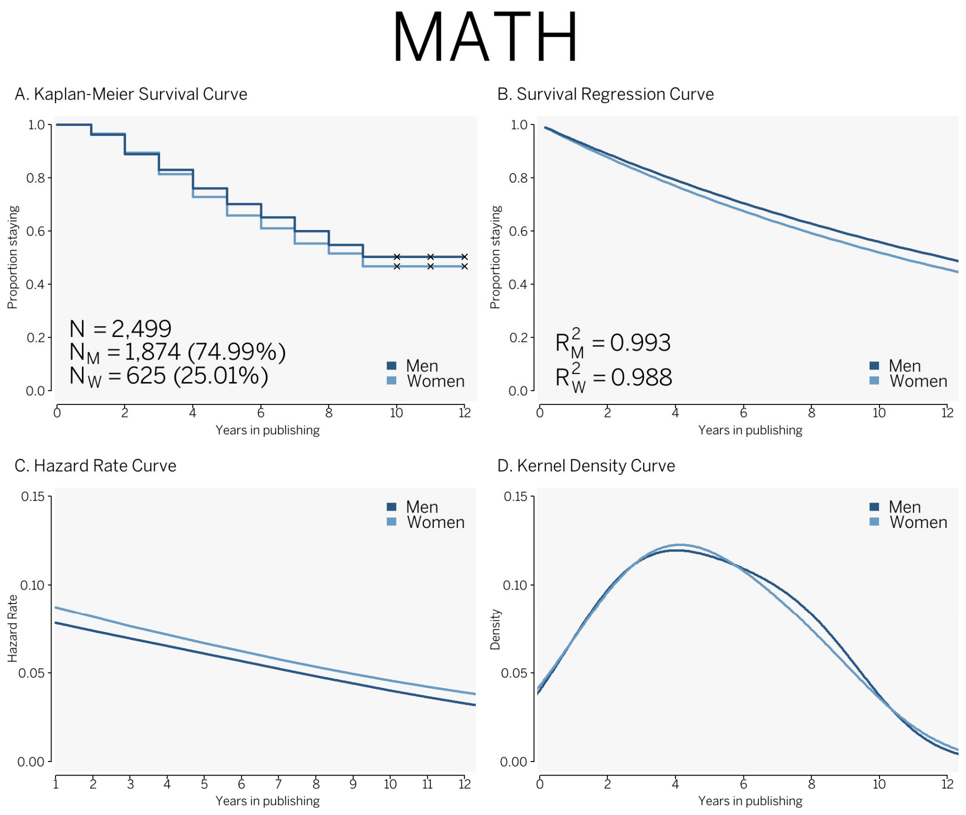
Supplementary Figure 3: Kaplan-Meier Curve, Survival Regression Curve, Hazard Rate Curve and Kernel Density Curve, COMP, scientists from the 2000 cohort (N= 6,424)



Supplementary Figure 4: Kaplan-Meier Curve, Survival Regression Curve, Hazard Rate Curve and Kernel Density Curve, COMP, scientists from the 2010 cohort (N=15,119)



Supplementary Figure 5: Kaplan-Meier Curve, Survival Regression Curve, Hazard Rate Curve and Kernel Density Curve, MATH, scientists from the 2000 cohort (N= 1,740)



Supplementary Figure 6: Kaplan-Meier Curve, Survival Regression Curve, Hazard Rate Curve and Kernel Density Curve, MATH, scientists from the 2010 cohort (N=2,499)

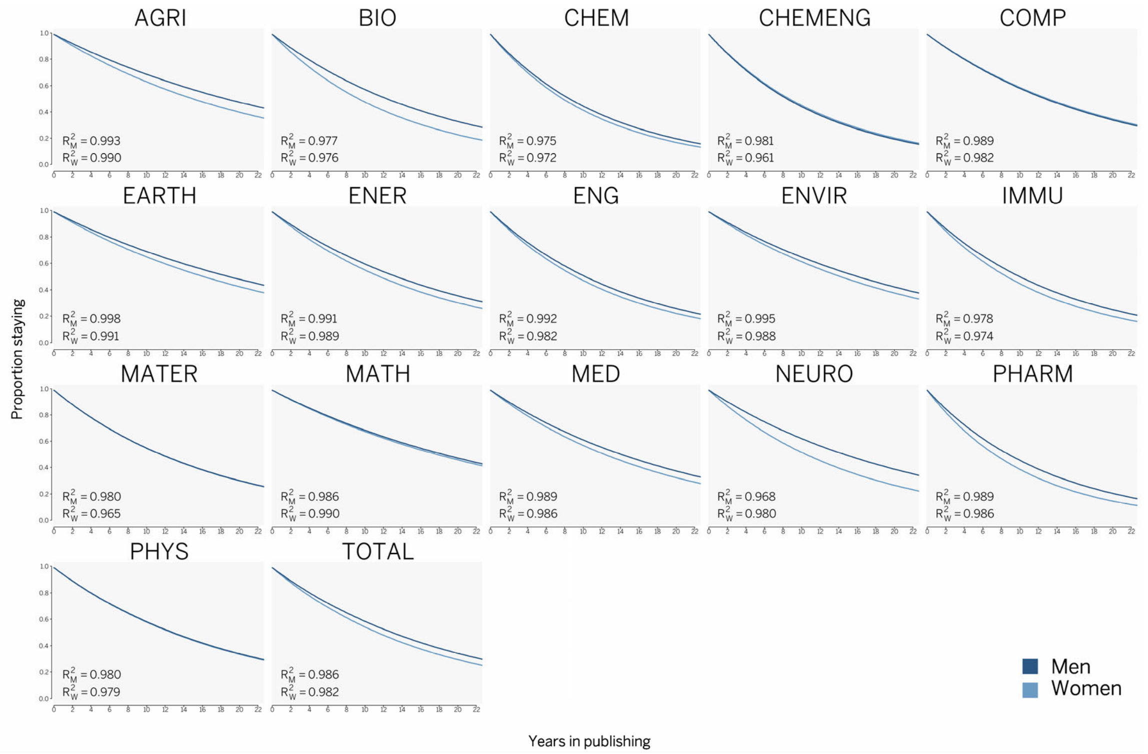
The 2000 and 2010 cohorts compared: Other survival perspectives (disciplinary variation)

In some disciplines, the changes are much more fundamental than in others. We restrict our cursory view first to the same disciplines selected for cohort 2000 (BIO and PHYS) (see Supplementary Figures 7 through 12). In BIO, the differences in the Kaplan–Meier survival curves and in the survival regression rate for men and women are considerably smaller for the 2010 cohort than for the 2000 cohort; both hazard rate curves and kernel density curves for men and women show similar differences, albeit with different intensities over the first 10 years. PHYS, in contrast, remains a discipline with minor (or even none at all) gender differences in attrition and with women showing a slightly higher probability of leaving science in early and late years than men. Kernel density distribution shows that men are much more likely to leave science than women in years 2–4, that is, very early in their careers.

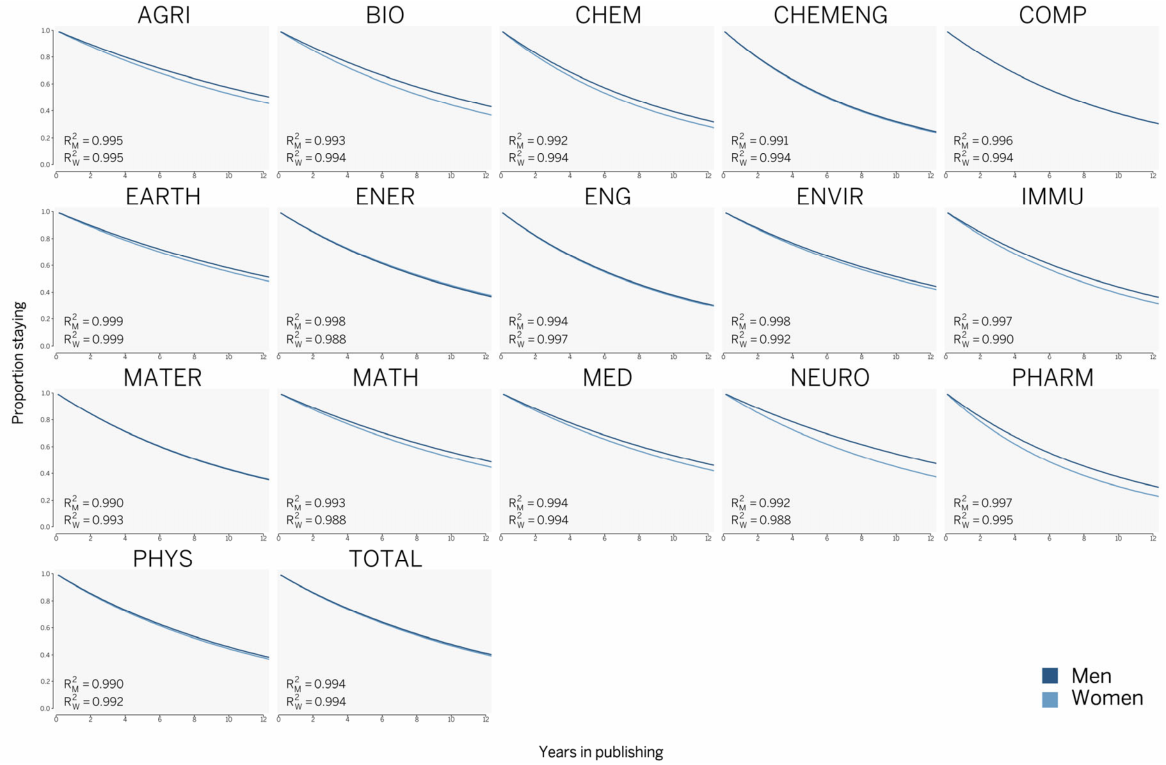
First, the survival regression curves for the two cohorts mirror the Kaplan–Meier curves for the two cohorts but are smoothed. For those disciplines for which there were differences between men and women for the 2000 cohort, these differences continue for the 2010 cohort but are smaller (see, e.g., the large disciplines of MED, BIO, and AGRI). For those disciplines where the differences were invisible—such as COMP and PHYS—they continue to be invisible. Overall, the attrition gaps between men and women have been closing across the board so that, for all disciplines combined, they are invisible.

Second, the hazard function curves by discipline for the two cohorts tend to confirm the findings based on Kaplan–Meier curves and survival regression curves. The chances to leave science are much less gendered for the 2010 cohort than for the 2000 cohort for most disciplines, especially in the early years of publishing careers. The differences between men and women are especially stark in the 2010 cohort for BIO, NEURO, and PHARM (as in the case of the 2000 cohort). However, for the three math-intensive disciplines of COMP, MATH, and PHYS, the generally overlapping curves for the 2000 cohort are more volatile for the 2010 cohort over time and still very similar for men and women.

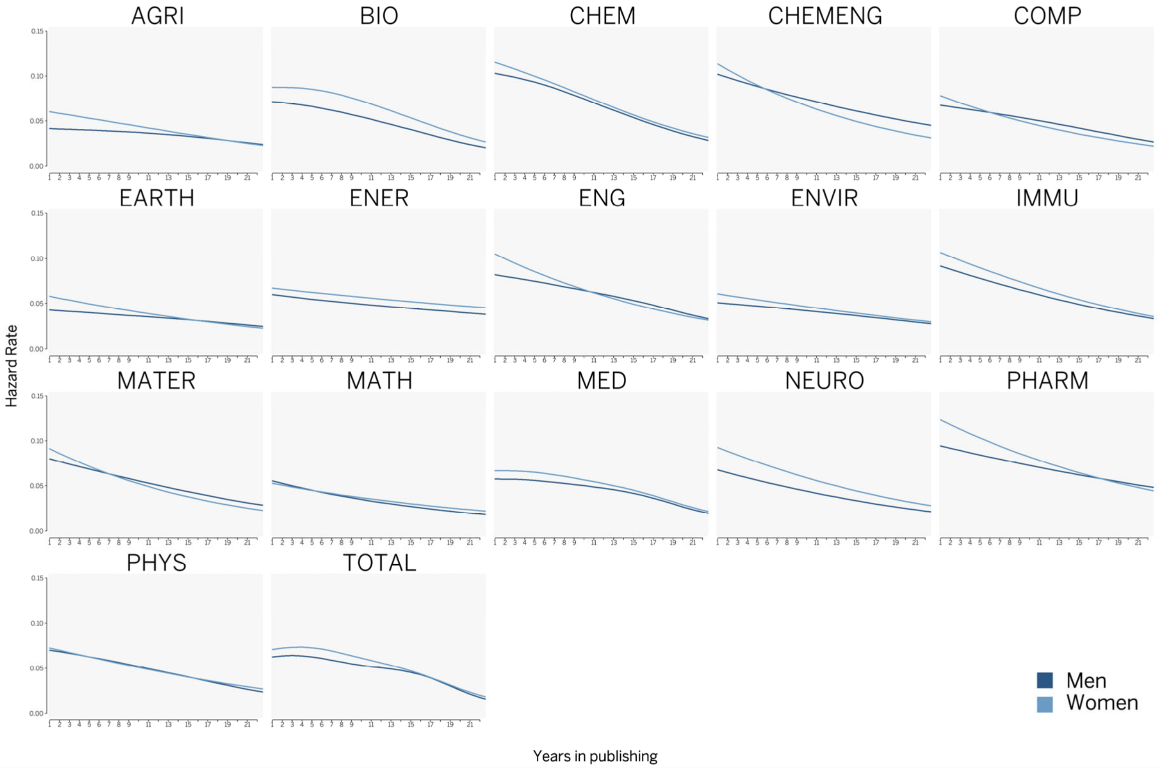
Finally, the kernel density curves for the two cohorts tell largely the same story; specifically, should the 2010 cohort be observed for another decade (i.e., in 2032), the distributions of those who leave science within the two cohorts may be much more similar. The early years (years 2–6) are those when large proportions of scientists leave, but gender differences in attrition for the 2010 cohort are considerably lower than for the 2000 cohort. The distribution for all disciplines combined is almost identical for men and women, and in several disciplines, more men leave early than women (e.g., in CHEMENG, COMP, MATER, and PHYS). Overall, the patterns of distribution of the leavers for the 2010 cohort are much less gendered than for the 2000 cohort, testifying to the observation that, over time, the differences in attrition in science between men and women are ever less pronounced.



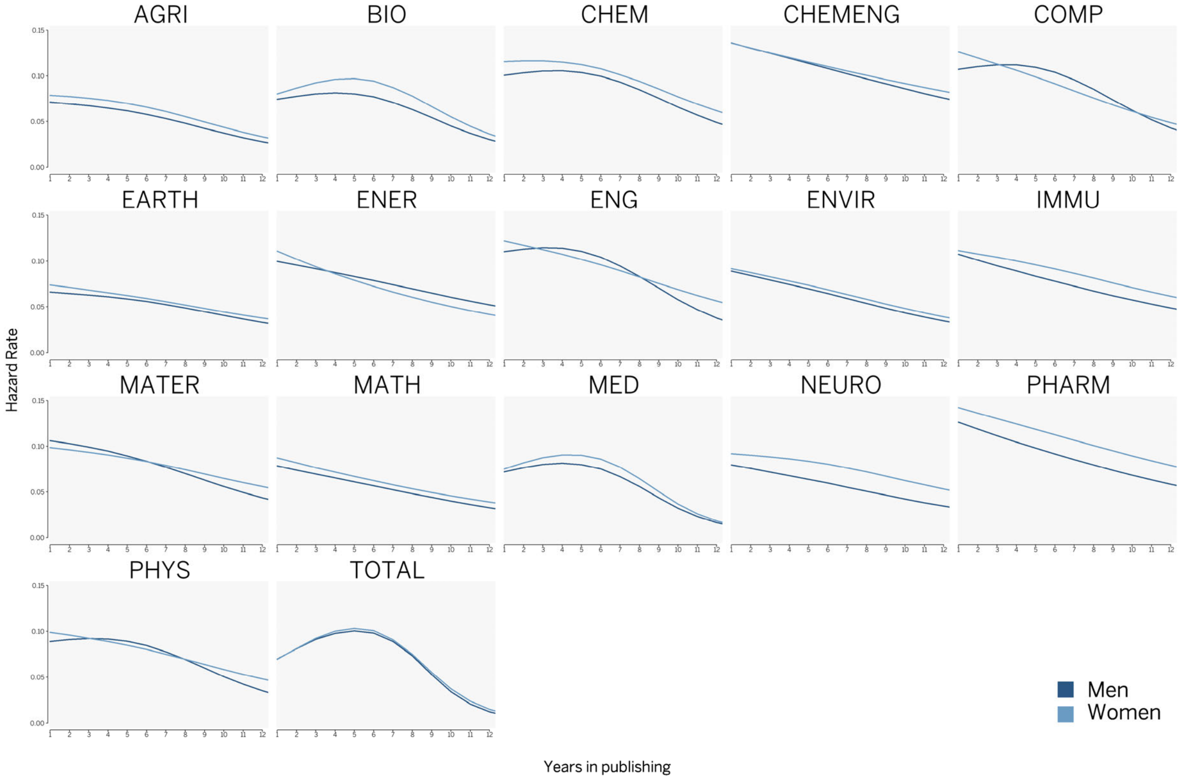
Supplementary Figure 7: Survival regression curve by discipline and gender, scientists from the 2000 cohort (N=142,776)



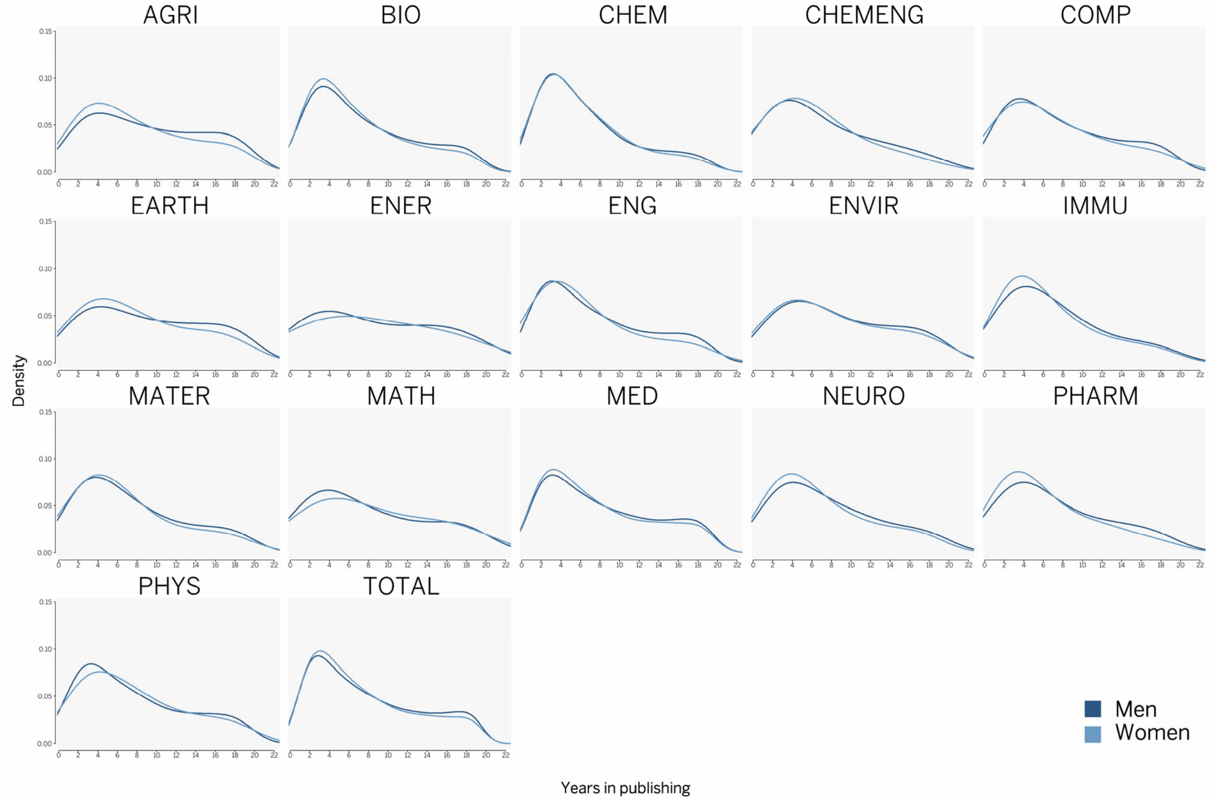
Supplementary Figure 8: Survival regression curve by discipline and gender, scientists from the 2010 cohort (N=232,843)



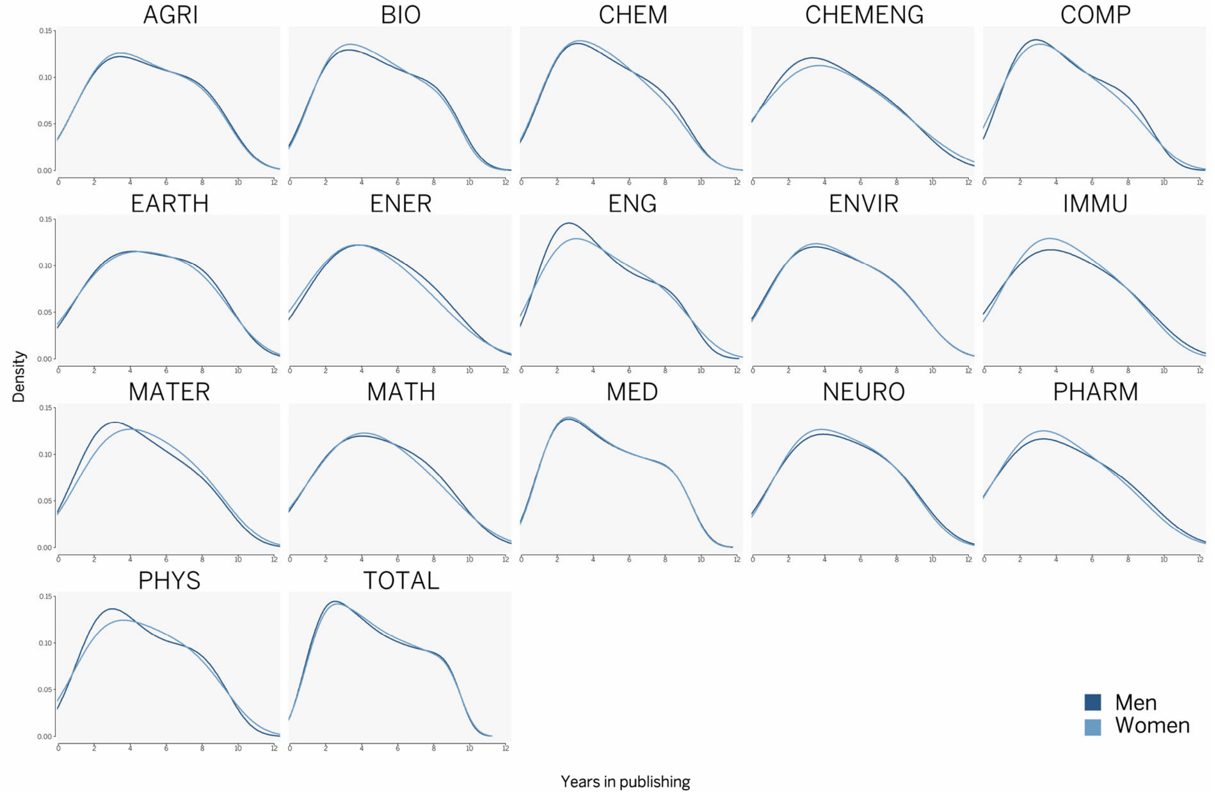
Supplementary Figure 9: Hazard function curve by discipline and gender, B-splines smoothing method, smoothing parameter 10k, scientists from the 2000 cohort (N=142,776)



Supplementary Figure 10: Hazard function curve by discipline and gender, B-splines smoothing method, smoothing parameter 10k, scientists from the 2010 cohort (N=232,843)

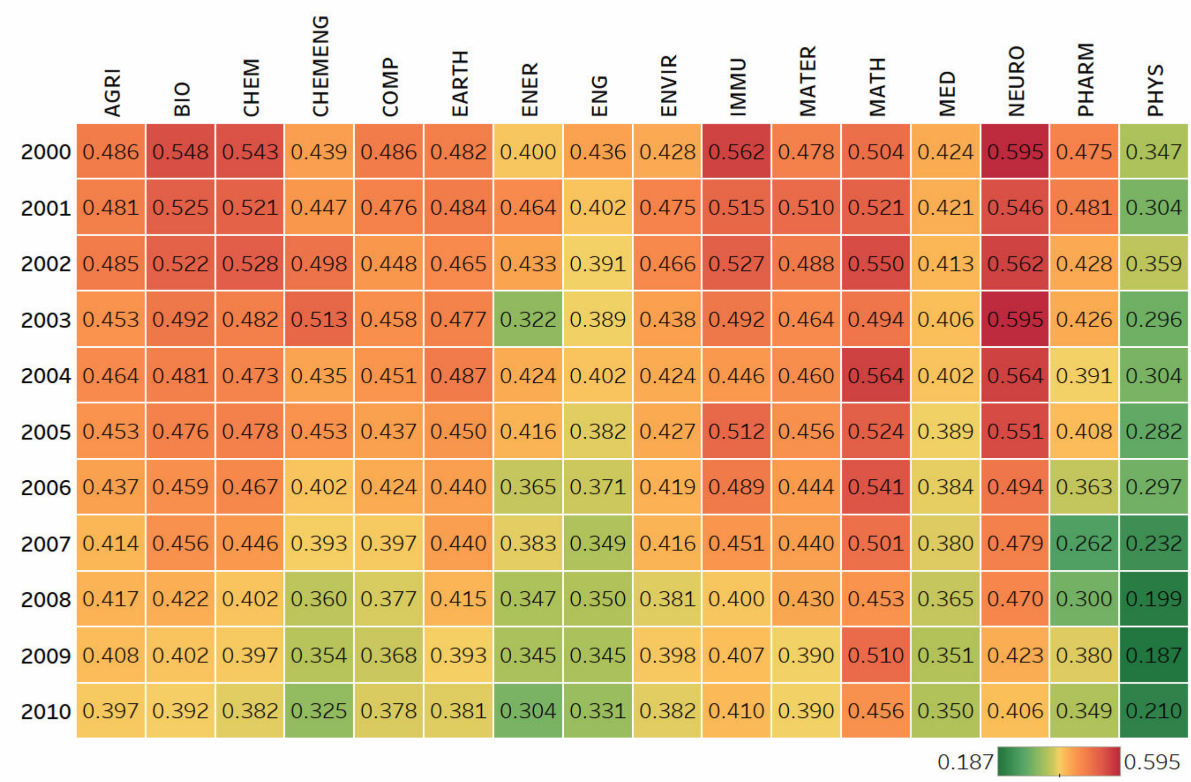


Supplementary Figure 11: Kernel density curve by discipline and gender, B-splines smoothing method, bandwidth 2, component per point based on Gaussian curve, scientists from the 2000 cohort (N=142,776)



Supplementary Figure 12: Kernel density curve by discipline, B-splines smoothing method, bandwidth 2, component per point based on Gaussian curve, scientists from the 2010 cohort (N=232,843)

Interestingly, the general model of staying in science we have constructed using mostly micro-level variables based on bibliometric data fits the phenomenon under exploration relatively well. We use pseudo-R^2 statistics because R^2 statistics are not appropriate for use with the logistic model. We used McFadden pseudo-R-squared, which represents the proportion of the variance for a dependent variable explained by independent variables (71). Although R^2 has been generally declining for each successive cohort, for the oldest cohort, in all disciplines, R^2 is in the 0.40–0.60 range—more than a half of the observed variation can be explained by inputs to our models for BIO, IMMU, MATH, and NEURO; and for the youngest cohort, although lower, R^2 is still in the range of 0.30–045 (except for PHYS), and it is about 0.40 for nine disciplines.



	AGRI	BIO	CHEM	CHEMENG	COMP	EARTH	ENER	ENG	ENVIR	IMMU	MATER	MATH	MED	NEURO	PHARM	PHYS
2000	0.486	0.548	0.543	0.439	0.486	0.482	0.400	0.436	0.428	0.562	0.478	0.504	0.424	0.595	0.475	0.347
2001	0.481	0.525	0.521	0.447	0.476	0.484	0.464	0.402	0.475	0.515	0.510	0.521	0.421	0.546	0.481	0.304
2002	0.485	0.522	0.528	0.498	0.448	0.465	0.433	0.391	0.466	0.527	0.488	0.550	0.413	0.562	0.428	0.359
2003	0.453	0.492	0.482	0.513	0.458	0.477	0.322	0.389	0.438	0.492	0.464	0.494	0.406	0.595	0.426	0.296
2004	0.464	0.481	0.473	0.435	0.451	0.487	0.424	0.402	0.424	0.446	0.460	0.564	0.402	0.564	0.391	0.304
2005	0.453	0.476	0.478	0.453	0.437	0.450	0.416	0.382	0.427	0.512	0.456	0.524	0.389	0.551	0.408	0.282
2006	0.437	0.459	0.467	0.402	0.424	0.440	0.365	0.371	0.419	0.489	0.444	0.541	0.384	0.494	0.363	0.297
2007	0.414	0.456	0.446	0.393	0.397	0.440	0.383	0.349	0.416	0.451	0.440	0.501	0.380	0.479	0.262	0.232
2008	0.417	0.422	0.402	0.360	0.377	0.415	0.347	0.350	0.381	0.400	0.430	0.453	0.365	0.470	0.300	0.199
2009	0.408	0.402	0.397	0.354	0.368	0.393	0.345	0.345	0.398	0.407	0.390	0.510	0.351	0.423	0.380	0.187
2010	0.397	0.392	0.382	0.325	0.378	0.381	0.304	0.331	0.382	0.410	0.390	0.456	0.350	0.406	0.349	0.210

Supplementary Figure 13: Comparison of R^2 values for 11 cohorts (2000 to 2010) for the model of odds ratio estimates of staying in science from the starting year of the cohort up to and including 2019 by discipline (11 cohorts times 16 disciplines = 176 models)