The Changing Demographics of Scientific Careers: The Rise of the Temporary Workforce

Staša Milojević[1], Filippo Radicchi[1], John Walsh[2]

[1]Indiana University, [2]Georgia Institute of Technology

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Contemporary science has been characterized by an exponential growth in publications and a rise of team science. At the same time, there has been an increase in the number of awarded Ph.D. degrees, which has not been accompanied by a similar expansion in the number of academic positions. In such a competitive environment, an important measure of academic success is the ability to maintain a long active career in science. In this paper, we study workforce trends in three scientific disciplines over half a century. We find dramatic shortening of careers of scientists across all three disciplines. The time over which half of the cohort has left the field has shortened from 35 years in the 1960s to only 5 years in 2010s. In addition, we find a rapid rise (from 25% to 60% since 1960s) of a group of scientists that spend their entire career only as supporting authors without having led a publication. Altogether, the fraction of entering researchers who achieve full careers has diminished, while the class of temporary scientists has escalated. We provide an interpretation of our empirical results in terms of a survival model from which we infer potential factors of success in scientific career survivability. Cohort attrition can be successfully modeled by relatively simple hazard probability function. Although we find statistically significant trends between survivability and author's early productivity, neither it nor the citation impact of early work or the level of initial collaboration can serve as reliable predictors of ultimate survivability.

scientific workforce | scientific careers | career success

Contemporary science has been characterized by an exponential growth in practitioners and publications (1) and a rise of team science, both in terms of the increasing prevalence of team-authored work and the growth of team sizes (2-4). The gradual shift from individual to team science is driven by a variety of factors, including increasing capital intensivity of science (5) and the increased need for technicians and staff scientists (6). At the same time, in recent decades there has been a substantial growth in the number of awarded Ph.D. degrees (7), which has not been accompanied by a similar growth in the number of academic positions (8), leading to a growing concern over the lack of opportunities for new PhDs in science (9, 10) and even warnings regarding possible scientific workforce bubbles (11, 12). In an environment with substantial growth in PhDs granted and only a modest growth in the number of faculty positions, the idea of each professor regularly reproducing herself in each cohort of graduate students becomes untenable. These and similar data have led to calls for rethinking academic careers, and to discussions of the need for policy interventions to address this growing problem (5, 9, 13).

How does the shifting landscape of science over the past half century affect the roles of new researchers and their overall careers? There is an abundance of studies that focus on the criteria that may affect researchers' success in terms of impact of their work, especially in terms of citations to publications. However, another, and perhaps more fundamental aspect of success is the ability to perform research over the full extent of someone's career, rather than to leave the field prematurely. A smaller fraction of literature focuses on understanding the factors leading to successful academic careers in this broader sense, and, more recently, the factors contributing to abandoning scientific careers (14-16). Prior work has identified productivity (14, 16-20), impact (20, 21), number of collaborators (14, 17), gender (22), prestige of PhD granting and hiring institutions (23, 24), prestige of the advisors (24, 25), gender of the advisors (16) and the level of specialization (26) as important factors correlated with career success. Some of these studies have found that these factors are correlated. For example, there is a correlation between citation success of early papers and later increase in productivity (27). There is also a reported correlation between gender and productivity (19, 28, 29), gender and citations, and gender and collaboration. Finally, there is a correlation between institutional prestige and productivity (30, 31) as well as institutional prestige and impact (32). Directionality of these correlation is difficult to establish and is not the focus of this paper.

On the other hand, there are relatively few studies that focus on modeling scientific careers (30, 33-38) in the context of survivability. An early study of this type (35) used a small sample (500 authors) during a seven-year period (1964-1970) and has established a division of all authors into transient and continuants and found that the levels of productivity are correlated with career length. Two recent studies (36, 37) used survival analysis and hazard models to examine gender differences in retention of science and social science assistant professors. These studies established that the chances of survival of assistant professors in science and engineering are less than 50% and that the "median time to departure is 10.9 years" (36) and that, in social sciences, "half of all entering faculty have departed by year 9" (37).

Despite various efforts there is a clear gap in our knowledge of careers of scientific workforce in general (and not only tenure-track scientists). Furthermore, large-scale investigation of the trends in careers of scientific workforce across entire disciplines and over long periods of time (many decades) is still in its infancy.

In this study we analyze the changing careers of the scientific workforce of entire disciplines without making assumptions regarding the position they have in the scientific community (i.e., not limiting to those who have tenure track jobs, as was the case in many of the earlier studies). We specifically focus on the role that different authors play in knowledge production. Furthermore, we investigate whether one can identify early factors (during a researcher's apprenticeship phase) that would indicate a scientist's ability to maintain a research active career over many years. Our big-data approach is facilitated by an extensive

Reserved for Publication Footnotes

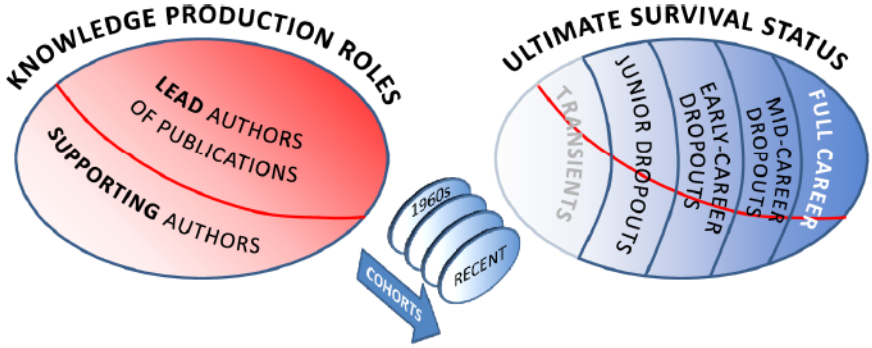

Fig. 1. Model of scientific careers. For each cohort of authors entering the field we determine the knowledge production role as a "lead author" (researcher who *at any time* in their careers leads a production of a scientific publication) or a "supporting author" (those who will have never lead a publication). Furthermore, each new author will fall into one of five categories of ultimate career status: transients (authors who only had one publication), dropouts (authors who leave the field prematurely at different levels of their careers) and full-career scientists (authors who ultimately survive in the field). In each survival category there will be some authors classified as lead and some as supporting (the repeated red curve). We follow fifty cohorts starting from the 1960s.

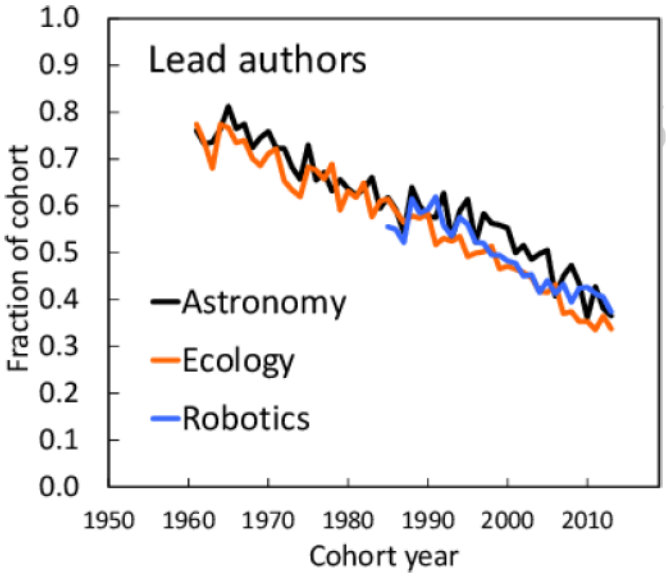

Fig. 2. The fraction of each cohort who contribute or will contribute to the knowledge production as lead authors. The status of lead author means that the author has led a publication at any time in his/her career. An increasing fraction of entering authors never acquire the lead role but participate in knowledge production solely as supporting scientists.

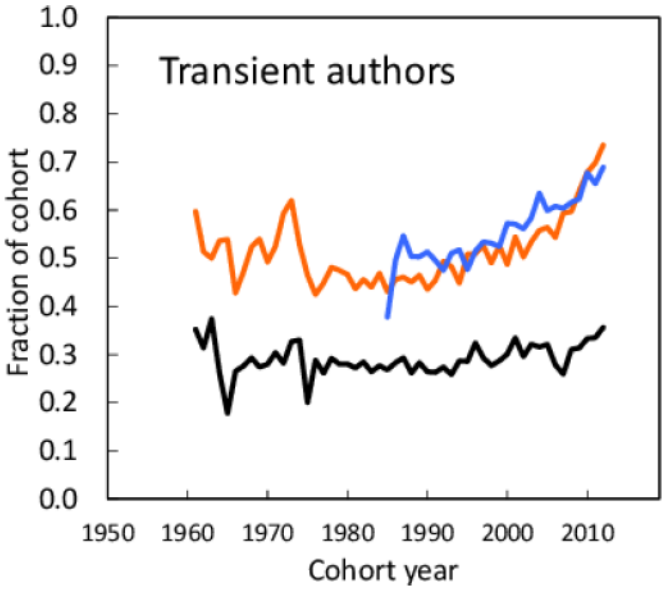

Fig. 3. The fraction of each cohort who has published only one paper (transient authors). The share of transients has increased in the last two decades, especially in ecology and robotics. The trend in the fraction of authors who are lead authors (Figure 1) remain similar when transients are excluded from cohorts.

longitudinal dataset containing millions of bibliographic items covering the entire period of contemporary science.

In order to capture the above-stated changes in the demographics of the scientific workforce, we created a survival model of authors based both on the primary role they play in the production of knowledge and their ultimate survival status in science. Each author is placed in one of two categories based on their primary authorship roles: lead authors and supporting authors. Lead authors are all authors who have lead a publication *at any time* in their career, whereas the supporting authors are the ones who have never that role in their career. Furthermore, we place each author (whether lead or supporting) into one of the five categories in terms of their ultimate survival status: transients (authors with a single publication), junior dropouts (multi-paper authors leaving after 0-10 years after the first publication), early-career dropouts (multi-paper authors leaving after 11-15 years after the first publication), mid-career dropouts (multi-paper authors leaving after 16-20 years after the first publication), and full career scientists (multi-paper authors who have careers longer than 20 years). This classification is presented schematically in Figure 1. The balance between supporting and lead authors in each of the survival categories is different, as indicated by the tilted curve in Figure 1. Most transient scientists belong to the supporting author group, whereas, as we move towards the full-career status, the proportion shifts in favor of lead authors. In order to study the changing landscape of scientific careers in terms of knowledge production roles and survivability, we focus our analysis on cohorts – a group of authors who first appear on the scientific stage at the same time (in the same year). Our study is facilitated by the availability of extensive longitudinal data allowing us to follow up half a century of cohorts and to assess their eventual careers.

In this study we focus on researcher cohorts in three scientific disciplines covering different areas of science: astronomy (physical sciences), ecology (life sciences) and robotics (engineering and computer science). We focus on researchers who have published in principal journals belonging to these fields (listed in *SI Appendix*). These are the journals that are well established, usually publish a large fraction of original research in

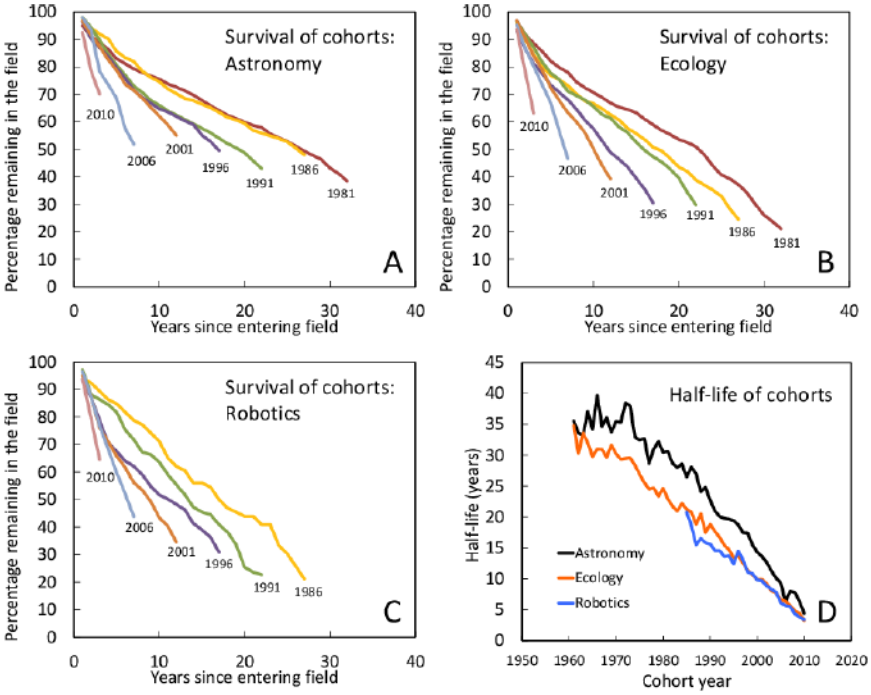

Fig. 4. Survival functions of select cohorts in three fields (panels A, B and C), and the half-life (time needed for half of the cohort to abandon the field) of all cohorts from all three fields (panel D). The decline in survivability over the last half century has been remarkable.

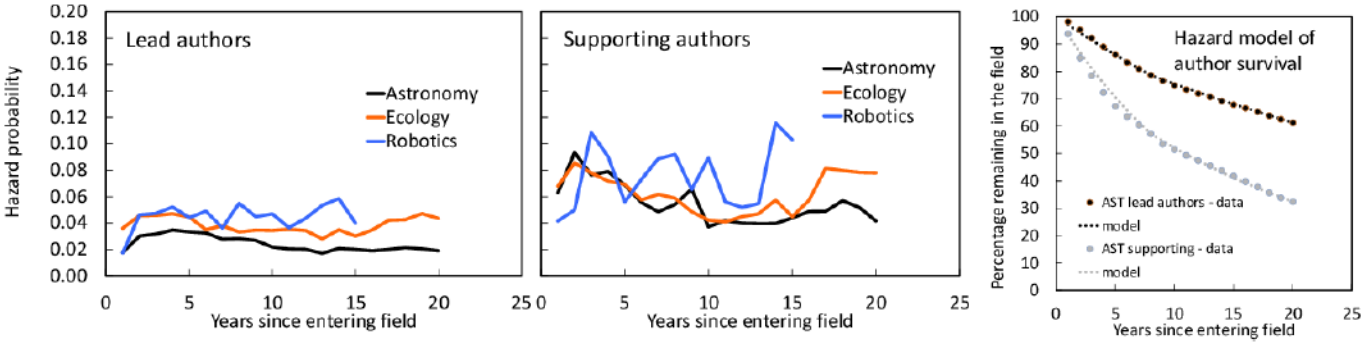

Fig. 5. Hazard model. Hazard probabilities for lead authors (panel A) and supporting authors (panel B). The comparison of linear + constant hazard model (dotted lines) with the empirical survival functions in the field of astronomy (panel C).

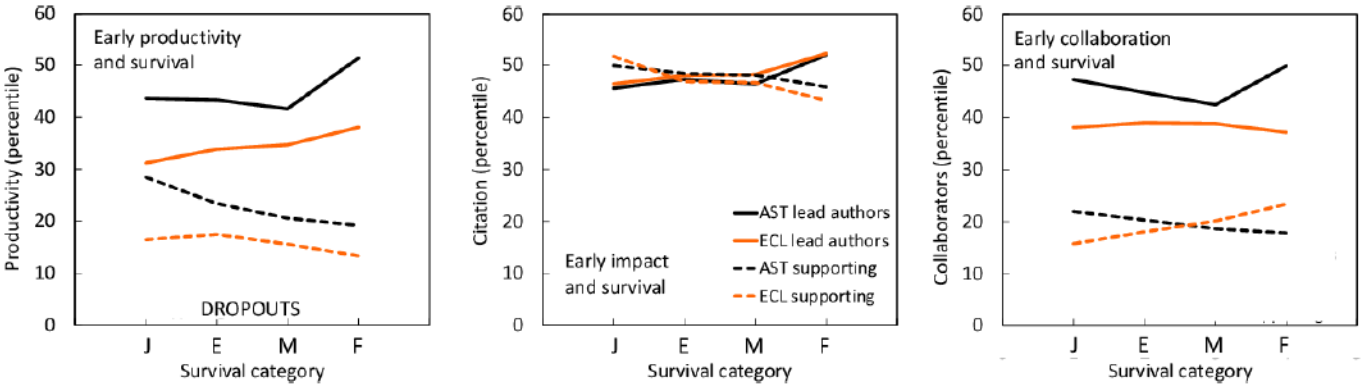

Fig. 6. Early predictors of survivability. Normalized productivity (panel A), impact (panel B) and collaboration (panel C) metrics based on the publication from the first five years of author's career are shown for lead (full lines) and supporting (dashed lines) authors in two disciplines for authors of different survival status: junior dropouts (J; leaving after 6-10 years after the first publication), early-career dropouts (E; 11-15 years), mid-career dropouts (M; 16-20 years), and finally the scientists who achieved full careers (F; >20 years).

a particular field, and are considered to be good representatives of those fields. We used a number of studies to identify the core journals for each discipline. For astronomy, (39) identified a list of core journals; the list used for ecology came from (40, 41). We define authors and derive their metrics from principal journals alone. Some of these authors may publish some fraction of their work in other journals (either other journals in the same or related area or in some cases in multidisciplinary journals).

Table 1. Cox proportional hazard regressions, for lead authors, by cohort.

Lead authors	(1) All	(2) All	(3) 1960s	(4) 1970s	(5) 1980s	(6) 1990s	(7) 2000s
Number of publications	.891*** (.004)	.891*** (.004)	.945* (.021)	.950*** (.013)	.925*** (.012)	.886*** (.008)	.857*** (.009)
Average citations per paper	.999 (.001)		.987** (.004)	.990*** (.003)	.994* (.002)	.998 (.001)	1.000 (.001)
Maximum citation on a paper		1.000 (.000)					
Number of collaborators	1.001 (.003)	1.001 (.003)	1.042 (.032)	.975 (.016)	.963** (.012)	.996 (.005)	.997 (.005)
Cases	34037	34037	1862	4764	6195	9511	11705
Exits	9034	9034	617	1227	1843	3531	1816
LR Chi-square	1111.49	1110.52	22.14	82.20	160.91	557.17	508.79
p>chi-square	.00	.00	.00	.00	.00	.00	.00

Productivity, citations and collaborators pertain to author's first five years of the career. ***p<0.001, **p<0.01, *p<0.05.

Table 2. Cox proportional hazard regressions, for supporting authors, by cohort.

Supporting authors	(1) All	(2) All	(3) 1960s	(4) 1970s	(5) 1980s	(6) 1990s	(7) 2000s
Number of publications	.966*** (.008)	.966*** (.008)	.972 (.099)	1.056 (.066)	1.042 (.046)	1.017 (.015)	.938*** (.012)
Average citations per paper	1.001** (.000)		1.003 (.007)	.990* (.005)	1.005* (.002)	1.001* (.000)	1.000 (.000)
Maximum citation on a paper		1.000* (.000)					
Number of collaborators	1.006 (.015)	1.003 (.016)	1.223 (.244)	1.038 (.093)	.964 (.061)	.942* (.023)	.994 (.024)
Cases	10677	10677	195	761	1540	3136	5045
Exits	4290	4290	91	308	767	1865	1259
LR Chi-square	59.16	55.76	1.83	7.70	5.77	10.95	103.24
p>chi-square	.00	.00	.61	.05	.12	.01	.00

Productivity, citations and collaborators pertain to author's first five years of the career. ***p<0.001, **p<0.01, *p<0.05.

Table 3. Cox proportional hazard regressions, for lead and supporting authors, by cohort.

	(1) All (lead authors)	(2) AST (lead authors)	(3) ECL (lead authors)	(4) ROB (lead authors)	(5) All (supporting authors)	(6) AST (supporting authors)	(7) ECL (supporting authors)	(8) ROB (supporting authors)
Number of publications	.891*** (.004)	.921*** (.005)	.867*** (.012)	.924 *** (.024)	.966*** (.008)	.969 *** (.082)	1.012 (.052)	1.105 (.099)
Average citations per publication	.999 (.001)	1.001 (.001)	.999 (.001)	.996 (.003)	1.001** (.000)	1.001** (.000)	1.009*** (.001)	.992 (.005)
Number of collaborators	1.001 (.003)	1.001 (.003)	1.018 (.009)	.979 (.018)	1.006 (.015)	1.019 (.016)	.937 (.066)	.822 (.102)
Cases	34037	22178	9499	2360	10677	6791	2988	898
Exits	9034	4613	3488	933	4290	2476	1409	405
LR Chi-square	1111.49	417.21	129.20	27.42	59.16	39.73	26.73	6.33
p>chi-square	.00	.00	.00	.00	.00	.00	.00	.09

Productivity, citations and collaborators pertain to author's first five years of the career. ***p<0.001, **p<0.01, *p<0.05.

This incompleteness will reduce the metrics, and in some cases may affect the determination of career length or authorship role. Quantifying the incompleteness and its effects is difficult given the lack of topical classification at the article level. However, since the analyses in the paper are relative (one time period vs. another, authors with one set of characteristics vs. another), the incompleteness will not affect some time periods or authors more than the others, so the relative trends should be unaffected. Our choice is conservative because the alternative, including all works that match some name, would greatly exacerbate the name disambiguation problem and potentially confounds the results.

All of the analyses are derived from the bibliographic data extracted from the full Clarivate Analytics Web of Science database. We used the entire temporal span of the database (from 1900 to 2015) to establish the starting and the ending years of activity of each author and thus identify the cohorts. For astronomy and ecology we follow cohorts from 1961 and for robotics since 1985 (none of the core robotics journals published prior to 1983). The number of authors belonging to these cohorts and included in the analysis is 71,164 in astronomy, 20,704 in ecology and 17,646 in robotics.

To identify unique authors, we perform, for each field-specific dataset separately, disambiguation of author names using the hybrid initials method. The scheme represents an improvement over standard initials methods because it either ignores or takes into account the middle initial depending on the name frequency (42), minimizing the splitting of unique authors due to inconsistent use of middle initial while maximizing the author separation. Percentages of authors whose identity has been compromised due to either splitting or merging have been estimated by simulation and are between 3 and 5% (42), which is below a level that would significantly affect our results. Ambiguity is relatively low because we focus on principal journals alone.

The roles that authors play in knowledge production (lead and supporting) are established from author lists in the following way. Authors on single authored papers are given a lead author status. In order to establish the roles in multi-authored papers, we have first verified that in all three disciplines under study the author lists are ordered by author contributions (with the first author almost always matching the corresponding author), except in rare cases when they are ordered alphabetically. We find no evidence for a deliberate alphabetical listing in papers with fewer than ∼4 authors, and in such cases we adopt the first author as a lead author. For longer lists of authors we check if the author list is alphabetical (based on up to seven first-listed authors), and if it is not, again take the first listed author as a lead author. If the list is alphabetical, we determine the lead author only if the corresponding author is not the first author. The fraction of articles for which the lead author could not be determined is relatively small (1.6%, 0.2% and 0.3% for astronomy, ecology and robotics, respectively).

For each unique author we establish the cohort year as the year when he or she first appeared as an author in any role (lead author or a supporting author). Since our data extend to periods prior to the starting time for the analysis, the cohort year, as well as the year of the departure from the field, can be established reliably. An author is considered currently active if he or she has published (in any role) in the last three years covered by database. Of the active authors some have achieved full-career status (defined as at least 20 years of active publishing), whereas for others their ultimate survival status is currently unknown and they are excluded from those analyses where such information is required.

RESULTS

Growth of supporting-author scientists

Previous results on the growth of team science and the changing structure of such teams allows us to propose that one component of the changing career demographics of scientist is a differentiation into heterogeneous career paths, with some scientists becoming lead authors and others specializing as non-lead supporting team members. Here we establish the extent to which each of these groups has contributed to the creation of knowledge over the last half century.

Figure 2 shows the fraction of authors from each cohort who at any point in their career will contribute to the field as lead authors. The fraction of lead authors has been experiencing a dramatic downward trend in all three disciplines since the 1960s, leading to a complementary increase in the share of supporting authors. Furthermore, the proportion of lead authors has been similar in all three fields, indicating that the shift of roles may follow a universal pattern. While in the early cohorts, from the 1960s and 1970s, the vast majority (∼75%) of entering authors had a lead-author role, this percent has dropped to less than 40% in most recent cohorts. The strong shift is unrelated to the presence of transient authors. If those were excluded from the cohort, the drop in the share of lead authors remains similar: from ∼85% in the 1960s to ∼50% in the current decade. Is the increasing fraction of supporting authors an inevitable outcome of increasing team sizes? To test this possibility we performed modeling in which we went through all the papers in each dataset and tried to replace coauthors (all authors except the lead author) who are classified as supporting authors with the authors who have the status of being lead authors and were active at the time of paper publication. In this modeling the number of authors per paper remains the same, as well as the individual (lead-author) productivity (because we only replace coauthors) and yet, we were able to populate mock author lists solely with lead authors. This demonstrates that having large team sizes does not automatically require the recruitment of supporting scientists. It also signifies that large teams are not entirely the product of collaboration among eventual full-role scientists (which may be more prevalent in small teams) but rather involves a recruitment of a special workforce of supporting scientists.

Survival function and the decreased half-life of cohorts

The minimal level of contribution to scientific knowledge is the production of a single paper. The existence of such authors was first pointed to by Price and Gürsey in 1976 (13), who named this type of author "transients" and established that they accounted for 25% of the population of scientists in the late 1960s. In Figure 3 we find that the fraction of transients has remained relatively constant in most cohorts, although this category of authors has started to increase in recent cohorts across all three fields (since about the 1990s), especially in robotics and ecology. Notably, we also find that unlike the fraction of lead authors which is universal, the number of transients is field dependent, with levels in astronomy similar to the ones Price & Gürsey found, and much higher rates (50-70%) in ecology and robotics. Interestingly, one quarter of recent transients, in all three fields, were lead authors. This fraction was as high as one half in the 1960s. This suggests that the threshold for lead authorship is often crossed even in the population that never genuinely embarks on a research path in that discipline.

For authors who persist after the initial publication we employ survival analysis to study their scientific career longevity. In Figure 4, we show the survival curves of select cohorts spanning the period of the most recent four decades. Survival curves are calculated as the fraction of a cohort remaining after x years. While the survival curves of contemporaneous cohorts in different fields have different slopes, we see that in each field the curves undergo a similar evolution: from relatively long survival times in the 1980s, to very rapid attrition of the scientific workforce in most recent times. We observe that until the 1980s (1990s for

astronomy), more than half of each cohort had "full" (20+ year) careers. However, in recent decades, it is no longer the case. The results correspond to a continuous decline in the expected career length.

In order to expand the survival analysis to every cohort and cover the full period from 1961, we calculate, for each cohort, its *half-life*, the time it takes to lose 50% of the cohort. Half-lives are determined from a linear fit to the survival function, regardless of whether the cohort has yet reached 50%. Half-lives for the three fields as a function of cohort year are shown in the lower right panel of Figure 4. In astronomy, the half-life has dropped from about 37 years in 1960s to just 5 years in 2007. In ecology and robotics, the half-lives are even shorter, and have also been decreasing at similar rates. When we analyze lead authors and supporting authors separately, we find that in ecology and robotics their half-lives are similar, whereas in astronomy the half-lives of supporting authors are shorter than those of the lead authors by about 5 years. Most recently (2010 cohort) half-lives are 9 and 4 years respectively.

Career progression model

To pave the way for a more fundamental understanding of the processes that lead to the attrition of the workforce, we describe the career trajectory of an individual researcher using a simplified version of the model of Figure 1. In the simplified version, we focus only on non-transient authors. Further we neglect the difference among types of dropouts. During a career, a researcher can be in one of the following four states: B the beginning of a career (defined with the first paper), S achievement of the supporting author role, L achievement of the lead-author role and X cessation of the career. An author can initially be in S state and transition into L state. The $S \rightarrow L$ transition is considered irreversible, i.e., $L \rightarrow$S is not allowed in the model. Authors continue in their states until reaching state X. We train the model using the data at our disposal. We find that the $S \rightarrow L$ transformation takes place in the first five years of a career: authors who become leads achieve this status quickly. For the survival model, we are interested in the likelihood of observing the transitions $S \rightarrow X$ or $L \rightarrow X$, i.e., the hazard probability. We show the hazard probability in Figure 5, separately for lead and supporting authors. For lead authors the hazard probability is relatively constant, at around 0.03. For supporting authors, the exit probability is higher and shows a two-mode behavior: decrease in the first 8 years and reaching a more stable value subsequently. We model the hazard function as a piece-wise linear+constant function:

$h = a(t - t_{\text{break}}) + b$ for $t \leq t_{\text{break}}$, and $h = b = \text{const.}$ for $t > t_{\text{break}}$,

where a b are constants and $t_{break} = 8$ is the time where the hazard function changes behavior. For astronomy, the model is tested against the data in the right panel of Figure 5. The survival curves are now based on all cohorts, so they represent time-averaged survival for lead and supporting authors. The model reproduces the salient features of the empirical curve. Remarkably, the analysis shows that the hazard is relatively constant throughtout the career, i.e., that there are no punctuated bottlenecks at which a large fraction of a cohort would leave the field.

Early indicators of scientific survivability

Given the increasing uncertainty of achieving a full career in science, one wonders whether there are any characteristics of scientists early in their careers that could indicate their survival status (38). We define 'early' as the first five years of a researcher's presence in the field (what we might call their "apprenticeship" years). Given our focus on the roles that scientists play in the production of knowledge, we focus on the variables that are directly related to this process: productivity, impact, and collaboration. These variables have been identified in prior work as correlated with career trajectories. We do not focus on some other variables that have been identified as important for career longevity and success: such as gender and the prestige of an institution a scientist is affiliated with, which are more pertinent in the context of studies that focus on career aspects that involve institutional and job roles (hiring, tenure and promotion). While our models do not explicitly control for gender, two recent studies analyzing career longevity of academic faculty found no differences in faculty attrition by gender (except in the field of mathematics) since 1990 (37, 43).

In this analysis we look at the total productivity in the first five years of a career (in any authorship role), and we examine two types of impact: average impact of early work (the number of citations per paper received in the first five years) and the peak impact (the maximum number of citations received in a 5-yr window to a single, early-career publication). Finally, for collaboration we focus on the number of direct collaborators in the first five years of the career. Direct collaborators are defined as coauthors on a paper led by the author in question, as well as all the unique lead authors of papers on which the author in question is a coauthor. If neither author is a lead on some publication they do not constitute direct collaboration.

In order to aggregate the data from cohorts that span a long time period, one needs to take into account that all three variables have significantly increased over time. For example, a researcher from the 1960 cohort who had 10 citations per paper may have been the most impactful in that cohort (~100 percentile), whereas the same number of citations for a cohort from 2000 may place the researcher in middle of the cohort (~50 percentile). Therefore, we establish normalized measures, by determining the percentiles for each variable and for each author in a given cohort.

Figure 6 shows mean productivity, citation and collaboration levels for authors of different survival categories: junior dropouts (J; leaving after 6-10 years after the first publication), early-career dropouts (E; 11-15 years), mid-career dropouts (M; 16-20 years), and finally the scientists who achieved full careers (F; >20 years). The values for robotics, which contains fewer cohorts and smaller sample size is noisier and we omit it for clarity. The trends are shown separately for lead and supporting authors. The trends are fairly consistent between astronomy and ecology (with the exception of collaboration). Furthermore we find that the trends involving average citation per paper and maximum citation are very similar and we show only the ones involving average citations. Figure 6 reveals that lead and supporting authors follow different trends. Overall, lead authors, regardless of survival category, have significantly higher production and collaboration levels than supporting authors, whereas their impact levels are similar. Supporting authors, while working on fewer papers and with fewer direct collaborators, nevertheless contribute to projects of similar impact. For lead authors there is a slight positive trend between the early level of all three metrics and the eventual survival (except for ecology and collaboration where there is no significant trend). In particular, based on the means comparisons, lead researchers who go on to full careers (F) tend to have, on average, higher levels of productivity, citation and (for astronomy) collaboration.

The four-state career model which provides an estimate of the career termination hazard rate by career state not only supports the empirical survival functions well, but shows that the hazard rate is relatively constant throughout the career, thus also supporting the model developed by Petersen et al. (34).

The above plots focused on individual variables. To quantify the effect of the variables on survival taking into account internal correlations we use the Cox proportional hazard survival model. For this analysis we use career lengths in annual increments (rather than grouping into only four categories) and the Efron method to correct for ties. Although many of the cases include careers of greater than 20 years, we recode career length as maximizing at 20 (hence all careers greater than 20 years, corresponding to full-career survival status, are treated as right

truncated). In addition, because we are testing the effects of the first five years of performance on subsequent exit, all our cases in this analysis have career lengths of at least 6 years. We are then testing, among the set of researchers who accumulate five years of background experience, how career lengths differ by publications, citations and number of collaborators during their first five years (net of the effects of the other variables). We use the untransformed publications and citations data, as we will be focusing on comparisons within cohorts.

Given the very different survival curves for the lead and supporting authors (Figure 5), we estimate the effects separately for each group. Tables 1 and 2 give the models. Column 1 in Tables 1 and 2 shows the effects of background characteristics (publications, citations, number of collaborators) on hazards of exit (with values greater than 1 increasing the rate of exit and values less than 1 decreasing the rate of exit). The first panel shows the results for lead authors and the second panel shows the results for supporting authors. Column 2 repeats this analysis using the maximum cites among the researchers first 5 years of publications. We see that, when we control for the net effects of the other indicators, across the 50 years (1960-2010) for lead authors, publications significantly reduce the hazard of exit, while there is little effect of citations (either measure) or number of collaborators. For supporting researchers, publications also have a negative effect on exit, although the effect is weaker than for lead authors. Citations (either measure) also have an effect, although the effect is positive (increasing exit). Number of collaborators has no effect.

A test of the proportional hazard assumption that the effects of the predictors are constant over time rejects the null hypothesis for publications (and is close to significant for citations). Furthermore, the data above suggest that the career conditions are changing over time. And, also, publications, citations and collaborations rates have all been changing over time. Hence, we estimate the effects across cohorts separately (columns 3 to 7). For lead authors, we see that publications have consistently been a significant predictor of career longevity. We also see that citations reduced the hazard of exit in the early cohorts, but more recently, the model is dominated by publications, with citations having little independent effect. In contrast, for supporting authors, publications have very weak effects until the most recent cohort. Table 3 shows that these effects are largely consistent across fields, although we find that the effect of publications is significant for supporting researchers in astronomy.

In Tables 1 and 2, we report the hazard ratios from a multivariate Cox proportional hazard model. We are estimating the relative hazard to exiting, truncating at 20 years (so we are estimating the relative hazard of leaving academic publishing before 20 years). The table is reporting the change in the hazard ratio for exiting from a one-unit change in each variable, controlling for the effects of all the other variables. These hazard ratios can be interpreted by estimating how far they are from 1.0. For example, for lead authors across all years, publications have a coefficient of .891 (column 1). This means that one publication reduces the hazard of exit by about 11% (1.0-0.891 = 0.109). In terms of the probability of achieving a full career, it grows gradually from 50% for authors with one early publication to 85% for authors with 20 publications. In contrast, one citation reduces the hazard very little (0.1%). Therefore, for lead investigators, each publication has substantially more impact on survival than does each citation (about 100 times greater). In contrast, for supporting authors, one publication reduces the hazard of exit by about 3% (1.0-0.966), while citations again have very little effect. However, looking across the cohorts, we see this effect for supporting authors is largely limited to the most recent cohort (column 7). We can also see that, for the full span of cohorts (column 1), the effect of publications for lead authors is much greater than that for supporting authors (11% v. 3%), and that when we compare across cohorts (columns 3 to 7), the effect of publications for reducing exit is stronger (the hazard ratio is lower) for lead than for supporting authors.

DISCUSSION

Recent work on the organization of science has focused on the internal structures of research teams, and argued that one likely outcome of this shift in the nature of scientific work has been the growth of dependent scientists, whose careers depend on being members of such teams (6, 13). And, less obviously, there has been a concomitant increase in high stakes evaluation and competition for funding, hence increasing emphasis on productivity (44-47). One solution to this new emphasis on productivity is increasing division of labor (48, 49). The growth of scientific team sizes is being accompanied by a transition in the organization of scientific work from craft to bureaucratic industrial principles, with increased division of labor and standardization of tasks (13, 50, 51). The result is a growth of scientists whose function is to support the projects that others are lead on. Our results confirm this scenario, showing that increasing fraction of entering authors never transition from supporting to lead author role. We also show that such trend is not an inevitable outcome of the increasing sizes of teams *per se*, but arises due to the different roles that some authors now have in large teams compared to the roles that members of smaller teams have (team *members* vs. collaborators). In some fields, such as ecology and robotics, lead and supporting authors have similar half-lives, while in others, such as astronomy, the half-lives of supporting authors is significantly shorter.

Of course, there are well known productivity advantages from organizing teams with a division of labor, and with some team members specializing in supporting roles (48). Hence, it is perhaps not surprising that science is shifting to larger teams, with more specialization, and that increasingly some scientists are specializing in supporting roles. Note that we are not assuming status or skill distinctions in our classification of lead and supporting authors (50). We are arguing that such supporting scientists are critical to the production of contemporary science (6). However, it is also the case that institutions, such as universities and funding agencies, build around these traditional status distinctions, for example, between post-doctoral scientists and tenure track professors (6). However, our survival analyses suggest that the criteria predicting longevity for supporting scientists is quite distinct from that for lead researchers and it may not be appropriate to impose similar criteria on both groups when making decisions about who to hire or renew. We argue there is a need to reform career structures in universities to account for the changing nature of the population composition and reproduction cycles in team science, with social insect colonies rather than parent-child reproduction as a more appropriate model.

While we cannot address this with our current data, we point to a tension between the research production and teaching functions that academic labs provide (5, 12, 44, 50, 52). These two trends are bringing fundamental changes to scientific careers, with decreasing opportunities for lead researcher positions and increasing production of, and demand for, a scientific workforce to fill positions as permanent supporting scientists. These trends together suggest downward pressure on career longevity (as more people exit the academic science labor force), and the growth of dependent supporting scientist positions to support the relatively shrinking share of lead researchers. However, one concern is that such supporting scientist positions do not fit well with the employment system in most universities, which are structured around a graduate apprenticeship, a short period of post-doctoral training, and then movement into a tenure track (and eventually tenured) professor position (5). Instead these support workers may be relegated to a series of short term post-doc contracts

or other forms of contingent academic work. While this model implies an up or out academic pipeline (with significant shares of the research workforce dropping out of research active academic positions at each stage), the growth of permanent supporting scientists may suggest an alternative career path that, while perhaps with shorter survival than the traditional lead researcher path, may be a growing share of the academic labor force. Furthermore, such careers may be premised on a different set of criteria than is typically predictive of the career survival of lead researchers.

Our findings show that the shift in the mode of knowledge production, from solo authors and small core teams (2) has coincided with a stratification in the scientific workforce in terms of their roles. The increased need for both the specialization and possession of specialized technical knowledge to manipulate increasingly complex instrumentation and data, has created an essential group of supporting contributors to knowledge. Unfortunately, the existing job roles and educational structures may not be responding to these changes. Our results suggest that, while essential, these supporting researchers are suffering from greater career instability and worse long-term career prospects in some fields.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grant EAGER SMA-1645585. This work uses Web of Science data by Clarivate Analytics provided by the Indiana University Network Science Institute and the Cyberinfrastructure for Network Science Center at Indiana University.

1. Fortunato S, *et al.* (2018) Science of science. *Science* 359:eaao0185.
2. Milojević S (2014) Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences (PNAS)* 111(11):3984-3989.
3. Wuchty S, Jones BF, & Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science* 316(5827):1036-1039.
4. Bordons M & Gomez I (2000) Collaboration networks in science. *The web of knowledge: A festschrift in honor of Eugene Garfield*, eds Cronin B & Atkins HB (Information Today, Medford, NJ), pp 197-213.
5. Stephan PE (2012) *How economics shapes science* (Harvard University Press, Cambridge, MA).
6. Barley SR & Bechky B (1994) In the backrooms of science: Notes on the work of science technicians. *Work and Occupations* 21:85-126.
7. Cyranoski D, Gilbert N, Ledford H, Nayar A, & Yahia M (2011) The PhD factory. *Nature* 472:276-279.
8. Schillebeeckx M, Maricque B, & Lewis C (2013) The missing piece to changing the university culture. *Nature Biotechnology* 31(10):938-941.
9. Stephan P (2012) Perverse incentives. *Nature* 484:29-31.
10. Alberts B, Kirschner MW, Tilghman S, & Varmus H (2014) Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences* 111(16):5773-5777.
11. Kaiser D (2012) Booms, busts, and the world of ideas: Enrollment pressures and the challenge of specialization. *Osiris* 27(1):276-302.
12. Teitelbaum MS (2014) *Falling behind? Boom, bust, and the global race for scientific talent* (Princeton University Press, Princeton).
13. Walsh JP & Lee Y-N (2015) The bureaucratization of science. *Research Policy* 44:1584-1600.
14. Geuna A & Shibayama S (2015) Moving out of academic research: Why do scientists stop doing research? *Global mobility of research scientists: The economics of who goes where and why*, ed Geuna A (Academic Press, Amsterdam), pp 271-300.
15. Preston AE (2004) *Leaving science: Occupational exit from scientific careers* (Russell Sage Foundation, New York).
16. Gaule P & Piacentini M (2018) An advisor like me? Advisor gender and post-graduate careers in science. *Research Policy* 47(4):805-813.
17. Allison PD (2014) *Event history and survival analysis* (Sage, Los Angeles, CA).
18. Allison PD & Stewart JA (1974) Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review* 39(4):596-606.
19. Long JS, Allison PD, & McGinnis R (1993) Rank advancement in academic careers: Sex differences and the effects of productivity. *American Sociological Review* 58(5):703-722.
20. Sugimoto CR, Sugimoto TJ, Tsou A, Milojević S, & Larivière V (2016) Age stratification and cohort effects in scholarly communication: A study of social sciences. *Scientometrics* 109(2):997-1016.
21. Clemens ES, Powell WW, McIlwaine K, & Okamoto D (1995) Careers in point: Books, journals, and scholarly reputation. *American Journal of Sociology* 101(2):433-494.
22. Xie Y & Shauman KA (2003) *Women in science: Career processes and outcomes* (Harvard University Press, Cambridge).
23. Allison PD & Long JS (1990) Departmental effects on scientific productivity. *American Sociological Review* 55(4):469-478.
24. Long JS, Allison PD, & McGinnis R (1979) Entrance into the academic career. *American Sociological Review* 44(5):816-830.
25. Long JS & McGinnis R (1985) The effects of the mentor on the academic career. *Sociometrics* 7:255-280.
26. Leahey E, Keith B, & Crockett J (2010) Specialization and promotion in an academic discipline. *Research in Social Stratification and Mobility* 28:135-155.
27. Cole JR & Cole S (1967) Scientific output and recognition: a study in the operation of the reward system in science. *American Sociological Review* 32:377-390.
28. Fox MF & Faver CA (1985) Men, women, and publication productivity: Patterns among social work academics. *The Sociological Quarterly* 26(4):537-549.
29. Xie Y & Shauman KA (1998) Sex differences in research productivity: New evidence about an old puzzle. *American Sociological Review* 63(6):847-870.
30. Way SF, Morgan AC, Clauset A, & Larremore DB (2017) The misleading narrative of the canonical faculty productivity trajectory. *Proceedings of the National Academy of Sciences* 114(44):E9216-E9223.
31. Fox MF (1983) Publication productivity among scientists: A critical review. *Social Studies of Science* 13(2):285-305.
32. Cole JR & Cole S (1973) *Social stratification in science* (The University of Chicago Press, Chicago).
33. Petersen AM, Riccaboni M, Stenley HE, & Pammolli F (2012) Persistence and uncertainty in the academic career. *Proceedings of the National Academy of Sciences* 109:5213-5218.
34. Petersen AM, Jung W-S, Yang J-S, & Stanley HE (2011) Quantitative and empirical demonstration of the Matthew effect in a study of career longevity. *Proceedings of the National Academy of Sciences* 108(1):18-23.
35. Price DJdS & Gürsey S (1976) Studies in scientometrics. Part I. Transience and continuance in scientific authorship. *International Forum on Information and Documentation* 1(2):17-24.
36. Keminski D & Geisler C (2012) Survival analysis of faculty retention in science and engineering by gender. *Science* 335(6070):864-866.
37. Box-Steffensmeier JM, *et al.* (2015) Survival analysis of faculty retention and promotion in the social sciences by gender. *PLoS ONE* 10(11):e0143093.
38. Sinatra R, Wang D, Deville P, Song C, & Barabási A-L (2016) Quantifying the evolution of individual scientific impact. *Science* 354(6312).
39. Henneken EA, *et al.* (2007) E-print journals and journal articles in astronomy: A productive co-existence. *Learned Publishing* 20(1):16-22.
40. Nobis M & Wohlgemuth T (2004) Trend words in ecological core journals over the last 25 years (1978-2002). *Oikos* 106:411–421.
41. Carmel Y, *et al.* (2013) Trends in ecological research during the last three decades – A systematic review. *PLoS ONE* 8(4):e59813.
42. Milojević S (2013) Accuracy of simple, initials-based methods for author name disambiguation. *Journal of Informetrics* 7(4):767-773.
43. Kaminski D & Geisler C (2012) Survival analysis of faculty retention in science and engineering by gender. *Science* 335(6070):864-866.
44. Hackett EJ (1990) Science as a vocation in the 1990s: The changing organizational culture of academic science. *The Journal of Higher Education* 61(3):241-279.
45. Hicks D (2012) Performance-based university research funding systems. *Research Policy* 41(2):251- 261.
46. Lewis JM (2015) Research policy as "carrots and sticks": Governance strategies in Australia, the United Kingdom and New Zealand. *Varieties of governance. Studies in the political economy of public policy*, eds G. C, M. H, & M. R (Palgrave Macmillan, London), pp 131-150.
47. Whitley R & Gläser J eds (2007) *The changing governance of the sciences: The advent of research evaluation systems* (Springer Dordrecht).
48. Becker GS & Murphy KP (1992) The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics* 107(4):1137-1160.
49. Smith A (1776) *Wealth of nations*.
50. Hagstrom WO (1964) Traditional and modern forms of scientific teamwork. *Administrative Science Quarterly* 9:241-263.
51. Hargens LL (1975) *Patterns of scientific research: A comparative analysis of research in three scientific fields* (American Sociological Association, Washington, DC).
52. Pavlidis I, Petersen AM, & Semendeferi I (2014) Together we stand. *Nature Physics* 10:700-702.