Publish or Perish

Selective Attrition as a Unifying Explanation for Patterns in Innovation over the Career

Huifeng Yu
Gerald Marschke
Matthew B. Ross
Joseph Staudt
Bruce A. Weinberg

ABSTRACT

Studying 5.6 million biomedical science articles published over three decades, we reconcile conflicts in a long-standing interdisciplinary literature on scientists' life-cycle productivity by controlling for selective attrition and distinguishing between research quantity and quality. While research quality declines monotonically over the career, this decline is easily overlooked because higher "ability" authors have longer publishing careers. Our results have implications for broader questions of human capital accumulation over the career and federal research policies that shift funding to early-career researchers—while funding researchers at their most creative, these policies must be undertaken carefully because young researchers are less "able" on average.

Huifeng Yu is a former doctorate student at the University at Albany, SUNY, Albany, NY. Gerald Marschke is an Associate Professor of Economics at the University at Albany, SUNY, Albany, NY and Research Associate at the National Bureau of Economic Research, Cambridge, MA (marschke@albany.edu). Matthew B. Ross is Associate Professor of Public Policy & Economics at Northeastern University, Boston, MA. Joseph Staudt is an Economist at the Center for Economic Studies at the U.S. Census Bureau, Suitland, MD. Bruce A. Weinberg is the Eric Byron Fix-Monda Professor of Economics at Ohio State

(continued on next page)

[Submitted December 2019; accepted March 2021]; doi:10.3368/jhr.59.2.1219-10630R1
JEL Classification: J24, I23, and J10
ISSN 0022-166X E-ISSN 1548-8004
Supplementary materials are available online at: https://jhr.uwpress.org/.
Huifeng Yu https://orcid.org/0000-0001-7505-6170
Gerald Marschke https://orcid.org/0000-0002-8231-5625
Matthew B. Ross https://orcid.org/0000-0002-7806-3138
Joseph Staudt https://orcid.org/0000-0002-5685-713X
Bruce Weinberg https://orcid.org/0000-0001-8856-1803

I. Introduction

The scientific workforce is aging (Blau and Weinberg 2017), and an ever-increasing share of resources are being devoted to older researchers. At the National Institutes of Health (NIH), the largest science funder, the share of senior investigators aged 56 or older has nearly doubled in the past 16 years. There are now nearly twice as many senior researchers funded by the NIH as the number of early-career investigators, that is, those aged 40 and younger. The share of older investigators is projected to exceed mid-career investigators, those aged between 41 and 55 years, in the next few years (Charette et al. 2016).[1] Even in the face of a wide range of pressing scientific and medical challenges, former NIH director Elias Zerhouni identified the aging of scientists as "the number-one issue in American science" (Kaiser 2008a). Concerns about the aging of the scientific workforce have prompted a series of policies to redirect resources from older to younger researchers. Key to the success of a policy that reallocates funding across researchers of different ages is an evidence-based understanding of how innovativeness varies over the career.

Unfortunately, despite nearly 150 years of research (for example, Beard 1874; Lehman 1953), no consensus has emerged on the relationship between age/experience and innovativeness. Consequently, there is no strong empirical foundation to inform science policy decisions on the allocation of resources across the age distribution of scientists. For example, psychologists and sociologists have largely focused on the lives of great

University, Columbus, OH and Research Associate at the National Bureau of Economic Research, Cambridge, MA. *The authors thank Mikko Packalen for helpful input on the text analysis and Neil Smalheiser and Vetle Torvik for the use of their Author-ity data and Clarivate for access to Web of Science records. The authors thank Chun-Yu Ho, Adrian Masters, Byoung Park, and seminar participants at Rensselaer Polytechnic Institute and SUNY Albany for helpful comments. All authors gratefully acknowledge support from the National Institute on Aging and the Office of Behavioral and Social Science Research via P01 AG039347. Weinberg is grateful for support from R24 AG048059, R24 HD058484, UL1 TR000090; NSF DGE 1760544, 1535399, 1348691; and SciSIP 1064220 and the Ewing Marion Kauffman and Alfred P. Sloan Foundations. Weinberg was supported on P01 AG039347 by the NBER directly and on a subaward from NBER to Ohio State. Marschke gratefully acknowledges support from NSF DGE 1661278 and NSCE 1918445. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the National Institutes of Health, or the U.S. Census Bureau. All errors are the responsibility of the authors. The research in this paper does not use any confidential Census Bureau information. This work was conducted with NBER IRB approval (IRB Ref#13_116). The results in the paper are from an analysis of 5.6 million publications from PubMed that have been matched to Clarivate's Analytics' Science Citation Index Expanded (SCIE) database. All programs and data files necessary to produce the results are permanently archived at https://www.nber.org/research/data/publish-or-perish-explaining-patterns-innovation-over-career. Some of the micro-level raw data upon which the results of this study are based were obtained under license from Clarivate Analytics (https://clarivate.com). Most of the data can be redistributed, but the citation metrics, which come from Clarivate Analytics' Science Citation Index Expanded, are proprietary and according to the terms of the authors' license with Clarivate cannot be redistributed in their entirety. These metrics have been removed from the data files. Readers can contact Jeffrey Clovis (IP&Science, jeff.clovis@Clarivate.com) for information on obtaining the same micro-level raw data used.*

1. Moreover, the average age for the first-time recipient of an independent NIH grant rose to 42 in 2007 from 37 in 1980 (Kaiser 2008b). Over roughly the same period, the average age among all NIH PIs increased to 49 from 39 (White, Rush, and Schaffer 2009) and the average age of new professors hired in medical schools increased to 37.5–40 from 34–36 years old, depending on the degree (Gingras et al. 2008).

innovators, such as Darwin, Newton, and Einstein. Based on these prominent examples, they theorize and find empirical evidence supporting the idea that innovation declines over the career (for example, Zuckerman and Merton 1973; Simonton 1997).[2] In contrast, economists generally subscribe to the logic of human capital models (for example, Becker 1975), which emphasize a concave relationship between experience and innovation. Empirical support for this relationship can be found in Levin and Stephan (1991), as well as work by Cole (1979); Falagas, Ierodiakonou, and Alexiou (2008); Horner, Rushton, and Vernon (1986); Costas, van Leeuwen, and Bordons (2010); and Kyvik and Olsen (2008). A third view, with prominent contributions from physicists, suggests that the probability of important contributions does not change over the career (for example, Sinatra et al. 2016; Simonton 1997).[3] Finally, a few researchers point to a "swansong" effect, where scientific output increases shortly before the end of the career (Lehman 1953; Davis 1954; Haefele 1962).[4] The contradictory findings in this sizeable interdisciplinary literature can largely be attributed to conceptual differences across fields, empirical methodologies of varying rigor, and the use of selected subsamples of particular scientists.

We use a data set covering nearly the universe of biomedical publications over a 30-year period from 1980 to 2009. These data include detailed information on the authors, citations, references, and text. Using this approximate population-level data set, we simultaneously combine three conceptual and empirical advances that, together, constitute a sharp departure from previous work and also reconcile the existing literature's inconsistent findings.

First, we leverage the longitudinal structure of our data to compare models that control for time-invariant researcher differences to those that do not, leading us to conclude that the vast majority of existing empirical work confounds the effect of aging with compositional changes in the quality of scientists over the career, that is, selective attrition. Indeed, we find that selective attrition, whereby the most innovative researchers tend to have longer careers, plays a decisive role in mediating the shape of scientists' life-cycle productivity profiles and explains much of the discrepancy in the prior literature. This finding helps explain why, for instance, studies using Nobel Laureates (for example, Weinberg and Galenson 2019) and other eminent researchers (often conducted by psychologists), for whom attrition is relatively small, tend to find earlier peak ages than studies using broader cross-sections of the population of scientists.

2. Supporting the psychologist's perspective, a number of explanations have been offered for a decline in scientific productivity over the life cycle, including declining cognitive ability (Verhaeghen and Salthouse 1997), the depletion of a fixed set of ideas (Simonton 1997), and competition with later generations of scholars who have access to better methods, tools, and theoretical training than earlier generations (Kyvik 1990). Other studies that have shown publication rates declining with age include Diamond (1984) and Levin and Stephan (1989). Interesting work in progress integrates the more economic and psychological views, focusing on human capital and fluid versus crystalized intelligence (Kaltenberg, Jaffe, and Lachman 2021).

3. Looking retrospectively at careers of physicists with publication careers of at least 20 years, with at least ten publications, and at least one paper every five years, Sinatra et al. (2016) found that a scientist's top article, their highest-impact work as defined by the number of citations received, could appear at any time during their career. They found the same random pattern for other subsamples of physicists, as well as for samples of chemists, cognitive psychologists, ecologists, economists, and neuroscientists.

4. Additionally, other age–productivity profiles have been reported. Gingras et al. (2008) and Over (1982) report publication rates increasing with age, and Feist (2006) finds a cubic relationship between productivity and age.

Second, we make a conceptual distinction between the quantity and quality of scientific output. We find that the estimated life-cycle productivity curve varies dramatically in shape depending on which type of measure is used, and thus we conclude that distinguishing between quantity and quality is essential for understanding how innovativeness varies over the life cycle. We operationalize the conceptual difference between quantity and quality by using publications, text, and citations to construct a wide range of metrics that richly characterize both the inputs and outputs of the scientific research process. Previous work has rarely used measures of scientific productivity that extend beyond publication counts (perhaps weighted by some measure of journal quality) or citation counts. However, scientific impact is a multidimensional construct (Bollen et al. 2009; Cronin and Sugimoto 2014, 2015), so we go beyond publication and citation counts by constructing additional metrics that capture the breadth of an article's impact (based on the range of fields that cite it), whether the article is employing the best and latest ideas, citing the best and latest research, and whether the article is drawing from multiple disciplines.[5]

Our third main contribution is to control for changes in a researcher's role in the laboratory over the life cycle. As in many scientific fields, biomedical research is increasingly performed by large teams (Wuchty, Jones, and Uzzi 2007), where different members play dramatically different roles and where a single member's role typically evolves with experience. In published biomedical research, the order in which authors are listed sheds light on these roles. The last author, who is generally older, is typically the principal investigator (PI) on a project and provides intellectual direction and resources for the team. In contrast, the first author tends to be the scientist in the last author's laboratory who designs and implements the research strategy, analyzes the data, and writes the manuscript, but under the last author's supervision. The first author is often young, either a graduate student or postdoctoral researcher. Successful researchers "graduate" from first to last author (Costas and Bordons 2011; Gingras et al. 2008; Marschke et al. 2018). The middle authors usually play much smaller roles in the research. Taking advantage of these uniform standards in biomedicine with respect to the ordering of author names, we construct measures that allow us to control for the role played by scientists in the laboratory as they mature. This allows us to disentangle the effects on productivity of role changes as a scientist ages from the effects of aging per se. For instance (and as we argue below), as a scientist ages, the quality of their coauthors may decline as they move from working with mostly (relatively) high-quality mentors as a young scientist to mostly (relatively) low-quality mentees as a more experienced scientist. This role-changing dynamic may be especially important for scientists who enjoy long, successful research careers. The richness of our data compared to other empirical work in

5. These input quality measures reflect the quality of the scientist's human capital and overlap with measures described in Staudt et al. (2018) that are designed to identify transformative science. Other recent work has sought to extract more information from citation records than is afforded by simple citation counts (for example, Wang, Song, and Barabasi 2013; Hutchins et al. 2016). Some of this work has been focused specifically on new citation-based metrics of novelty. Acemoglu, Akcigit, and Celik (2015) used a range of rich characterizations of citations to identify the most innovative work. Wang, Veugelers, and Stephan (2017) identified novel research from unique combinations of citations. Funk and Owen-Smith (2016) used shifts in citation patterns to identify work that consolidates or destabilizes existing technologies. Foster, Shi, and Evans (2021) reviewed approaches to identifying novelty and developed a unifying simulation approach. Packalen and Bhattacharya (2019) used the use of important new concepts measured using text analysis.

this literature allows us to estimate a series of high-dimensional fixed-effect models that include author-by-position controls.[6] Thus, relative to the existing literature, we are able to better identify the pure effect of age on scientific output because these models allow us to simultaneously control for role changes and selective attrition.

No previous work has simultaneously accounted for selective attrition, the quantity–quality distinction, and the changing role of scientists over the life cycle. The paper most closely related to our work, and the only other to account for selective attrition, is Levin and Stephan (1991). However, their work focuses exclusively on the quantity of scientific output, and they are unable to control for systematic changes in the role that scientists play in the laboratory. Given our findings (discussed below) that selective attrition has a much greater effect on the life-cycle pattern of quality than quantity and that accounting for role changes is important for isolating age/experience effects, our empirical framework marks an important advancement.[7]

In addition to our empirical contributions, we make a novel conceptual contribution through a formal overlapping generations model that illustrates how selection and changing lab roles can differentially impact the quantity and quality of work over the life cycle. In the model, young researchers are a mix of high and low ability, but only the high-ability researchers continue to publish when old, illustrating the importance of selection. Young researchers work with a single (high-ability) older researcher, but older researchers supervise a cadre of (mixed high- and low-ability) young researchers, illustrating the changing roles played over the life cycle. Finally, we work out the different implications of the model for the quantity and quality of research output. This theoretical contribution is particularly relevant for science policymakers, who are charged with making difficult funding decisions in an increasingly resource-constrained world.

Our empirical results suggest that, when we do not control for unobserved individual ability (selective attrition), the *quality* of scientific output (as measured by citation counts) has an inverted-U shape over the life cycle. In contrast, controlling for unobserved individual ability with author-by-position fixed effects produces estimates showing that the average quality of articles declines substantially and uniformly over the life cycle. Relative to the oldest first and last authors, articles written by the youngest first and last authors receive 19 and 14 more citations, respectively. Relative to the mean, these translate into experience gradients of 65 percent and 48 percent over the life cycle.[8] Given the multifaceted nature of scientific quality, we also examine how alternative measures of

6. Packalen and Bhattacharya (2019) study the relationship between age and innovativeness among first and last authors, but without any fixed effects or other controls for ability. See also Bhandari et al. (2004) and Baerlocher et al. (2007).

7. Levin and Stephan (1991) used multiple waves of the Survey of Doctorate Recipients, with repeated observations of the same scientists, and found an inverted-U shape between age and publications in five out of six subfields of physics and earth sciences. As noted, Levin and Stephan improve upon *quantity*-based studies in their use of a more-representative sample for scientists and sophisticated treatment of the attrition problem.

8. Though not the primary focus of the paper, our results on the quality of output also indirectly shed some light on how academic tenure affects the productivity of scientists. Tenure's detractors note the disincentive to exert research effort, while its advocates emphasize the freedom it affords researchers to pursue more transformative albeit riskier research programs (Franzoni and Rossi-Lamastra 2017; Brogaard, Engelberg, and Van Wesep 2018). Our main results show a continuous decline in *average* publication quality over the life cycle but also show that the quality of researchers' *best work* changes little over the career, suggesting tenure does not cause them to "swing for the fences" (Brogaard, Engelberg, and Van Wesep 2018), at least not successfully.

quality vary over the life cycle.[9] We find that, relative to younger authors, older authors use fewer new scientific concepts (identified using text analysis), cite older and less impactful references, and cite a narrower range of references across fewer fields. Thus, it appears that older scientists do not use the newest ideas, do not build on the most promising recent research, and are less interdisciplinary than when they were young.[10]

Though our results make it clear that citations (and other measures of quality) decline over the life cycle, they do not necessarily reflect a decline in a given researcher's cognitive ability over time. For instance, a researcher's average citations may decline if the researcher becomes more likely over time to coauthor on relatively low-quality publications produced in collaboration with an increasing number of mentees. Indeed, our theoretical model allows for both lower-quality coauthorships and skill depreciation to contribute to the decrease in the quality of a researcher's output over time. Our main empirical models control for the variation in last author's publication quality due to variation in coauthor quality through first-author fixed effects. Thus, for a given last author, our main specifications control for one of the most important sources of variation in author quality. To further adjudicate between the age and mentorship stories, we show that conditioning on the number of publications a researcher coauthors at a given career age does not alter the measured quality–experience profile. Thus, it appears that an increased share of coauthorships with lower-quality junior scholars and a rising burden of mentorship cannot alone explain the estimated decline in average publication quality.

Consistent with models of human capital accumulation, we find evidence that the *quantity* of scientific output (that is, publication counts) is concave over the life cycle. Our finding persists regardless of whether we control for unobserved individual ability. Thus, we conclude that selective attrition does not qualitatively alter the experience profile with respect to the quantity of scientific output. In our regressions with person fixed effects, the quantity of output peaks at 21–25 years of experience, at which time researchers produce about 1.2 additional publications relative to the oldest researchers and 1.7 relative to the youngest. Our results are consistent with our overlapping generations model, in which the count of publications (that is, output quantity) rises over a scientist's career as they collaborate with an increasing number of mentees, even as quality declines.

Our results suggest that federal science funders (most notably the National Institutes of Health) face an important trade-off when deciding whether to shift resources from older to younger scientists. While the quality of an individual's research declines with experience, the average "ability" of a researcher who remains active is higher than the average young researcher. Thus, reallocating resources from the average senior to the average junior scientist has benefits in terms of funding people at a more productive point in the typical career but also has risks in terms of potentially funding a low-ability researcher.

The setting in this paper also provides a valuable laboratory for examining how productivity evolves over the career because, unlike most settings that rely on wages to proxy for productivity, we have direct measures of the quantity and quality of output.

9. Interestingly, when used as control variables, differences in these measured research inputs do not account for much of the experience-related variation in the quality of research output.

10. About this resistance, Max Planck famously wrote, "A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it" (Planck 1949).

Though we hesitate to extrapolate to other scientific fields (or the workforce more generally), given a presumption that scientific creativity peaks earlier in more mathematical or abstract fields (Weinberg and Galenson 2019), our finding of early creativity peaks in biomedicine (not a highly theoretical, abstract, or mathematical field) is suggestive evidence that the quality of research may decline with experience across a wide range of innovative disciplines.[11] In any case, regardless of the generalizability of our results, as STEM innovators become a more important segment of the workforce, they become more important to analyze per se (see Deming and Noray 2018).

II. Conceptual and Empirical Frameworks

A. Conceptual Framework

To motivate the analysis, we outline an overlapping generations model of scientific careers (additional details are in Online Appendix A). A new generation of young scientists arrives at the beginning of each period and conducts research as scientists for one or two periods, after which they retire. Thus, young scientists overlap with older scientists who were young in the previous period. Each young scientist writes one paper under the guidance of an older researcher, an arrangement reflecting the roles played by younger and older researchers within real authorship teams.

We assume scientific creativity or ability is mixed among the young. Let α denote the scientist's (fixed) ability, with $\alpha = \{\alpha_L, \alpha_H\}$ and $\alpha_H > \alpha_L$. By assumption, only H-type scientists survive to become older scientists in the next period. That is, all L-type scientists are assumed to attrit. Let N_{ot} and N_{yt} be the number of old and young researchers entering period t, respectively, and let p be the share of N_{yt} that are H-type scientists. In a steady state, $N_{ot} \leq pN_{yt} < N_{yt}$. Thus, while each young scientist writes one article, each older scientist supervises several. Note the number of articles published by a (two-period) scientist increases over their career from one to N_{yt}/N_{ot}.

If each older scientist has one unit of effort to allocate across their papers, the amount of effort they devote to a single paper, ε_t, is $N_{ot}/N_y < 1_t$. Moreover, age may change the scientist's effectiveness: a high-ability researcher will contribute α_H when young, but $\delta\alpha_H$, when old. If age reduces effectiveness, then $\delta \in [0, 1)$. The scientific contribution embodied by a paper is assumed linear in the scientist's age and effort-adjusted abilities: $\delta f(\varepsilon_t)\alpha_H + \alpha_L$ and $\delta f(\varepsilon_t)\alpha_H + \alpha_H$ for papers with low- and high-ability young scientists, respectively, where $f(\varepsilon_t)$ is assumed rising in ε at a decreasing rate. Because a scientist works only with high-ability researchers as a young scientist and with a mix of abilities as an older scientist, their average research quality must fall. In particular, the average quality of their papers falls by $(\alpha_H - \alpha_L)(1-p)$. That is, the quality decline over the career is greater the higher their chances of matching to low-ability workers when older and the greater the ability difference between high- and low-ability workers. Note that this result holds regardless of δ (because when they were junior, their senior coauthor's productivity was lowered by δ also). Their effectiveness on any paper also falls: from α_H when

11. While we are cautious about extrapolating to other fields, note that the life sciences and biomedicine constitute the largest scientific sector by scholarship, accounting for nearly one-half of all published scientific articles by U.S. authors (National Science Board 2019, Table S5a-17).

young to $\delta f(\varepsilon_t)\alpha_H$ when older, declining both because they are spread more thinly in their supervisory role and because of the age-related decline in their skills, though they may still be contributing more to the paper than their young coauthor. If $\delta > 1$ because of human capital accumulation, their effectiveness may still decline if δ is not too high.

At the cohort level, selective attrition (only H-type young scientists remain when old) assures that the cohort's mean ability rises with age. The mean ability of the cohort *adjusted for age-related decline* rises with age as long as $(1-\delta)\alpha_H < (1-p)[\alpha_H - \alpha_L]$, that is, as long as the age-related loss in effectiveness is small relative to the number of low-ability researchers and their ability disadvantage. The cohort's mean adjusted ability rises of course when $\delta > 1$.

Thus, the model predicts: (i) the average quality of a cohort increases over time (due to selective attrition); (ii) while publications rise over a scientist's career, their quality declines (because papers are written with younger researchers who are, on average, less "able"); and (iii) a scientist's effectiveness on papers declines as well due both to aging and their transition to a supervisory role (assuming no offsetting human capital accumulation).

B. Empirical Framework

At the article level, we estimate scientific innovativeness as

$$(1) \quad \tilde{Y}_a = \alpha_0 + E_{Fa}\theta_F + E_{La}\theta_L + X_a\alpha_X + \nu_F + \nu_L + \xi_a$$

where $\tilde{Y}_a$ is a measure of the "quality" of an article a, capturing reception, novelty, or the inputs into the article (detailed in Section III). The variables of interest are the experience (years since first publication) of first and last authors, E_{Fa} and E_{La}. We follow Levin and Stephan (1989) in using dummy variables for five-year increments.[12] We capture unobserved differences in individual productivity in different roles by including first (ν_F) and last (ν_L) author fixed effects. We estimate a separate fixed effect for each author who appears as a first or last author when appearing in each position. These are more general versions of author fixed effects because the impact of a team member's experience on article quality may vary with role. Specifically, these "author-position" fixed effects allow the effects to vary with job responsibilities: persons who are good at conducting research may or may not be as good at directing the lab and attracting resources, and vice versa. Our results are robust to specifying author fixed effects that do not vary with author position (Ross et al. 2018 presents such estimates). Lastly, X_a includes controls for article characteristics (detailed in Section III.A.3), and ξ_a is an i.i.d. error term.

Our measures of research quality vary across fields and over time. In the case of citations, for instance, even important papers in small or mature fields can receive considerably fewer citations than less important works in large or growing fields (Seglen 1997; Althouse et al. 2009), and the number of citations tends to increase over time as the scientific enterprise expands and shrinks in the most recent years (beyond which citations are not measured). For these reasons, it is important to control for differences in

12. The intervals are zero years, 1–5 years, 6–10 years, 11–15 years, 16–20 years, 21–25 years, 26–30 years, and more than 30 years. However, since we drop articles that have a last author with zero experience (which is an unusual pattern), last authors have only seven indicator variables. Online Appendix Table B.2 gives definitions of all variables.

quality across subfields and over time, but we cannot directly include time effects since they are collinear in a model of experience that includes individual fixed effects (Deaton and Paxon 1994). To overcome this identification problem, we adopt a two-step procedure adapted from Aguiar and Hurst (2013), where we first regress our raw measure of scientific output on time and subfield fixed effects and then use the residuals from that regression ($\tilde{Y}_a$), which are now orthogonal to both time and subfield (to account for field productivity differences), as the dependent variable in Equation 1.[13] The risk with this approach is that some factors that should be attributed to age will be attributed to time in the first stage, but that risk is attenuated by including all variation, including across cohort variation in the first stage.

In addition to examining the quality of scientific output, we are also interested in estimating how the quantity of research varies over the career. Thus, we also estimate models where the unit of analysis is an individual researcher–year pair. Here, our model takes the form

$$(2) \quad \tilde{Y}_{it} = \beta_0 + E_{it}\theta + \nu_i + \pi_{it}$$

where $\tilde{Y}_{it}$ is a measure of the quantity of individual i's output in year t, specifically, publication counts, sometimes broken down by author position. Details about these outcomes are contained in Section III. As in Equation 1, the principle variables of interest are included in E_{it}, a vector of five-year experience indicators. ν_i is a researcher fixed effect. Here too, we first regress outcomes on field and year dummy variables and estimate Equation 2 using residuals from those equations. Note that, while Equation 1 allows us to examine how the *average* quality of an author's work varies over the career, Equation 2 can also be used to examine how the quality of an author's *best* work or works vary over the life cycle by using the number of citations received by the author's most highly cited article each year as the outcome.

III. Data

Our data are derived from the MEDLINE 2014 baseline files distributed by the National Library of Medicine (NLM). These files contain information on more than 21 million biomedical articles published between 1946 and 2014, including article title, journal title, publication year, author names, author position, and publication type. In addition, each article is tagged with Medical Subject Headings (MeSH) that describe its content. To identify citation linkages across articles, we match the MEDLINE articles to the Clarivate Analytics' Science Citation Index Expanded (SCIE) database. The SCIE data contain MEDLINE articles published between 1950 and May 20, 2014, along with a list of their references. To identify the "experience" (or "career

13. As discussed by Deaton and Paxon (1994), any linear trend in output can be arbitrarily attributed to either the time fixed effects, experience fixed effects, or cohort fixed effects (which are absorbed by the person fixed effects). One recent stream of literature on innovation suggests bounding such estimates using a nested test where experience, time, and person/cohort effects are sequentially dropped from the model (Hall, Mairesse, and Turner 2007; Stephan 2010). In a series of robustness checks contained in Online Appendix F, we provide additional estimates where, instead of first residualizing the outcomes with respect to time and subfield fixed effects, we residualize with respect to subfield fixed effects and a linear time trend. These alternative specifications result in estimates comparable to those from our main specification.

age") of each article's first and last authors, we match articles to the Author-ity database (Torvik et al. 2005; Torvik and Smalheiser 2009), which contains disambiguated authors linked to their MEDLINE articles.[14] The Author-ity disambiguation permits the identification of each author's first article in MEDLINE (and the tracking of all subsequent articles), and thus the calculation of each author's experience. These data also allow us to calculate author network size and citation counts that omit self-citations.

Here, we briefly describe the sample restrictions made to the underlying data, but Online Appendix Table B.1 provides a detailed outline of how we arrive at our final analytical sample. To summarize, we begin with the 15,085,762 articles indexed in both in SCIE and MEDLINE. We then limit our sample to the 9,897,775 articles published between 1980 and 2009.[15] Since both data sets contain publications that would not be considered scientific contributions, we further limit our analysis to 7,198,087 research articles published within this period.[16] As our focus is on teamwork, and a sizeable majority of biomedical science is conducted by teams, we restrict our analysis to the 6,648,200 articles with two or more authors and without a truncated author list.[17] Further sample restrictions include omitting articles that are missing first or last author experience,[18] articles with many authors,[19] articles with a first or last author experience over 40,[20] and articles with a last author experience of zero.[21] After imposing these

14. Technically the data contain author clusters (that is, probable authors). The author name disambiguation is based on the "Author-ity" model (Torvik et al. 2005; Torvik and Smalheiser 2009). The resulting data set contains more than nine million identity clusters covering MEDLINE records up to July 2009. The overall recall is 98.8 percent, and precision is about 98 percent. While this performance compares favorably to other disambiguations at this scale, about 2 percent of articles belonging to a given investigator are misassigned to a second predicted individual. These splitting errors can occur because of very common names (for example, John Smith) or radical career changes (an investigator might abruptly change topic areas, affiliations, and sets of coauthors). Nonetheless, the Author-ity data set has already demonstrated broad scientific, social, and commercial impact. Numerous scholars have obtained the data set to facilitate their own research, and the National Library of Medicine (NLM) is using the data set in its PubMed/Entrez/Medline databases as the starting point for a scheme to assign Author IDs to all publications. An earlier version, the 2008 baseline, is freely available online (http://arrowsmith.psych.uic.edu, accessed December 6, 2022).

15. Articles published before 1980 are dropped because abstracts are important for generating our text metrics, and MEDLINE's coverage of abstracts is poor prior to 1980. Articles published after 2009 are dropped because Author-ity disambiguates MEDLINE only through July 2009. Note that career age is calculated using the first publication, even if it is before 1980.

16. We exclude MEDLINE articles of the following article types: Review, English Abstract, Case Reports, Historical Article, Comment, Portrait, Biography, Guideline, News or Conference.

17. MEDLINE provides only the first ten authors for articles published between 1984 and 1995 and the first 25 authors for articles published between 1996 and 1999. For articles published after 1999, MEDLINE does not truncate author lists.

18. Career age may be missing for articles outside of Author-ity, articles with institutional first or last authors that are not disambiguated, or because of disambiguation errors.

19. All articles with team size greater than 13 (the 99th percentile of team size) are removed because in very large teams, the impacts of first and last authors may be small.

20. If scientists publish their first paper at biological age 25 (this coincides with the middle of most scientists' Ph.D. training), a career age of 40 years coincides with a biological age of 65. Dropping articles for which either the first or last author has a career age older than 40 helps to avoid "lumping errors" in the Author-ity data. Note that a lumping error occurs when two or more separate authors (with separate publication histories) are erroneously lumped into the same identity cluster.

21. Last authors are usually experienced team leaders, so a zero-aged last author raises the likelihood that they are subject to a "splitting error" or that team roles are assigned unconventionally. Note that a splitting error occurs when a single author is erroneously split into two or more identity clusters (with erroneously separate publication histories).

additional restrictions, we are left with 5,613,189 articles, about 84 percent of the multiauthored research articles published between 1980 and 2009 that are in the SCIE–MEDLINE intersection.

A. Variable Construction

1. Measuring article quality

We construct two outcome variables to characterize article quality. The first is total citations ever received by an article.[22] Though citations are the conventional bibliometric artifacts by which influence on future researchers and works is assessed and tracked, particularly important scientific works not only receive many citations, but also impact many fields. Thus, our second measure of quality captures the disciplinary diversity of the citations an article receives. Formally, disciplinary diversity for article a is defined as

$$\textit{Disciplinary Diversity of Citations}_a = \left(1 - \sum_{\underline{f}} s_{a\underline{f}}^2\right) \times 1000$$

where $\underline{f}$ is the scientific field of each citing article, and $s_{a\underline{f}}$ is the share of citations received by article a from field $\underline{f}$.[23] The disciplinary diversity of citations is between zero and 1,000 and increases in the breadth of an article's impact across fields. If an article is cited by articles from more fields, its impact is wider, and the disciplinary diversity of citations is larger (closer to 1,000).

2. Measuring article inputs

In addition to examining the quality–experience profile, we also examine whether the measured inputs into articles can explain the shape of this profile. To characterize an article's inputs, we construct seven variables, which can be broadly separated into two categories: three variables constructed using the text in articles' titles and abstracts (text based) and four variables constructed using articles' references (reference based).

The first and second text-based metrics are the number of new and the number of old important concepts appearing in an article's title or abstract. A concept is said to be "important" if, relative to other concepts introduced to the MEDLINE corpus in the same year, it is above the 99.9th percentile in terms of total lifetime occurrences in titles

22. The residualization of outcomes with respect to subfield and year fixed effects, outlined in Section II, adjusts for differences in citations due to variations in publication years and in practices across fields.

23. Readers will recognize this as one minus the Herfindahl–Hirschman index. In addition to characterizing disciplinary diversity of citations (Rafols and Meyer 2010), the scientometric literature has also used this metric to measure the breadth of a patent's impact on future technology (Trajtenberg, Henderson, and Jaffe 1997). Online Appendix C describes how we assign fields to articles based on the Medical Subject Headings (MeSH terms) with which the articles are tagged. The MeSH terms are broadly hierarchical, ranging from the general (for example, Body Regions) to the specific (for example, Peritoneal Stomata). Online Appendix C describes how we aggregate these terms to a similar level of specificity, resulting in a division of biomedical science into approximately 6,200 comparable fields. Note that because the typical article is tagged with multiple MeSH codes, the typical article is assigned to more than one field (see the example in Online Appendix C).

or abstracts of articles.[24] An article is said to use a "new" concept or an "old" concept if it uses a word or word combination less than or more than five years after it was introduced to the MEDLINE corpus. We view the use of new important concepts as measuring an article's contribution to important new lines of scientific inquiry. The third text-based metric is the mean age of all concepts (regardless of whether they are important) used in the title or abstract of an article. See Online Appendix D for details on processing the text of MEDLINE titles and abstracts.

The first reference-based metric is the mean age of an article's references, which we view as a measure of an article's distance from the scientific frontier. The second reference-based metric captures the disciplinary diversity of an article's references, which we view as capturing the multidisciplinarity of an article. This metric is analogous to the disciplinary diversity of citations discussed above, but it is defined over the articles that a focal article references rather than the citations that a focal article receives. As with the disciplinary diversity of citations, the disciplinary diversity of references ranges between zero and 1,000 and increases in the breadth of fields referenced. The third and fourth reference-based metrics are the (log) mean number of citations received by the articles that a focal article references, which are less than or more than five years old, respectively.[25] These metrics capture the quality of the articles that a focal article builds on, distinguishing between the quality of newer and older references.

Unlike the variables that measure quality in Section III.A.1, which are not directly under the author's control, the variables that describe inputs to research are directly controlled by authors. We use the latter variables in two distinct ways. In Section IV.B, we use them as outcomes and examine how they change over the career. In Online Appendix Table E.8, we use them as control variables to examine whether they explain the quality–experience profiles identified in Section IV.A.

3. *Article-level control variables*

Finally, we construct a variety of article-level control variables. To proxy for the human resources and research effort applied in the production of the article, we compute the number of authors listed on each article. In addition, we compute the size of the author team's network—the total number of distinct coauthors of all authors in the three years preceding publication—which is an extension of the authorship team that can contribute research inputs and provide feedback. Also, with larger networks, the article's impact and reputation may diffuse more broadly through the profession (Singh 2005; Sorenson, Rivkin, and Fleming 2006; Azoulay, Zivin, and Wang 2010), possibly leading to higher and more diverse citations. To proxy for the geographical dispersion of the research team, we compute the number of distinct cities among the authors' affiliations. Evidence from the economics literature on technological diffusion suggests that the diffusion of ideas is limited by geography (for example, Jaffe, Trajtenberg, and Henderson 1993; Zucker, Darby, and Brewer 1998), and thus a more geographically dispersed set of authors bring a more diverse set of ideas into the production of research, which may alter article quality. A more geographically dispersed set of authors may also broaden the

24. Because of publication lags and because our data, while comprehensive, do not include all publications where a concept may first arise (for example, publications outside of biomedicine), we do not seek to identify each concept's "origin" article.
25. Note that any article that a focal article cites has at least one citation by construction.

Table 1
Summary Statistics

	All Sample		Number of Citations > 0	
	Mean	SD	Mean	SD
Observations	5,613,189		5,417,616	
Number of citations	29.49	77.27	30.55	78.44
Disciplinary diversity of citations	946.12	33.34	946.12	33.34
Number of references	30.19	17.77	30.67	17.68
Mean age of references	8.83	5.84	8.78	5.65
Disciplinary diversity of references	948.36	35.23	948.91	33.95
Number of important new concepts used	0.33	0.97	0.34	0.99
Number of important old concepts used	16.67	12.06	16.98	12.06
Mean age of concepts used	53.78	14.11	53.64	13.92
Mean citation of recent references (log)	4.374	1.127	4.399	1.108
Mean citation of old references (log)	5.550	1.707	5.570	1.698
Diff. in citations to recent and old references (log)	−1.188	1.630	−1.183	1.622
Number of authors	4.68	2.26	4.71	2.27
Size of team's network	82.61	109.99	83.71	110.78
Number of authors' cities	1.59	0.95	1.59	0.96
First author experience	7.85	8.21	7.82	8.17
Last author experience	17.66	9.69	17.72	9.66
Mean experience of middle author	10.65	7.53	10.67	7.51

geographic reach and speed of recognition of an article's contributions. In some specifications, we also use the mean experience of middle authors as a control variable.

B. Summary Statistics

Table 1 presents summary statistics of the variables defined above, separately for all articles and for articles with nonzero citations. Nearly 97 percent of the articles have been cited at least once, and the mean number of citations is 29.5 with a standard deviation of 77.3. The mean article references about 30 prior works, and the mean age of these references is about 8.8 years. The disciplinary diversity of citations and references average 946 and 948, respectively, with standard deviations of 33 and 35.[26] The mean numbers of new and old important concepts appearing in a title or abstract are 0.33 and 17. Both measures have large standard deviations of 0.97 and 12.1. The mean number of authors on an article is 4.7, and the mean number of cities represented by

26. In part, these relatively high degrees of dispersion (recall that the maximum possible dispersion is 1,000) reflect the fact that most articles are tagged with many fields, so even an article that is cited once can have a relatively high degree of citation dispersion.

author affiliations is 1.6. First authors have a mean experience of about 7.8 years, and last authors have a mean experience of 17.7 years, which is consistent with the conventional story about who occupies these roles.

Online Appendix Table B.3 presents all pairwise correlations between our outcome and control variables. The most important takeaway is that citation counts, the traditional measure of impact in the literature, and our new outcome metrics move in the same direction but have correlations that are always less than 0.5 in magnitude (and often much smaller), suggesting these new metrics are capturing aspects of influence and input quality that citations do not. Indeed, citation counts are positively correlated with disciplinary diversity of citations (0.47), disciplinary diversity of references (0.14), and number of new important concepts used (0.22). The negative correlation between citations and mean age of references (–0.18) suggests that the most influential works draw on more recent articles. The table also shows that articles drawing on a wider range of research fields (more multidisciplinary) impact a wider range of fields ($\rho = 0.34$) but are only slightly more impactful ($\rho = 0.14$).

Online Appendix Figure B.1 displays article counts for different experience combinations. Most articles are written by first authors with an experience of zero to ten years and last authors with an experience of one to 30 years, suggesting that younger first authors tend to team with middle-aged last authors. Note also that a disproportionate number of articles have first authors with experience zero, which is consistent with a high hazard rate early in the career and creates scope for important selection effects.

IV. Results

A. Article Quality and Author Experience

In this section, we use citation counts and the disciplinary diversity of citations (defined in Section III.A.1) as measures of an article's quality and examine how quality changes with author experience. Recall that, in order to identify the effects of experience while controlling for unobserved time and individual/cohort effects, all outcome variables are residuals obtained by regressing the raw outcomes on time and field fixed effects. Columns 1 and 2 of Table 2 show the results of estimating Equation 1 when the dependent variable is citation counts. To better illustrate the dynamics of citation counts over the career, Figure 1A plots the point estimates of θ_F and θ_L from Equation 1, along with their 95 percent confidence intervals.

Column 1 uses cross-person variation in experience (omits author-position fixed effects). For both first and last authors, the coefficient estimates imply an inverted U-shaped relationship between experience and the number of citations, although neither pattern is particularly strong. Citations peak for articles written by first authors with six to ten years of publishing experience and last authors with 11–15 years of experience. Articles written by first authors and last authors at these peaks receive, on average, 3.4 and 3.7 more citations than articles written by first and last authors in the highest experience categories. These differences amount to 12 percent and 13 percent of the sample mean of 29.5 and are statistically significant.

Column 2 eliminates cross-person variation in experience by including author-position fixed effects, and this dramatically changes the experience–citation profiles—for both

Table 2
Experience and Article Quality

Dependent Variable	Number of Citations		Disciplinary Diversity of Citations	
	(1)	(2)	(3)	(4)
Mean	29.49		946.12	
Standard Deviation	(77.27)		(33.34)	
First author experience				
0	0.103	19.146***	1.635***	5.448***
	(1.343)	(1.933)	(0.456)	(0.585)
1–5	2.949**	15.536***	2.743***	4.290***
	(1.417)	(1.894)	(0.578)	(0.482)
6–10	3.417**	8.872***	1.846***	2.729***
	(1.511)	(1.317)	(0.399)	(0.260)
11–15	1.867*	6.317***	1.461***	2.005***
	(1.036)	(0.983)	(0.293)	(0.177)
16–20	0.831	5.043***	0.929***	1.371***
	(0.675)	(1.053)	(0.140)	(0.152)
21–25	–0.257	2.532***	0.393***	0.463**
	(0.335)	(0.561)	(0.115)	(0.192)
26–30	0.716***	3.777***	0.759***	0.925***
	(0.248)	(0.864)	(0.210)	(0.308)
Last author experience				
1–5	1.486***	14.149***	–1.871***	3.709***
	(0.416)	(1.883)	(0.178)	(0.489)
6–10	2.667***	9.803***	–0.586***	2.336***
	(0.443)	(1.333)	(0.107)	(0.342)
11–15	3.672***	6.727***	0.558***	1.718***
	(0.436)	(1.026)	(0.091)	(0.205)
16–20	3.020***	4.012***	0.662***	1.102***
	(0.315)	(0.655)	(0.078)	(0.158)
21–25	1.561***	1.781***	0.339***	0.498***
	(0.223)	(0.385)	(0.051)	(0.089)
26–30	1.992***	2.255***	0.863***	0.797***
	(0.168)	(0.341)	(0.065)	(0.112)

(*continued*)

Table 2 (*continued*)

Dependent Variable	Number of Citations		Disciplinary Diversity of Citations	
	(1)	(2)	(3)	(4)
Number of authors	1.056***	3.389***	0.540***	0.744***
	(0.108)	(0.220)	(0.046)	(0.018)
Number of authors' cities	1.787***	0.678***	0.065	–0.080***
	(0.300)	(0.263)	(0.184)	(0.027)
Size of team's network	0.034***	0.007***	0.011***	0.005***
	(0.002)	(0.001)	(0.001)	(0.001)
Author–position FE		Yes		Yes
Observations	4,372,875	3,248,324	4,259,127	3,163,030
R^2	0.008	0.428	0.009	0.514

Notes: Standard errors in parentheses are clustered by field. The dependent variables are residuals from first-stage regressions that control for year and field fixed effects. The omitted groups are 31–40 years of experience for both first and last authors. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

first and last authors, citations steadily decline with experience. These effects are remarkably large—articles written by first and last authors in the youngest experience category receive 19 and 14 more citations, respectively, compared to those written by first and last authors in the oldest experience category. Since the mean article receives 29.5 citations, the experience gradients are 65 percent (first author) and 48 percent (last author). These results suggest that the upward-sloped portion of the citation profile in Column 1 is due to less innovative scientists exiting a research career at younger ages.

Particularly important works not only receive many citations, but also impact many fields. Columns 3 and 4 of Table 2 and Figure 1B show the results of estimating Equation 1 when the dependent variable is the disciplinary diversity of citations, which measures the breadth of an article's impact. Column 3 omits author-position fixed effects and shows that the breadth of scientific impact varies moderately over the careers of both first and last authors. For first authors, breadth rises, peaking early in the career (1–5 years), and then gradually declines. For last authors, breadth rises, peaks at 11–15 years, and then levels off, slightly rising later in the career (26–30 years) before declining for the omitted, oldest category. Column 4 includes author-position fixed effects, and, as with the results for citation counts, this dramatically changes the dynamics of the disciplinary diversity of citations over the career—for both author types, the breadth of scientific impact decreases with experience. These results indicate that the increase in observed breadth of impact with last author experience is due to changes in the ability composition along the career path and not a pure experience effect. Older scientists who still publish are among the high-ability survivors and produce more broadly impactful articles.

The results presented so far indicate that, in the cross-section, article quality displays a hump-shaped pattern over the career, but for a given researcher, article quality typically

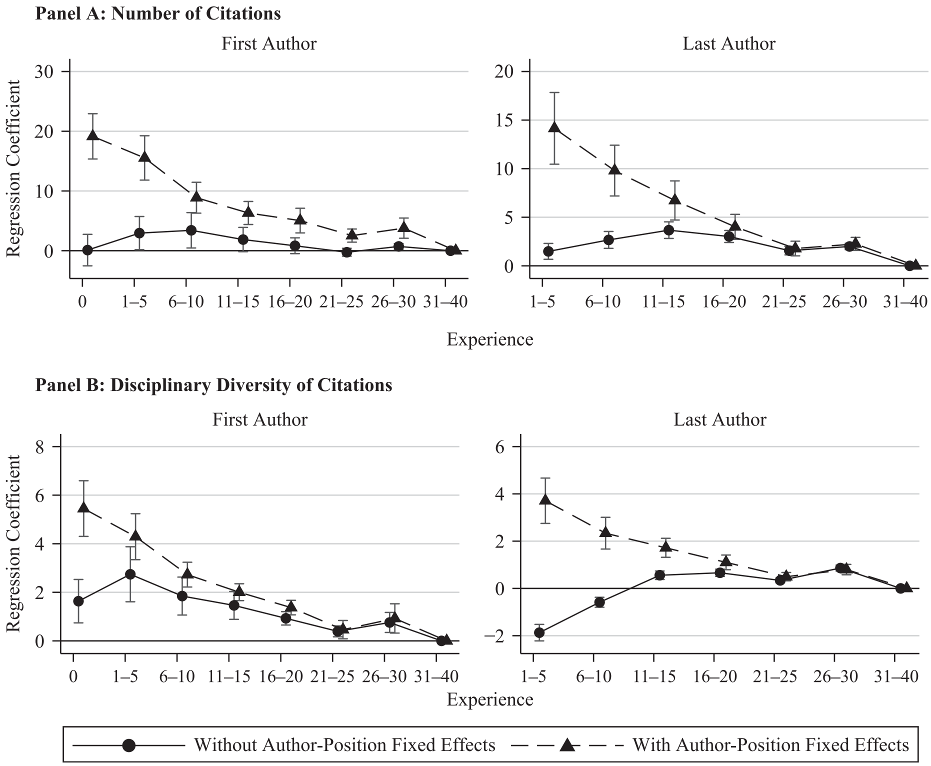

Figure 1

Experience and Article Quality

Notes: The figure plots the coefficients from Table 2.

declines over the career. In Online Appendix E, we show that the peaks identified for both the citation counts and the disciplinary diversity of citations are statistically significant[27] and that the shapes of the experience profiles are not driven by self-citations[28]

27. In Online Appendix Table E.1, we use one-sided *t*-tests to formally test whether the coefficient estimates for the peak experience groups identified in Table 2 are different from the coefficient estimates for the nonpeak experience groups. For instance, we test whether the coefficient estimate of 3.42 for first authors with 6–11 years of experience in Column 1 is statistically significantly larger than the coefficients for all other first author experience groups in Column 1. These tests confirm that all peaks identified in Table 2 are statistically significant.

28. Citations to one's own work may reflect self-serving motives and not scientific linkages, and when they do reflect linkages, they indicate internal knowledge transfers as opposed to spillovers to other researchers. To probe the robustness of our results we create a citation measure that excludes self-citations and reestimate the forward citations with self-citations removed. The results for both total citations and the disciplinary diversity of citations are reported in Online Appendix Table E.2. Dropping self-citations moderately changes the magnitudes of the differences across experience groups but does not qualitatively alter the results. When author-position fixed effects are omitted, the outcome–experience profiles have inverted U-shapes and when author-position fixed effects are included, the outcomes decline in experience.

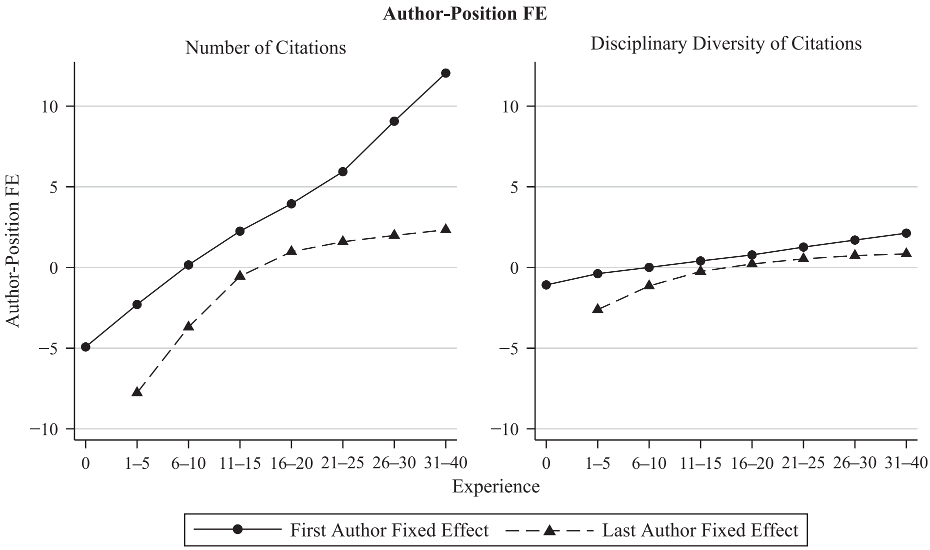

Figure 2

Experience and Author–Position Fixed Effects for Publishing Researchers

Notes: The figure plots the mean fixed effect of the people publishing at each level of experience from the regressions in Table 2 Columns 2 and 4. These can be thought of as a measure of the mean "ability" of the people publishing at each experience.

or the experience of middle authors.[29] As previously suggested, these results can be reconciled if the average "ability" of authors that continue to publish increases with experience—in other words, if lower "ability" authors cease to publish earlier in their careers. Using the estimated author-position fixed effects from Equation 1, $\hat{\nu}_F$ and $\hat{\nu}_L$, we can explicitly examine how the mean fixed effect ("ability") of authors who continue to publish changes with experience. Figure 2 plots mean author fixed effects from Columns 2 and 4 of Table 2 by experience.[30] The left panel shows that the fixed effects for last authors who publish at one to five years of experience is –8 citations, rising to +3 citations by the end of the career, a substantial change (the change for first authors is even larger, but late career first authors are rare). These figures directly confirm a striking improvement in the composition of authors at each level of experience for both total citations and disciplinary diversity of citations.

Our results thus far make it clear that research quality tends to decline over the life cycle of a researcher. However, as our model in Section II.A suggests, two possible

29. To examine this possibility, we calculate the mean experience of middle authors for articles with at least three authors. We report the results of the effect of middle authors' experience on article impact in Online Appendix Table E.3. Including this variable has little effect on the first and last author experience profiles, and the coefficient on mean author experience is economically and statistically insignificant in models with author-position fixed effects.

30. To construct this measure, we average the author-position fixed effects estimated in Table 2 for all authors publishing in each five-year experience group. Note that authors are implicitly weighted by their number of publications in each age bin, so that an author who publishes n times in a given experience bin receives n times as much weight as an author publishing one time.

mechanisms underlying this measured decline are depreciating skills/declining "ability" or increases in the number of low-quality coauthored articles driven by an increasing number of collaborations with mentees over the life cycle. Our fixed effects estimate control for first author fixed effects, so for people who are last authors, our main specifications control for one of the most important sources of variation in the "quality" of their coauthors. To shed additional light on these two possibilities, Online Appendix Table E.4 and Figure E.1 show quality–experience profiles that condition on the number of coauthored publications at a given experience level. Including this control does not alter the downward sloping quality–experience profile, suggesting it is not driven by authors taking on an increasing number of low-quality coauthorships as they age.

If we are correct that the conflicting quality-experience profiles from specifications with and without author-position fixed effects are driven by lower-ability researchers exiting research careers at earlier ages and thus increasing the average ability of researchers who continue to publish, then restricting our sample to authors with long careers should produce similar profiles whether or not we include author-position fixed effects in our regressions. Online Appendix Table E.5 examines this possibility by reproducing Table 2 but restricting the sample to articles with last authors who have career lengths of 20 years or more (no restriction is put on the career lengths of first authors). When author-position fixed effects are included, the experience profiles of last authors are very similar to those in Table 2. However, when fixed effects are not included, the results for last authors differ markedly from those in Table 2. Specifically, rather than increasing initially, the number of citations and the disciplinary diversity of citations are constant for the first two experience categories and then decline monotonically. Thus, whether we control for selective attrition using author-position fixed effects or through sample selection, the results strongly suggest that article quality tends to decline over the course of an author's career.

These regressions include characteristics of the researcher team to control for variation in human capital employed in the production of the research. The results show that teams with more authors who are more connected generate higher quality research as measured by citations and citation diversity. More geographically diverse teams also produce more cited work, though with less breadth in impact.[31] All regressions reported in Table 2, and in Tables 4 and 5 of Section IV.B, were repeated without team measures but the results are not reported for parsimony. While omitting the team-based measures changes the height of the profiles, the relationship between publication output (and input) quality and experience is robust to this exclusion.

In Table 3, we examine whether the relationships between publication quality and experience are pervasive across biomedicine by showing the distribution of the peak experience group over subfields.[32] When author-position fixed effects are excluded, first author citation counts peak at six to ten career years (the peak in the full sample) in 50.5 percent of subfields, and disciplinary diversity of citation peaks occur at one to five

31. Such measures may be related to experience (for example, more experienced researchers may be more connected than researchers fresh from graduate school), and thus these measures may be capturing some of the experience effects we observe.

32. We define subfields using level four MeSH fields (see Appendix C) and run the regressions separately for each subfield. To produce samples of sufficient size to estimate models with fixed effects, we include only subfields with at least 57,905 articles (the 99th percentile of the distribution of subfield size).

Table 3
Distribution of Peak Experience Groups for Subfields

Variable		Without Fixed Effects: Subfield Count	Without Fixed Effects: Subfield Percent	With Fixed Effects: Subfield Count	With Fixed Effects: Subfield Percent
Number of citations	First author experience				
	0	0	0.00%	275	95.16%
	1–5	82	28.37%	2	0.69%
	6–10	146	50.52%	5	1.73%
	11–15	10	3.46%	1	0.35%
	16–20	5	1.73%	1	0.35%
	21–25	12	4.15%	2	0.69%
	26–30	30	10.38%	1	0.35%
	31–40	4	1.38%	2	0.69%
	Last author experience				
	1–5	11	3.81%	253	87.54%
	6–10	31	10.73%	21	7.27%
	11–15	163	56.40%	8	2.77%
	16–20	36	12.46%	1	0.35%
	21–25	6	2.08%	1	0.35%
	26–30	42	14.53%	4	1.38%
	31–40	0	0.00%	1	0.35%
Disciplinary diversity of citations	First author experience				
	0	3	1.04%	235	81.31%
	1–5	203	70.24%	8	2.77%
	6–10	48	16.61%	22	7.61%
	11–15	12	4.15%	6	2.08%
	16–20	5	1.73%	4	1.38%
	21–25	5	1.73%	4	1.38%
	26–30	11	3.81%	3	1.04%
	31–40	2	0.69%	7	2.42%
	Last author experience				
	1–5	2	0.69%	235	81.31%
	6–10	7	2.42%	27	9.34%
	11–15	95	32.87%	7	2.42%
	16–20	54	18.69%	5	1.73%
	21–25	17	5.88%	2	0.69%
	26–30	113	39.10%	9	3.11%
	31–40	1	0.35%	4	1.38%

Notes: For this analysis, we focus on 289 subfields defined by level 4 MeSH terms (see Online Appendix C.1) fields with at least 57,905 articles (the 99th percentile of the distribution of subfield size). We then estimate the models in Table 2 for each subfield. The table reports the number and share of subfields that peak at each age bin for each specification for first and last authors.

career years (again, the peak in the full sample) in 70.24 percent of subfields. Last author citation count peaks occur at 11–15 years in 56.4 percent of cases, while the disciplinary diversity of citation peaks occur at 26–30 years in 32.9 percent of subfields. When author-position fixed effects are included, the results are remarkably consistent across subfields. First author citation count and disciplinary diversity of citation peaks occur in the youngest experience category (zero years) in 95.1 percent and 81.3 percent of subfields. Similarly, last author citation count and disciplinary diversity of citation peaks occur in the youngest category (1–5 years) in 87.5 percent and 81.3 percent of subfields.[33] Overall, these results show that the quality–experience profiles identified using the entire sample are remarkably consistent across subfields of biomedicine and confirm that the shapes of these profiles are not being driven by a small number of large subfields.

Taken together, the results in this section strongly suggest that article quality declines over the course of a scientist's career. This decline in quality exists whether measured by total citations or the disciplinary diversity of citations. It exists for both first and last authors, and whether we control for publication volume, and it exists whether we account for attrition using author-position fixed effects or by limiting the sample to authors with long careers. Moreover, the decline is remarkably consistent across subfields of biomedicine and is not impacted by the ages of middle authors or the elimination of self-citations. In addition, these results convincingly reconcile long-standing inconsistencies in the literature. When attrition is not accounted for, and the quality of articles produced by older scientists is compared to the quality of articles produced by younger scientists (ordinary least squares regressions without author-position fixed effects), the quality–experience profile is, indeed, moderately hump-shaped. Only by appropriately dealing with this attrition, using either author-position fixed effects or samples of authors with long careers, are we able to confirm that this hump-shape is simply due to changes in the ability composition along the career path and has little to do with experience per se.

B. Article Inputs and Author Experience

In the previous section, we demonstrated that, for a given researcher, the average quality of research—whether measured by total citations or the disciplinary diversity of citations—declines with experience. We now examine why this decline occurs. To do this, we examine how the text-based (Table 4) and reference-based (Table 5) variables, which characterize article inputs (and are defined in Section III.A.3), vary over the career. Text and references are inputs to the production of an article, influencing the direction the research takes, the scale of impact, and whether it is transformative or incremental. Ultimately, these variables help us identify whether younger authors incorporate into their work more of the raw materials that ultimately generate impact and transformation in science.

Columns 1 and 2 of Table 4 and Panel A of Figure 3 display regression results when the outcome variable in Equation 1 is the number of new important concepts used. With

33. Note that when the peak for a subfield differs from the peak in the full sample, the difference may be imprecise. We examine this possibility in Online Appendix Table E.6. For each subfield with a peak experience group different from the full sample result, we perform a one-sided *t*-test comparing the two peaks. With author-position fixed effects, only a small portion of subfields have a peak experience group statistically significantly greater than the peak experience group in the full sample.

Table 4
Experience and Article Inputs (Concept Use)

Dependent Variable	Number of New Concepts Used		Number of Old Concepts Used		Mean Age of Concepts Used	
	(1)	(2)	(3)	(4)	(5)	(6)
Mean	0.33		16.67		53.78	
Standard Deviation	(0.97)		(12.06)		(14.11)	
First author experience						
0	0.074***	0.179***	1.444***	2.086***	−1.587***	−2.170***
	(0.014)	(0.028)	(0.335)	(0.356)	(0.213)	(0.326)
1–5	0.099***	0.137***	2.273***	2.273***	−2.168***	−1.886***
	(0.015)	(0.025)	(0.409)	(0.305)	(0.253)	(0.266)
6–10	0.066***	0.070***	1.780***	1.536***	−1.597***	−1.181***
	(0.010)	(0.015)	(0.292)	(0.219)	(0.160)	(0.179)
11–15	0.041***	0.036***	1.491***	1.241***	−1.263***	−0.789***
	(0.007)	(0.010)	(0.192)	(0.164)	(0.106)	(0.149)
16–20	0.019***	0.015**	0.955***	0.874***	−0.790***	−0.493***
	(0.004)	(0.007)	(0.085)	(0.120)	(0.050)	(0.106)
21–25	0.004	0.007	0.445***	0.441***	−0.392***	−0.261***
	(0.003)	(0.005)	(0.051)	(0.077)	(0.047)	(0.076)
26–30	0.010	0.016**	0.525***	0.669***	−0.473***	−0.409***
	(0.006)	(0.007)	(0.098)	(0.071)	(0.098)	(0.118)

(continued)

Table 4 (*continued*)

Dependent Variable	Number of New Concepts Used		Number of Old Concepts Used		Mean Age of Concepts Used	
	(1)	(2)	(3)	(4)	(5)	(6)
Last author experience						
1–5	0.100***	0.239***	−0.591***	0.553**	−1.082***	−1.612***
	(0.011)	(0.038)	(0.157)	(0.223)	(0.155)	(0.116)
6–10	0.086***	0.159***	−0.115	0.432***	−0.950***	−1.039***
	(0.008)	(0.025)	(0.085)	(0.131)	(0.101)	(0.085)
11–15	0.073***	0.103***	0.592***	0.436***	−0.922***	−0.744***
	(0.008)	(0.018)	(0.048)	(0.096)	(0.045)	(0.067)
16–20	0.047***	0.056***	0.585***	0.297***	−0.662***	−0.449***
	(0.006)	(0.011)	(0.046)	(0.070)	(0.029)	(0.041)
21–25	0.025***	0.029***	0.347***	0.136***	−0.329***	−0.200***
	(0.003)	(0.006)	(0.027)	(0.040)	(0.027)	(0.032)
26–30	0.028***	0.031***	0.522***	0.189***	−0.559***	−0.360***
	(0.003)	(0.006)	(0.034)	(0.068)	(0.027)	(0.031)

(*continued*)

Table 4 (*continued*)

Dependent Variable	Number of New Concepts Used		Number of Old Concepts Used		Mean Age of Concepts Used	
	(1)	(2)	(3)	(4)	(5)	(6)
Number of authors	0.017***	0.022***	0.614***	0.542***	–0.393***	–0.223***
	(0.001)	(0.001)	(0.042)	(0.018)	(0.035)	(0.013)
Number of authors' cities	–0.014***	–0.007***	–0.581***	–0.110***	0.444***	0.108***
	(0.004)	(0.002)	(0.104)	(0.016)	(0.076)	(0.030)
Size of team's network	0.001***	0.000***	0.012***	0.004***	–0.010***	–0.002***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Author–position FE		Yes		Yes		Yes
Observations	4,488,900	3,328,811	4,488,900	3,328,811	4,423,970	3,279,249
R^2	0.012	0.469	0.054	0.583	0.022	0.528

Notes: Standard errors in parentheses are clustered by field. The dependent variables are residuals from first-stage regressions that control for year and field fixed effects. The omitted groups are 31–40 years of experience for both first and last authors. $*p<0.10$, $**p<0.05$, $***p<0.01$.

Table 5
Experience and Article Inputs (References)

Dependent Variable	Mean Age of References		Disciplinary Diversity of References		Mean Citations of Recent References (Log)		Mean Citations of Old References (Log)		Difference in Citations to Recent and Old References (Log)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Mean	8.83		948.36		4.44		5.59		−1.17	
Standard Deviation	(5.84)		(35.23)		(1.09)		(1.65)		(1.58)	
First author experience										
0	−0.543***	−1.152***	1.116**	2.053***	0.124***	0.211***	0.167***	0.192*	−0.050***	0.021
	(0.084)	(0.076)	(0.440)	(0.514)	(0.029)	(0.046)	(0.039)	(0.104)	(0.012)	(0.065)
1–5	−0.933***	−1.129***	3.063***	2.975***	0.154***	0.137***	0.207***	0.147	−0.047***	−0.005
	(0.111)	(0.079)	(0.493)	(0.412)	(0.032)	(0.042)	(0.040)	(0.095)	(0.009)	(0.060)
6–10	−0.583***	−0.661***	2.036***	2.044***	0.120***	0.071**	0.125***	0.096	0.001	−0.017
	(0.084)	(0.039)	(0.400)	(0.280)	(0.026)	(0.028)	(0.030)	(0.073)	(0.006)	(0.049)
11–15	−0.483***	−0.531***	1.766***	1.571***	0.083***	0.039*	0.089***	0.039	−0.002	0.007
	(0.053)	(0.031)	(0.288)	(0.254)	(0.018)	(0.022)	(0.022)	(0.051)	(0.006)	(0.033)
16–20	−0.336***	−0.378***	1.228***	1.013***	0.041***	0.020*	0.045***	0.007	−0.001	0.018
	(0.031)	(0.026)	(0.189)	(0.197)	(0.008)	(0.012)	(0.014)	(0.034)	(0.007)	(0.026)
21–25	−0.189***	−0.197***	0.544***	0.385***	0.015***	0.009	0.014	−0.005	0.004	0.016
	(0.021)	(0.031)	(0.164)	(0.127)	(0.005)	(0.008)	(0.008)	(0.018)	(0.006)	(0.013)
26–30	−0.297***	−0.337***	0.986***	0.704***	0.020**	0.017	0.041***	−0.008	−0.023***	0.021**
	(0.041)	(0.050)	(0.149)	(0.193)	(0.010)	(0.011)	(0.009)	(0.012)	(0.006)	(0.010)

(continued)

Table 5 (*continued*)

Dependent Variable	Mean Age of References		Disciplinary Diversity of References		Mean Citations of Recent References (Log)		Mean Citations of Old References (Log)		Difference in Citations to Recent and Old References (Log)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Last author experience										
1–5	−0.562***	−0.823***	−4.355***	−0.370	−0.051***	0.226***	−0.148***	0.164**	0.102***	0.070
	(0.042)	(0.085)	(0.450)	(0.448)	(0.011)	(0.020)	(0.018)	(0.072)	(0.009)	(0.054)
6–10	−0.498***	−0.525***	−1.497***	0.683***	0.027***	0.146***	−0.053***	0.133***	0.079***	0.019
	(0.033)	(0.071)	(0.158)	(0.223)	(0.009)	(0.016)	(0.009)	(0.049)	(0.008)	(0.035)
11–15	−0.526***	−0.384***	0.084	0.758***	0.071***	0.098***	0.017**	0.109***	0.055***	−0.007
	(0.034)	(0.056)	(0.079)	(0.159)	(0.009)	(0.011)	(0.007)	(0.038)	(0.005)	(0.028)
16–20	−0.396***	−0.254***	0.355***	0.507***	0.055***	0.055***	0.019***	0.069***	0.037***	−0.011
	(0.028)	(0.043)	(0.060)	(0.114)	(0.007)	(0.007)	(0.005)	(0.025)	(0.005)	(0.019)
21–25	−0.222***	−0.127***	0.249***	0.265***	0.031***	0.028***	0.008***	0.035***	0.024***	−0.006
	(0.017)	(0.029)	(0.047)	(0.070)	(0.003)	(0.004)	(0.003)	(0.013)	(0.004)	(0.010)
26–30	−0.289***	−0.144***	0.512***	0.339***	0.043***	0.036***	0.024***	0.032*	0.020***	0.007
	(0.014)	(0.019)	(0.053)	(0.078)	(0.003)	(0.004)	(0.003)	(0.016)	(0.003)	(0.015)

(*continued*)

Table 5 (*continued*)

Dependent Variable	Mean Age of References		Disciplinary Diversity of References		Mean Citations of Recent References (Log)		Mean Citations of Old References (Log)		Difference in Citations to Recent and Old References (Log)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Number of authors	–0.125***	–0.071***	0.747***	0.560***	–0.051***	0.226***	0.029***	0.029***	–0.014***	–0.008***
	(0.015)	(0.006)	(0.053)	(0.038)	(0.011)	(0.020)	(0.004)	(0.002)	(0.002)	(0.002)
Number of authors' cities	0.120***	0.041***	–0.439**	–0.043	0.027***	0.146***	–0.024**	–0.010***	0.013**	–0.002
	(0.019)	(0.005)	(0.209)	(0.050)	(0.009)	(0.016)	(0.011)	(0.002)	(0.006)	(0.001)
Size of team's network	–0.004***	–0.001***	0.008***	0.002***	0.071***	0.098***	0.001***	0.000***	0.001***	0.000
	(0.000)	(0.000)	(0.001)	(0.001)	(0.009)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)
Author–position FE		Yes		Yes		Yes		Yes		Yes
Observations	4,402,841	3,263,662	4,430,478	3,293,024	4,329,432	3,219,993	4,307,461	3,192,511	4,198,607	3,118,025
R^2	0.017	0.533	0.012	0.583	0.031	0.575	0.010	0.550	0.002	0.510

Notes: Standard errors in parentheses are clustered by field. The dependent variables are residuals from first-stage regressions that control for year and field fixed effects. The omitted groups are 31–40 years of experience for both first and last authors. Note that in Columns 9 and 10 the difference is log(new) – log(old). *$p<0.10$, **$p<0.05$, ***$p<0.01$.

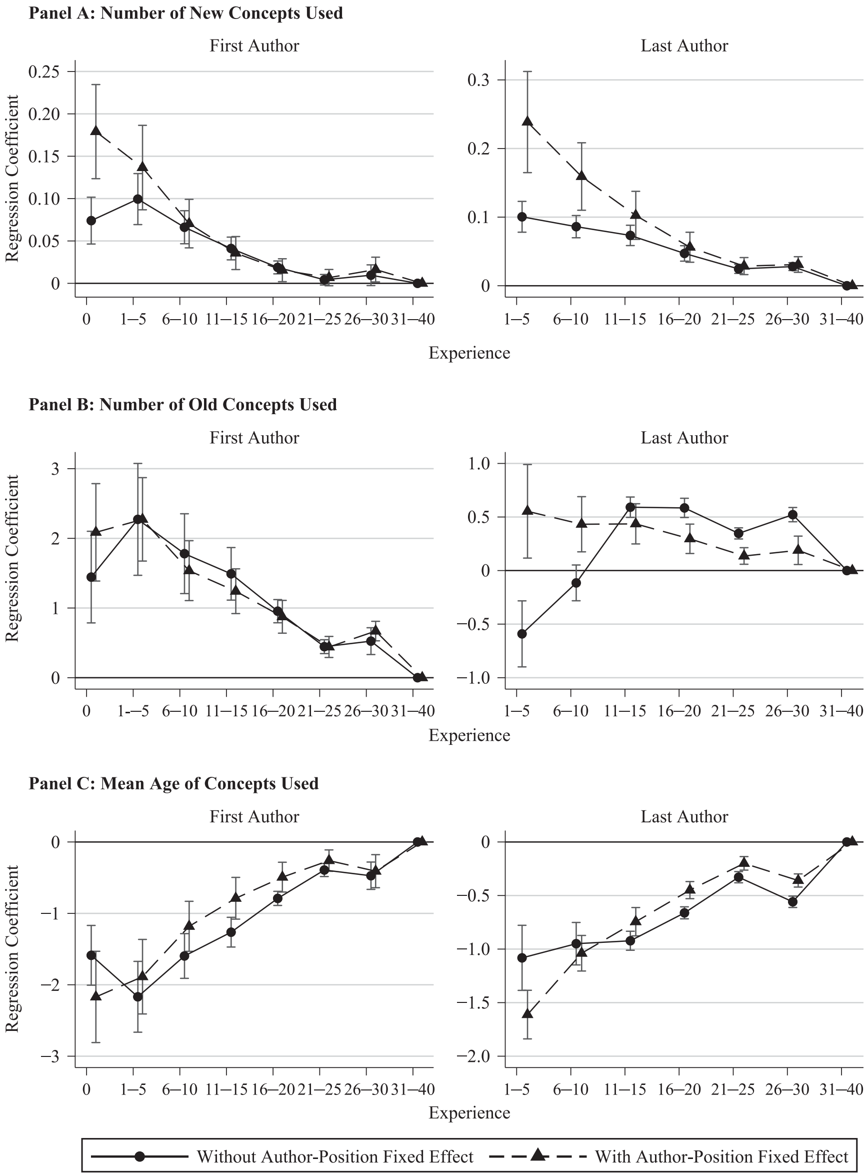

Figure 3

Experience and Article Inputs (Concept Use)

Notes: Estimates from Table 4.

or without author-position fixed effects, an article's use of new important concepts decreases with experience for both first and last authors. The fixed effects estimates suggest that, on average, the youngest first and last author groups use 0.18 and 0.24 more new important concepts than the oldest first and last author groups. Given a sample mean of 0.33, the decline in the use of new important concepts is quite large—55 percent and 73 percent, respectively. Column 4 of Table 4 and Panel B of Figure 3 show that younger authors (first and last) also use more old important concepts than older authors (when author-position fixed effects are included).

The previous results show that the propensity to use new or old *important* concepts declines over a researcher's career. Columns 5 and 6 of Table 4 and Panel C of Figure 3 display the regression results when the outcome variable is the mean age of *all* concepts an article uses in its title or abstract, *regardless of the concepts' importance.* For both author types and regardless of whether author-position fixed effects are included, the mean age of concepts tend to increase with author experience. The fixed effects estimates suggest that the youngest first and last author groups use concepts that are, on average, two and 1.6 years newer than those used by the oldest first and last author groups. While this is a meaningful difference, compared to a standard deviation of 14 years, it does indicate that researchers update the concepts they use substantially over their careers. These differences are about 14 percent and 10 percent of the standard deviation in mean concept age.

Columns 1 and 2 of Table 5 and Online Appendix Figure E.2 Panel A display regression results when the outcome variable is the mean age of an article's references. Like the mean age of concepts used, the mean age of references tends to increase over the career for both first and last authors and for specifications with and without author-position fixed effects. The fixed effects estimates suggest that the youngest first and last author groups reference articles that are, on average, 1.15 and 0.82 years newer than those referenced by the oldest author groups, which are 19 percent and 14 percent of the standard deviation in the mean reference age. Note that the experience of first authors is more strongly related to the age of references than the experience of last authors, which is consistent with the conventional view that first authors play a more dominant role in drafting manuscripts. Columns 1 and 2 of Online Appendix Table E .7 report similar results excluding self-citations.

The range of disciplines cited by an article is a measure of multidisciplinarity. Columns 3 and 4 of Table 5 and Panel B of Online Appendix Figure E.2 display regression results when the outcome variable is the disciplinary diversity of an article's references. For first authors, the specifications with and without author-position fixed effects yield similar results: the multidisciplinarity of articles decreases with experience. The fixed-effects estimates show that the difference in the multidisciplinarity measure between the youngest and oldest first author groups is three, which is about 8 percent of the standard deviation. For last authors without author-position fixed effects, the multidisciplinarity of articles increases with the experience over the first 15 career years and then flattens out. With fixed effects, articles with last authors in the 11–15-year experience group cite the widest range of literature. Our finding of greater experience responsiveness for first authors is again consistent with the conventional role of first authors as the author who writes the manuscript. Columns 3 and 4 of Online Appendix Table E.7 report results excluding self-citations, which increases the gaps between the youngest and oldest experience categories, especially for the last author.

Thus far, we have shown that the references of younger authors tend to be newer and more multidisciplinary than the references of older authors. However, it is worth asking: Are younger authors also more likely to reference "better" research than older authors? If so, are they particularly likely to reference better *recent* research (research that is less than five years old) compared to older research? If there are differences in the quality of work that young and old researchers draw on, it might indicate that younger researchers are better able to distinguish the quality of research or that they take approaches more in keeping with current trends. The last six columns of Table 5 address these questions by using the natural logarithm of the mean number of citations received by all of an article's references as the outcome variable in Equation 1.[34] We regard this outcome as a measure of the quality of research on which an article draws.

Columns 5 and 6 present regression results when the outcome variable is the log of mean citations received by an article's *recent* references (that is, references that are less than five years old) and Columns 7 and 8 show analogous results for citations received by an article's *older* references (that is, references that are more than five years old). For both first and last authors, when author-position fixed effects are omitted, the quality of recent and older references initially rises and then either declines or stabilizes, with peaks occurring earlier for first authors (1–5 years). When author-position fixed effects are included, we see near monotonic declines in the quality of recent and older references over the course of both first and last author careers, with authors in the youngest experience group referencing articles that receive about 16–23 percent more citations than articles referenced by the oldest experience group.

Columns 9 and 10 display regression results when the outcome variable is the log difference in the quality of recent and old references; that is, log(new) – log(old). This allows us to examine how author experience is related to the gap in the quality of references to older and more recent works—a positive coefficient for an experience group suggests that the articles published by this group have recent references that are higher quality relative to older references. When author-position fixed effects are omitted, young first authors tend to have higher quality old references relative to recent references, and young last authors tend to have higher quality recent references relative to old. However, when author-position fixed effects are included, the quality gap does not substantially vary across either first or last author experience groups, suggesting that young authors are not particularly likely to reference important recent works relative to older works.

Though our focus in this section has been on the relationship between article inputs and experience, Tables 4 and 5 also shed light on how article inputs vary across differently structured teams. Thus, teams with more authors, broader academic networks, and authors from less dispersed research locations are more likely to use important concepts (both new and old), use newer concepts (regardless of impact), cite newer references, and have more disciplinary diversity in their references.

Overall, the results in this section suggest that article inputs differ systematically by experience. Specifically, we find that articles written by younger authors demonstrate input choices and strategies that are associated with impactful and original work—they

34. Note that the articles cited by a focal article all have at least one citation by construction.

build on new, promising avenues of inquiry, are expansive in drawing from preceding work, and tap ideas that are fresh and from diverse sources.[35]

C. The Quantity of Output, Best Works, and Author Experience

As noted in the introduction, much of the empirical literature on scientists' careers report an inverted-U shape relation between experience and productivity, which is consistent with a human capital model that predicts worker productivity should increase with age, at least for a portion of the career. However, this literature focuses on the relationship between experience and output *quantity* (for example, Levin and Stephan 1991) or the timing of high-ability researchers' highest quality works (for example, Weinberg and Galenson 2019). In contrast, our work has thus far focused on the relationship between experience and research output and input *quality*, with our quality measures derived from citations, title and abstract text, and references.

To connect our work with previous empirical literature, we use our data to construct an author–year panel that allows us to track the productivity of individual authors through time and to examine the relationship between output quantity (as measured by publication counts) and experience. Specifically, we segment each researcher's career into consecutive five-year intervals, and, for each interval, we construct the following two variables to measure quantity: (i) "unprorated" publication counts in which each author receives full credit for each article and (ii) "prorated" publication counts in which an author receives credit for $1/n$ of an article that has n total authors.[36] These data allow us to estimate the author-level Equation 2, with publication counts as the outcome variable.[37] We also use this author–period data set to identify each author's most highly cited work in each five-year period and assess how the quality of the author's *best* work changes over the career.

Figure 4 and Online Appendix Tables E.9 and E.10 report the results. Whether or not we include individual fixed effects, the estimates show that quantity is hump-shaped in experience. For the unprorated publication counts, productivity peaks are at 31–35 years and 21–25 years, without and with individual fixed effects (Figure 4, Panel A). The

35. In Online Appendix Table E.8, we examine whether measured input choices of authors can explain the negative experience–quality profiles identified in Section IV.A. If so, adding these variables to regressions should attenuate the experience–quality gradient. To do this, we reproduce the regression estimates in Table 2 using total citations and the disciplinary diversity of citations as outcomes, but include, as additional control variables, the number of references, mean reference age, disciplinary diversity of references, new important concept count, old important concept count, and mean concept age as control variables. Adding these control variables has a small effect on the experience–quality gradient, whether measured by total citations or the disciplinary diversity of citations. Thus, although younger authors use newer concepts and reference more recent, important, and diverse works, this behavior does not explain why articles written by younger authors generate more citations and generate citations from a wider variety of fields. This suggests that unobserved differences in the quality of work must account for the decline in the quality of innovation over the career.

36. Our main sample consists of all authors who have at least one research article during this period, and researchers remain in the analytic sample through 2009 or their 41st career year, whichever arrives first. We assign a zero to any gaps in publications up to the earlier of 2009 and the last five-year interval with any publications.

37. As in the article-level regressions above, we first regress our outcome variables (publications) on year and field fixed effects, where each author is assigned the fixed effect associated with the level 4 MeSH with which they were most strongly associated in their first year as a researcher. The left-hand side of the regression is the residual from that regression.

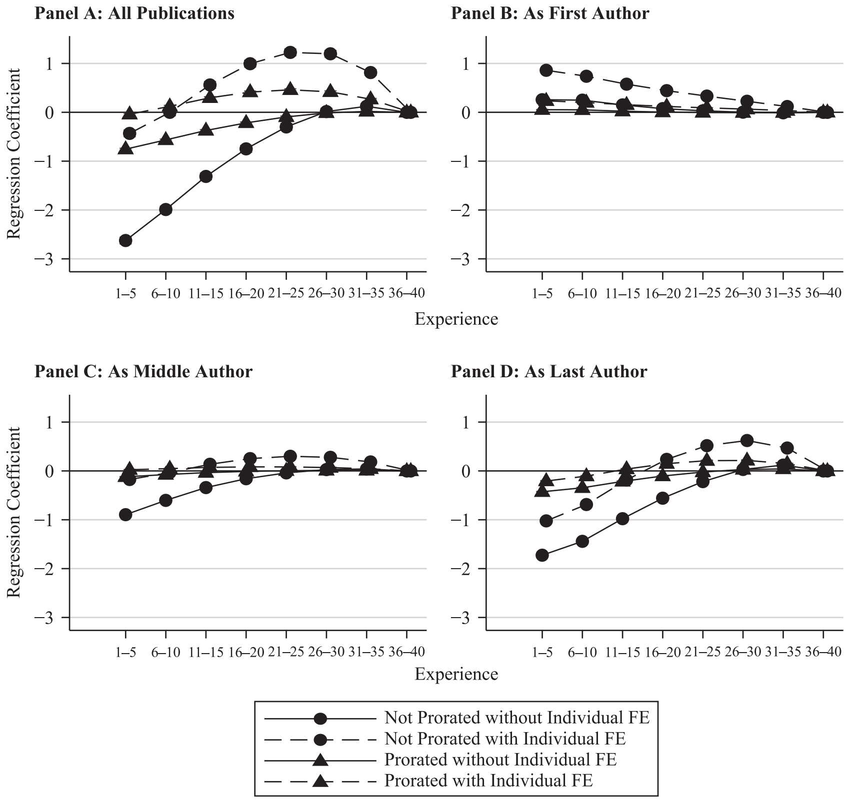

Figure 4

Experience and Quantity (Publication Counts)

Notes: Based on estimates in Online Appendix Tables E.9 and E.10. In the prorated estimates, articles are prorated by the inverse of the number of authors (that is, if an article has five authors, the publication counts as 0.2 publications for all authors).

specification with individual fixed effects implies that, at their peak, authors publish 1.2 more articles than they do at the oldest experience level (36–40 years). When prorated publication counts are used, the profiles remain similarly shaped, but scaled down by the size of teams. The remainder of Figure 4 shows that publications as a first author decline over the career, while publications as middle and last author are hump-shaped, which is consistent with first authors transitioning to last authors as their careers progress.

We can also test whether the effect of experience is different for more accomplished researchers. Online Appendix Table E.11 reproduces the regressions in Columns 1 and 2 of Online Appendix Table E.9 for two subgroups of authors: (i) those who appear at least once as a first author and (ii) those who appear at least once as a last author. The first group excludes authors who never are first authors. The second group is an even more exclusive group of researchers that excludes all authors who never attain last author

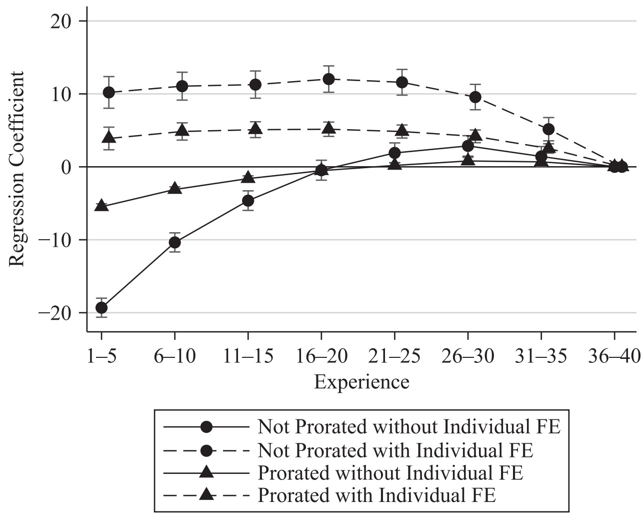

Figure 5

Experience and Article Quality (Best Paper)

Notes: Based on estimates in Online Appendix Tables E.12 and E.13 Columns 1 and 2. The dependent variable is the number of citations received by the most cited paper in each five-year period.

status. The results obtained using these more exclusive groups of authors are qualitatively similar to results in Online Appendix Table E.9, though the peak for authors who attain last author status occurs earlier.

Figure 5 and Online Appendix Table E.12 and E.13 Columns 1 and 2 report results for a researcher's best article in each five-year period. These estimates show that the quality of researchers' *best* work increases over the career when fixed effects are not included, but that the quality of the best work, like the quality of the *average* work, declines over the career once fixed effects are included. The decline is quite large—ten citations (when publications are not prorated)—but the curve is relatively flat for much of the career. The large decline over the last two experience intervals may be partly due to the decline in publishing rates and thus may be partly mechanical (compare to Figure 4).

In addition to results for the best paper, we present results for the second and third best paper and the best "initial *N*" papers in Online Appendix Tables E.12 and E.13, as well as Online Appendix Figure E.3.[38] "Initial *N*" is the researcher's publication count in the first five years of their career. The idea here is to see whether the researcher is producing the same quantity of high-level work each year, but as their experience rises, begins supplementing the core work with additional low-quality research. Overall, the profiles for these papers look similar to the profile for the best paper. However, we are concerned about selection polluting these estimates because there are many authors who do not

38. In the case where an author does not have a second (or third, or *N*th article, we exclude them from the sample for that period.

have multiple papers within a given time period, thus causing this outcome to be undefined for such author–period pairs.

In sum, our analysis of publication rates over the career yields results that are consistent with what others have found in other contexts: the relationship between experience and output *quantity* (publication rates) is hump-shaped. By contrast, the quality of the best work(s) is not hump-shaped—they are roughly flat for the early part of the career and then declining later in the career.

D. Fund Young or Old? The Policy Trade-Off

We have shown that creativity falls with experience and that attrition is ability based. What does this imply about how scarce research positions and resources should be allocated between younger and older scientists? Consider the theoretical framework outlined in Section II. We now outline a simple extension of this model in which science is dependent on funding, with salaries being paid by a single funding agency. The funding agency distributes an annual time-invariant budget that supports N researchers, both young and older, to maximize scientific output over an infinite horizon, the quality-weighted sum of publications. That is, the agency chooses $\left\{N_{yt}, N_{ot}\right\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t F\left(N_{yt}, N_{ot}\right),$$

where

$$\begin{aligned} (4) \qquad F\left(N_{yt}, N_{ot}\right) &= \left[\delta f(\varepsilon_t)\alpha_H + \alpha_L\right] N_{yt}(1-p) + \left[\delta f(\varepsilon_t)\alpha_H + \alpha_H\right] N_{yt} p \\ &= \left[\delta f(\varepsilon_t)\alpha_H + \mu_y\right] N_{yt} \end{aligned}$$

Here β is the agency's discount rate, and μ_y is the mean ability for young researchers. Optimization is subject to $N_{yt}+N_{ot}=N$, which must hold in each period, and old researchers must be recruited from the high-ability young researchers hired in the previous period (so $N_{ot} \le pN_{yt-1}$). The first term on the right-hand side of Equation 4 is the quality of a paper when the young author is an L-type times the number of such papers, and the second term is analogously defined for papers from H-type young authors. Thus, F is the average quality of articles ($[\delta f(\varepsilon_t)\alpha_H + \mu_y]$) times the number of articles written (N_{yt}). As in Section II, only H-type scientists continue to a second period, but here the funding agency actively decides how many to extend and faces a trade-off that the more older scientists it funds, the smaller is the in-coming "class" of young researchers.

This is a dynamic problem in which choices in each period resonate in the future. An exact solution for the optimal paths of N_{yt} and N_{ot} is beyond the scope of the paper, but the solution is nonetheless simple to characterize through inspection of Equation 4, and the constraint that the old researchers must be recruited from the previous period's young. In each period, the funding agency balances the gains against the losses from increasing funding to young scientists. As Equation 4 shows, enlarging N_{yt} compared to N_{ot} increases the quantity of articles published. The benefit of this rises with the expected ability of the young researcher, μ_y. Enlarging N_{yt} also increases the stock of scientists from which to draw the scientific leaders in the next period, which is valuable if the constraint $N_{ot+1} \le pN_{yt}$ binds. But diverting resources from older researchers decreases the effort older scientists allocate to each paper, lowering each paper's quality (because

$f'(\varepsilon_t) > 0$). The loss due to effort reduction is less of a deterrent the more diminished the older researcher's creativity—all else equal, therefore, the lower is δ, the more the funding agency will favor young researchers.

The upshot is that as science is team-based, where teams comprise young researchers working under the guidance of high-ability, older researchers, funding agencies have to be wary of increasing funding to the young. On the one hand, diverting funding to the young increases the quantity of research and increases resources to researchers undiminished by age. It is also an investment in future scientific leadership. On the other hand, it may reduce research quality by diverting resources from researchers who are more tested and may also reduce the effectiveness of their leadership by spreading it more thinly.

V. Discussion and Conclusion

This study systematically investigates the effect of experience on the quality and quantity of scientific output. Our data are more comprehensive and detailed than previous work on scientific productivity, as they include information on the publications, citations, references, and text of almost all biomedical scientists publishing between 1980 and 2009. This allows for relatively broad generalization from our results and a rich characterization of output using a variety of metrics. In addition, the longitudinal structure of our data allows us to control for time-invariant researcher differences, which turns out to be crucial for accurately understanding how scientific productivity evolves over the career. Finally, our data also allow us to estimate the effect of experience separately for different roles played in the research team, for first authors, who design, implement, and write up the work and are often at the beginning of their careers, and last authors, who conceive the research and provide the intellectual guidance and the funding for the research.

Our analysis of the quantity of scientific output yields results that are consistent with the human capital model and much of the empirical literature. Specifically, the quantity of output in biomedical science (as measured by publications) follows an inverted-U relationship in experience, with publication rates peaking well into the career. This result holds regardless of whether we control for time-invariant researcher differences using author fixed effects and mirrors the findings of Levin and Stephan, who, using a similar approach and the same publication count–based measure of productivity from a different data set, find this life-cycle pattern for earth scientists and physicists (though not particle physicists).

A contribution of this work is the employment of data that enables a much richer description of the quality, impact, and composition of the science produced over the life cycle than is possible with publication counts alone. Our analysis of the quality of scientific output tells a more nuanced story of how productivity and human capital change over the career compared to the analysis of publication counts. When we do not control for time-invariant researcher differences, by omitting author-position fixed effects, the quality (like the quantity) of scientific output follows a slight inverted-U relationship with experience. However, when we include author-position fixed effects, the quality of scientific output declines with experience. This highlights the importance of the longitudinal structure of our data and its ability to control for time-invariant researcher differences; the absence of such controls is likely responsible for the hump-shaped or flat quality–experience profiles found in past research. We show that the conflicting results produced from regression specifications with and without author-position fixed effects are the result of ability-biased attrition rates. Lower-ability researchers tend to cease publishing at earlier points in the career, leaving higher-ability researchers to produce a growing share of total publications (which tend to be higher quality). In other words, the mean person publishing when young has a lower-ability fixed effect than the mean person publishing when old, and the improving ability distribution offsets the decline in productivity.

Our analysis of how article inputs change over the career shows that, compared to when they are old, younger authors are more likely to use new important concepts and reference more recent, important, and diverse works. These input choices and strategies are associated with works that have a higher, broader, and more transformative impact. Though we do not find that measured changes in article inputs account for the negative experience–quality profiles, we interpret these results as evidence that, like the quality of scientific output, the quality of scientific human capital falls with experience.

Is the experience–quality relationship we are finding because older researchers are PIs whose responsibilities (fund-raising, mentoring, and administration) have them spread thin? Our finding that age affects even the first authors (who are not PIs) substantially implies that the experience-related decline in quality is not simply due to an increase in distractions that come with transitioning to PI. Of course, we are not able to say to what extent the age effects are due to other life-cycle events, such as family commitments, but it is noteworthy that quantity does increase for much of the career. Because researchers in our data are coming up for tenure at very different times in their careers—due to when they first published, the length of post-docs, and differences and resets in tenure clocks—our data are not well suited to study the impact of academic tenure on research productivity. But the timing of the quantity decline and the continuity of the decline in both input and output quality beyond even after the second decade of publication suggest that other factors are behind these dynamics. Our findings that output and input quality measures are lower for older researchers and that the quality of researchers' best work(s) changes little over most of the career are also inconsistent with a "swing for the fences" effect of tenure.

These results suggest that policies attempting to shift resources from established older scientists to younger scientists should be designed with care. Though the skills of younger scientists may be undiminished by age, the average older scientist has proved their mettle by surviving the test of time, whereas the set of younger scientists to which resources will be shifted are a mixed bag of quality.

Our finding that the experience gradient is steeper for first authors suggests that first authors' contributions are more decisive for the outcome of the project and reception of the work, or that the talents or skills called upon in that position are more affected by experience. It may also reflect the fact that the last author, because they must have attracted funding for the project to go forward, is in a sense prescreened (low-ability, older authors, whose abilities have declined, would not have secured the resources for the project in the first place) in a way that the first author is not.

We recognize that while the metrics that we construct are richer than in most previous studies, they are still limited in that they do not capture all inputs and outputs of the scientific production process. For instance, they do not capture the potentially important effect of older researchers mentoring younger researchers. Nor do these estimates capture any effect of older researchers on the reception of younger researchers' works.

References

Acemoglu, Daron, Ufuk Akcigit, and Murat Celik. 2015. "Young, Restless, and Creative: Openness to Disruption and Creative Innovations." NBER Working Paper 19894. Cambridge, MA: NBER.

Althouse, Benjamin M., Jevin D. West, Carl T. Bergstrom, and Theodore Bergstrom. 2009. "Differences in Impact Factor across Fields and over Time." *Journal of the American Society for Information Science and Technology* 60(1):27–34.

Aquiar, Mark, and Erik Hurst. 2013. "Deconstructing Life Cycle Expenditure." *Journal of Political Economy* 121(3):437–92.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. 2010. "Superstar Extinction." *Quarterly Journal of Economics* 125(2):549–89.

Baerlocher, Mark Otto, Marshall Newton, Tina Gautam, George Tomlinson, and Allan S. Detsky. 2007. "The Meaning of Author Order in Medical Research." *Journal of Investigative Medicine* 55(4):175–80.

Beard, George Miller. 1874. *Legal Responsibility in Old Age: Based on Researches into the Relation of Age to Work.* New York: Russells' American Steam Printing House.

Becker, Gary S. 1975. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition.* Cambridge, MA: NBER.

Bhandari, Mohit, Jason W. Busse, Aabhaya V. Kulkarni, P.J. Devereaux, Pamela Leece, and Gordon H. Guyatt. 2004. "Interpreting Authorship Order and Corresponding Authorship." *Epidemiology* 15(1):125–26.

Blau, David M., and Bruce A Weinberg. 2017. "Why the US Science and Engineering Workforce Is Aging Rapidly." *Proceedings of the National Academy of Sciences of the United States of America* 114(15):3879–84.

Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, and Ryan Chute. 2009. "A Principal Component Analysis of 39 Scientific Impact Measures." *PLoS One* 4(6):e6022.

Brogaard, Jonathan, Joseph Engelberg, and Edward Van Wesep. 2018. "Do Economists Swing for the Fences after Tenure?" *Journal of Economic Perspectives* 32(1):179–94.

Charette, Marc F., Young S. Oh, Christine Maric-Bilkan, Lindsey L. Scott, Charles C. Wu, Matthew Eblen, Katrina Pearson, H. Eser Tolunay, and Zorina S. Galis. 2016. "Shifting Demographics among Research Project Grant Awardees at the National Heart, Lung, and Blood Institute (NHLBI)." *PLoS One* 11(12):e0168511.

Cole, Stephen. 1979. "Age and Scientific Performance." *American Journal of Sociology* 84 (4):958–77.

Costas, Rodrigo, and Maria Bordons. 2011. "Do Age and Professional Rank Influence the Order of Authorship in Scientific Publications? Some Evidence from a Micro-Level Perspective." *Scientometrics* 88(1):145–61.

Costas, Rodrigo, Thed N. van Leeuwen, and María Bordons. 2010. "A Bibliometric Classificatory Approach for the Study and Assessment of Research Performance at the Individual Level: The Effects of Age on Productivity and Impact." *Journal of the American Society for Information Science and Technology* 61(8):1564–81.

Cronin, Blaise, and Cassidy R. Sugimoto, eds. 2014. *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. Cambridge, MA: MIT Press.

Cronin, Blaise, and Cassidy R. Sugimoto, eds. 2015. *Scholarly Metrics under the Microscope: From Citation Analysis to Academic Auditing*. Medford, NJ: Information Today, Inc./ASIST.
Davis, Robert A. 1954. "Note on Age and Productive Scholarship of a University Faculty." *Journal of Applied Psychology* 38(5):318–19.
Deaton, Angus S., and Christine H Paxson. 1994. "Saving, Growth, and Aging in Taiwan." In *Studies in the Economics of Aging*, ed. David A. Wise, 331–62. Chicago: NBER and University of Chicago Press.
Deming, David J., and Kadeem L. Noray. 2018. "STEM Careers and Technological Change." NBER Working Paper 25065. Cambridge, MA: NBER.
Diamond, A.M. 1984. "An Economic Model of the Life-Cycle Research Productivity of Scientists." *Scientometrics* 6(3):189–96.
Falagas, Matthew E., Vrettos Ierodiakonou, and Vangelis G. Alexiou. 2008. "At What Age Do Biomedical Scientists Do Their Best Work?" *FASEB Journal* 22(12):4067–70.
Feist, Gregory J. 2006. "The Development of Scientific Talent in Westinghouse Finalists and Members of the National Academy of Sciences." *Journal of Adult Development* 13(1):23–35.
Foster, Jacob G., Feng Shi, and James Evans. 2021. "Surprise! Measuring Novelty as Expectation Violation." https://doi.org/10.31235/osf.io/pnm9q
Franzoni, Chiara, and Cristina Rossi-Lamastra. 2017. "Academic Tenure, Risk-Taking and the Diversification of Scientific Research." *Industry and Innovation* 24(7):691–712.
Funk, Russel J., and Jason Owen-Smith. 2016. "A Dynamic Network Measure of Technological Change." *Management Science* 63(3):791–817.
Gingras, Yves, Vincent Larivière, Benoî Macaluso, and Jean-Pierre Robitaille. 2008. "The Effects of Aging on Researchers' Publication and Citation Patterns." *PloS One* 3(12):e4048.
Haefele, John W. 1962. *Creativity and Innovation*. New York: Reinhold Publishing Corporation.
Hall, Bronwyn H., Jacques Mairesse, and Laure Turner. 2007. "Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists." *Economics of Innovation and New Technology* 16(2):159–77.
Horner, Karen L., J. Philippe Rushton, and Philip A. Vernon. 1986. "Relation between Aging and Research Productivity of Academic Psychologists." *Psychology and Aging* 1(4):319–24.
Hutchins, B. Ian, Xin Yuan, James M. Anderson, and George M. Santangelo. 2016. "Relative Citation Ratio (RCR):A New Metric that Uses Citation Rates to Measure Influences at the Article Level." *PLoS Biology* 14(9):e1002541.
Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108(3):577–98.
Kaiser, Jocelyn. 2008a. "Zerhouni's Parting Message: Make Room for Young Scientists." *Science* 322(5903):834–35.
———. 2008b. "The Graying of NIH research." *Science* 322(5903):848–49.
Kaltenberg, Mary, Adam B. Jaffe, and Margie F. Lachman. 2021. "Invention and the Life Course: Age Differences in Patenting." NBER Working Paper 28769. Cambridge, MA: NBER.
Kyvik, Svein. 1990. "Age and Scientific Productivity. Differences between Fields of Learning." *Higher Education* 19(1):37–55.
Kyvik, Sevin, and Terje Olsen. 2008. "Does the Aging of Tenured Academic Staff Affect the Research Performance of Universities?" *Scientometrics* 76(3):439–55.
Lehman, Harvey C. 1953. *Age and Achievement*. Princeton, NJ: Princeton University Press.
Levin, Sharon G., and Paula E. Stephan. 1989. "Age and Research Productivity of Academic Scientists." *Research in Higher Education* 30(5):531–49.
———. 1991. "Research Productivity over the Life Cycle: Evidence for Academic Scientists." *American Economic Review* 81(1):114–32.

Marschke, Gerald, Allison Nunez, Bruce A. Weinberg, and Huifeng Yu. 2018. "Last Place? The Intersection of Ethnicity, Gender, and Race in Biomedical Authorship" *American Economic Review Papers and Proceedings* 108(5):222–27.

National Science Board. 2019. "Publications Output: U.S. Trends and International Comparisons." Science & Engineering Indicators 2020. NSB-2020-6 2020. Alexandria, VA: National Science Board, National Science Foundation. https://ncses.nsf.gov/pubs/nsb20206/ (accessed December 6, 2022).

Over, Ray. 1982. "Does Research Productivity Decline with Age?" *Higher Education* 11(5): 511–20.

Packalen, Mikko, and Jay Bhattacharya. 2019. "Age and the Trying Out of New Ideas." *Journal of Human Capital* 13(2):341–73.

Planck, Max. 1949. *Scientific Autobiography and Other Papers.* New York: Philosphical Library.

Rafols, Ismael, and Martin Meyer. 2010. "Diversity and Network Coherence as Indicators of Interdisciplinarity: Case Studies in Bionanoscience." *Scientometrics* 82(2):263–87.

Ross, Matthew B., Gerald Marschke, Huifeng Yu, Joseph Staudt, and Bruce A. Weinberg. 2018. "The Effects of Own and Field Career Age on Research Productivity in Biomedical Science." Working paper.

Seglen, Per O. 1997. "Why the Impact Factor of Journals Should Not Be Used for Evaluating Research." *British Medical Journal* 314(7079):498–502.

Simonton, Dean K. 1997. "Creative Productivity: A Predictive and Explanatory Model of Career Trajectories and Landmarks." *Psychological Review* 104(1):66–89.

Sinatra, Roberta, Dashun Wang, Pierre Deville, Chaoming Song, and Albert-László Barabási. 2016. "Quantifying the Evolution of Individual Scientific Impact." *Science* 35(6312):aaf5239–1-aaf5239–8

Singh, Jasjit. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51(5):756–70.

Sorenson, Olav, Jan W. Rivkin, and Lee Fleming. 2006. "Complexity, Networks and Knowledge Flow." *Research Policy* 35(7):994–1017.

Staudt, Joseph, Huifeng Yu, Robert P. Light, Katy Borner, Gerald Marschke, and Bruce A. Weinberg. 2018. "High-Impact and Transformative Science (HITS) Metrics: Definition, Exemplification, and Comparison." *PLoS One* 13(7):e0200597

Stephan, Paula. 2010. "The Economics of Science." In *Handbook of the Economics of Innovation,* Volume 1, ed. Bronwyn H. Hall and Nathan Rosenberg, 217–73. New York: Elsevier.

Torvik, Vetle I., and Neil R. Smalheiser. 2009. "Author Name Disambiguation in MEDLINE." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 3(3):11.

Torvik, Vetle I., Marc Weeber, Don R. Swanson, and Neil R. Smalheiser. 2005. "A Probablistic Similarity Metric for Medline Records: A Model for Author Name Disambiguation." *Journal of the American Society for Information Science and Technology* 56(2):140–58.

Trajtenberg, Manuel, Rebecca Henderson, and Adam Jaffe. 1997. "University versus Corporate Patents: A Window on the Basicness of Invention." *Economics of Innovation and New Technology* 5(1):19–50.

Verhaeghen, Paul, and Timothy A. Salthouse. 1997. "Meta-Analyses of Age–Cognition Relationships in Adulthood: Estimates of Linear and Nonlinear Age Effects and Structural Models." *Psychological Bulletin* 122(3):231–49.

Wang, Dashun, Chaoming Song, and Albert-Laslo Barabasi. 2013. "Quantifying Long-Term Scientific Impact." *Science* 342(6154):127–32.

Wang, Jian, Reinhilde Veugelers, and Paula Stephan. 2017. "Bias Against Novelty in Science: A Cautionary Tale for Users of Bibliometric Indicators." *Research Policy* 46(8):1416–36.

Weinberg, Bruce, and David Galenson. 2019. "Creative Careers: The Life Cycles of Nobel Laureates in Economics." *De Economist* 167:221–39.

White, J. Chris, Margaret Rush, and Walter T. Schaffer. 2009. "Workforce Modeling for the National Institutes of Health (NIH)." Report. https://report.nih.gov/sites/report/files/docs/Workforce%20Modeling.pdf (accessed December 6, 2022).

Wuchty, Stephan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in the Production of Knowledge." *Science* 316(5827):1036–39.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review* 88(1):290–306.

Zuckerman, Harriet A., and Robert K. Merton. 1973. "Age, Aging and Age Structure in Science." In *The Sociology of Science: Theoretical and Empirical Investigations*, ed. Norman Storer, 497–560. Chicago: University of Chicago Press.