Productivity Differences in Fundamental Life-Sciences Discovery*

Amitabh Chandra† Connie Xu‡

March 20, 2025

Abstract

Fundamental knowledge in the life sciences is generally discovered in universities and hospitals, and has profound implications for welfare by its effect on clinical medicine, new medical innovations, and subsequent scientific knowledge. To assess whether the institution where research is produced causally increases the productivity of scientists, we employ a scientist-mover design and find that $50-60\%$ of the variation in research productivity between researchers is causally attributable to institutions, with more than half of the effect driven by the presence of star researchers. Our results suggest that institutional advantages such as research funding, facilities, collaboration networks, and mentorship are positively correlated with scientific output. The magnitude of these effects has not decreased in the wake of technologies that make cross-institutional collaborations easier.

*We thank, without implicating, George Daley and Amy Wagers from Harvard Medical School for helpful guidance on measuring of knowledge production in the life-sciences, and Harvard Business School for funding this research.

†Harvard University and NBER, amitabh_chandra@harvard.edu

‡Harvard University, conniexu@g.harvard.edu

Fundamental breakthroughs in the life-sciences—including discoveries of DNA, genes, mRNA, antibodies, and CRISPR—have enabled new medicines and deepened our understanding of physiology and disease (Zerhouni, 2005). These advances originate primarily from academic institutions and research hospitals, raising the question of whether some institutions are causally more productive than others in generating fundamental scientific knowledge. This question is especially salient because fundamental knowledge is likely to be underprovided by the market. Understanding the forces that produce these limited discoveries has major implications for social welfare, and can inform how limited public and philanthropic resources should be allocated.

Research location is one mechanism that can enhance scientific productivity by providing scientists with easier access to critical inputs such as talented students, cutting-edge facilities, equipment, mentors, and rich collaboration networks (Furman and Stern, 2011). Geographic proximity to other researchers and institutions facilitates knowledge spillovers and research collaborations (Zucker et al., 1998), suggesting potential benefits from concentrated research ecosystems. However, more productive scientists might self-select into locations based on personal or professional preferences—such as where they trained, desirable climate, or dual-career opportunities—making some places appear more productive simply because of who chooses to work there. Such sorting patterns mean that raw differences in institutional research output could be confounded by individual locational preferences rather than true place-based advantages.

To determine whether the association between where scientists work and their productivity is causal, we implement a movers design. Specifically, we estimate an event study specification of an individual's research productivity around job moves, exploiting variation in the "move size", defined as the difference in average research productivity of the origin and destination institutions. If institutional factors explain 100% of a scientist's productivity, then moving to an institution with (say) a 20% higher average output should increase that scientist's output by 20%. Conversely, if research productivity is entirely an individual-level phenomenon, then relocation to a more or less productive institution will have no effect on one's output. Since the movers design links changes in an individual scientist's productivity to changes in the productivity of their location, we can control for time-invariant individual characteristics and overall time trends. This design also allows us to test whether more productive scientists tend to move to more productive institutions, which would confound the identification of true place effects.

Our study is closely related to Lerner et al. (2024), who pioneered a scientist-mover design of "wandering scholars" to estimate that university-level factors account for one-fifth of the variation across universities in the production of commercially relevant knowledge, as measured by the degree to which patents cite academic research. While we share their focus on life-sciences research, our scope extends beyond commercial applications to examine knowledge creation more broadly. We recognize that fundamental research holds intrinsic value regardless of commercial potential. Moreover, even when research does eventually influence patents, clinical practice, or therapeutics, these impacts often emerge over decades through complex, non-linear pathways (Myers and Lanahan, 2022). This suggests that Lerner et al.'s estimates, which primarily capture direct and immediate connections, may be conservative in measuring the full influence of university environments on scientific productivity. [1]

[1]Another reason for studying knowledge creation in the life-sciences is that it is an industry where the complete life-cycle of R&D– from academic research to patents to clinical trials, to new drug approvals– is observed. This

To perform this analysis, we need to define fundamental science research, measure it in a systematic manner, and attribute it to the institution where it was done. We adopt the fundamental research definition from the National Academy of Medicine[2] This definition encompasses traditional wet laboratory research (such as cell division and protein degradation) as well as early-stage translational research (such as initial testing of new therapies in humans). The definition highlights why fundamental research may be underprovided relative to the social optimal: it often uncovers natural phenomena such as transcription and genes for which patents are impermissible (Kesselheim et al., 2013) or it may be too removed from immediate commercial applications to satisfy the 'useful' criterion (also known as Section 101 criteria) for patentability (Budish et al., 2015). Moreover, disclosing these discovers in scientific journals, creates positive externalities, as other researchers can freely build upon them. The combination of limited appropriability and positive externalities implies that profit-seeking firms will underinvest in fundamental science research, underscoring the importance of understanding the public and nonprofit institutions that produce such research(Nelson, 1959; Akcigit et al., 2021; Azoulay and Li, 2022; Myers and Lanahan, 2022).

We measure research productivity as the frequency with which a scientist is listed as the first two or last two authors on a paper, adjusting for paper citation counts and journal impact factors (we show that these weighting choices are not consequential for our results). Our publication data come from *OpenAlex*, a comprehensive open catalog of global research (Priem et al., 2022). Our sample consists of over 560,000 life-science papers (published 1945-2023) across 15 journals, collectively cited over 65 million times. We track the location of over 891,000 authors (including 331,846 US-based researchers) and their research productivity before and after moving between institutions (we validate this approach by examining the coauthorship patterns of scientists and whether a move predicts a change in coauthorship).

These data reveal that the production of fundamental life-science research is highly concentrated (Carlino and Kerr, 2015). At an aggregate level, the United States (US) dominates the field, producing more than half of the world's output. Notably, the gap in output between the US and its closest competitor in recent times, China, is more than five-fold. Three countries (the US, China and the UK) account for 70 percent of the world's research output. Within the US, aggregate research output is heavily concentrated in a few key regions. California and Massachusetts emerge as dominant hubs, with Boston and the Bay Area producing a disproportionate share of fundamental life-science research. Notably, this concentration reflects aggregate production rather than exclusively individual researcher productivity. While these cities house some of the most productive researchers, our analysis of geographic variation in author-level productivity shows that highly productive scientists are distributed across a diverse set of institutions, including universities, government labs, and biotechnology firms. Moreover, we find that cities producing the most fundamental research also receive more citations in patents down the road, indicating that this research is not only of higher scientific quality (as measured by citations) but also of greater commercial relevance (as measured by paper-to-patent citations). This reinforces the role

would not be true for industries that rely on trade secrets over patents or where fundamental knowledge is not published in journals.

[2]The National Academy of Medicine (NAM) as "new understandings of disease mechanisms gained in the laboratory into the development of new methods for diagnosis, therapy, and prevention and their first testing in humans Sung et al. (2003).

of fundamental research in driving subsequent innovation in the life-sciences.

Our causal estimates show that location indeed matters for scientific productivity. Roughly $50 - 60\%$ of the geographic variation in researcher productivity is attributable to institutional-specific factors. In other words, a scientist's research location has a large independent influence on their research output, even after accounting for individual-level factors like experience and talent. The effect is asymmetric: researchers who relocate to more productive environments experience a slightly larger increase in their productivity than the decline in productivity when a researcher locates to a less productive institution. We further find that over half of the estimated institutional effect is driven by the presence of star researchers, defined here as scientists in the top 5% of their institution's productivity distribution. This prominence of "stars" aligns with prior research that top scientists disproportionately drive knowledge creation. For example, (**?**) finds that the dismissal of star scientists during Nazi Germany suffered especially severe and persistent declines in research output. Similarly (Azoulay et al., 2019) finds a lasting decline in quality-adjusted publication rates for collaborators of deceased stars—a result which we replicate with our data. Finally, we observe that our estimated institutional fixed effects are positively correlated with factors like an institution's R&D spending, its success in attracting research grants and large instrumentation.

Our findings have several implications for science policy and the geography of innovation. An immediate application is to the active debate on the allocation of research funding. Policymakers often face the choice between spreading limited funds broadly across many institutions versus concentrating resources in a few top performing places. Our results indicate that treating all locations as equal recipients would likely be inefficient if the goal is to maximize knowledge production. An institution's causal effect on productivity implies that a dollar of research funding can produce more output at a truly high-performing institution than at an average on. Thus, efforts to distribute research funding without regard to institutional productivity would be an inefficient allocation relative to the goal of producing the most research (but may achieve other social objectives that are unrelated to producing knowledge). At the same time, our evidence shows that individual institutions in smaller or non-traditional locations can be consequential producers of fundamental science— there is little evidence that one must be in a mega-cluster to succeed. This echos the pattern documented by (Ellison et al., 1997) in manufacturing: almost every industry exhibits some geographic clustering, but in many cases the degree of clustering is modest. In the context of science, location matters, but extreme agglomeration is not a universal prerequisite for high research productivity. A broader implication, therefore, is that national scientific output might be raised by cultivating or recruiting star scientists in a wider range of institutions rather than allowing them to concentrate exclusively in a few elite hubs. Spreading top talent more evenly across regions could, in principle, bolster research activity in areas that historically lag behind major knowledge centers. We emphasize however, that such a redistribution is challenging. The leading institutions enjoy profound advantages—prestige, abundant funding, state-of-the-art facilities, and dense networks of junior talent and collaborators— that naturally attract and retain star researchers.

Finally, our study contributes to the growing recognition of the need for targeted support to the scientific workforce, especially early-career researchers (Azoulay et al., 2011, 2019, 2021). If institutional environments causally affect scientists' productivity, as our evidence indicates, then

initiatives like young investigator awards, mentorship programs, or talent grants could yield the greatest returns if focused on enabling talent at the most productive institutions.

I. Data and Measurement

A. Measuring Fundamental Science Research

The NAM definition of fundamental science research encompasses two distinct categories: (1) basic science research, which examines biological and chemical phenomena without specific therapeutic or disease aims—covering studies of anatomy, organisms, and biological processes such as RNA, DNA, genes, cell signaling, and viral replication—and (2) translational science research, which applies these basic insights to medical development and therapeutic applications.[3]

To operationalize this definition of fundamental science research, we query for Medical Subject Headings (MeSH) assigned to every publication in *PubMed*. These are hierarchically organized classifications, created by the NIH National Library of Medicine and serve as a standardized taxonomy for indexing biomedical literature. We query specifically for the MeSH terms *Anatomy, Chemicals and Drugs, Organisms, Phenomena and Processes, Diseases, Anesthesia and Analgesia,* and *Therapeutics*. The predominant MeSH terms represented in our sample are *receptors, DNA, RNA, transcription factors*, and *neurons*, validating our approach of finding fundamental science research.

Our dataset includes journal articles published between 1945[4] and 2023, excluding publications classified as comments, case reports, technical notes, letters, and reviews. We refine the sample to ensure high research quality in a simple and transparent manner by limiting our query to publications from fifteen top life-sciences journals: *Cell*, *Nature*, *Science*, *Nature Cell Biology*, *Nature Genetics*, *Nature Medicine*, *Nature Biotechnology*, *Nature Chemical Biology*, *Nature Neuroscience*, *Neuron*, *Cell Stem Cell*, *PLOS One*, *Oncogene*, *Journal of Biological Chemistry* and *the FASEB Journal*.[5] This selection process yields a core sample of 563,702 publications, authored by 1,433,962 scientists. The proportion of fundamental science articles published in these journals ranges from 37% (Science) to 98% (Oncogene), with remaining articles pertaining to physical and social sciences (see Appendix Table A1).

After obtaining the relevant set of fundamental science papers from *PubMed*, we gather detailed metadata on the authors of these articles using *OpenAlex*, a comprehensive and openly accessible

[3]The NAM definition is focused on the life-sciences and *excludes* all social science research such as health-services research, public-health research, methods research including econometrics, biostatistics and epidemiology, and economics research such as this paper and most of its references. Many excluded articles have public good attributes that are similar to the definition of fundamental research used by NAM.

[4]This is the earliest year in which MeSH terms are reliably assigned to articles in *PubMed*.

[5]We determined this list by asking 5 leading scientists at Harvard Medical School to list the 10 leading fundamental science in the life-sciences and took the intersection of their responses. There was unanimous agreement on the inclusion of 9 journals including *Cell*, *Nature*, and *Science*, and most disagreement on the inclusion of *PLOS One* and *Oncogene*, which were endorsed by only 1 scientist. Articles in clinical journals such as the *Journal of the American Medical Association* (JAMA) or the *New England Journal of Medicine* are excluded from our analyses because the NAM definition of fundamental science excludes clinical studies such as advanced Phase 3 trials.

catalog of the global research system developed by the nonprofit *OurResearch* (Priem et al., 2022). *OpenAlex* catalogs scholarly entities including academic works, authors, institutions, journals, and research concepts, all interlinked for ease of analysis. These data come from aggregating and standardizing information from more than 249,000 sources, such as publication databases like *PubMed* and Microsoft Academic Graph (MAG) as well as metadata database organizers like *Crossref* and *ORCID*.[6] As of the latest data release, *OpenAlex* indexes over 250 million works by 95 million authors affiliated with 109,000 different institutions. These data are updated frequently, often every few hours through its API.

OpenAlex offers several advantages over comparable databases like *PubMed* or *Web of Science* for identifying the geography of research. It employs intentional efforts to disambiguate authors and standardize institutional affiliations. Using the author information from *MAG*, *Crossref*, *PubMed*, *ORCID*, and publisher websites, *OpenAlex* uses an algorithm based on an author's name, their publication record, citation patterns, and (where available) their ORCID to identify all publications belonging to a specific author. For instance, J. Smith and John Smith writing about structural biology would be treated as the same author, while John J. Smith who writes about mathematics would be treated as a different author. OpenAlex also parses the raw affiliation strings from the metadata and employs a machine learning algorithm to extract geographical and institutional information. This allows consistent treatment of affiliations like "MIT, Boston, US" and"Massachusetts Institute of Technology". These features of OpenAlex provide a more comprehensive set of author affiliations throughout our analyzed time frame relative to other sources.

Finally, we use *OpenAlex* to impute the age of fundamental science researchers to understand how institutions affect life-cycle productivity. This is done by identifying the first time that a researcher appears in a paper indexed by *OpenAlex*. We assign a researcher's age to be 25 in the year that they first publish a paper[7].

B. Measuring Scientist's Research Productivity

We move from publications to the individual scientists' research productivity by quantifying an "effective" research contribution measure for each author on a given paper. To do so, we first allocate credit by dividing each paper equally among the number of authors. We then adjust for research impact using paper-specific weights that account for both journal quality and citation influence. Specifically, we assign weights based on the journal's most recent five-year impact factor from Clarivate (2023) and the average annual citation count since publication. These weighting adjustments ensure that papers published in higher-impact journals receive greater weight, as do papers that accrue more citations over time. Importantly, the weighting scheme is designed such that the total sum of research contributions across all authors still equals the total number

[6]Crossref is a nonprofit organization that provides Digital Object Identifiers (DOIs) to index scholarly content which facilities linking metadata, while ORCID (Open Researcher and Contributor ID) is a unique alphanumeric identifier system that helps researchers mange their professional information and publication records.

[7]We validate our 25 at age of first publication assumption with a few checks. Scraping through *OpenAlex* for papers that are listed as dissertations, we find a very linear relationship between our definition of publication birth and the researcher's PhD graduation year. Moreover, the productivity age profile using our definition matches that found in previous research (Jones, 2009)

of papers in our dataset. The mathematical details of this weighting scheme are provided in Appendix Section B. Additionally, the top panel of Appendix Table A2 compares various weighting methods, demonstrating that all approaches yield highly correlated results at the country, city, and institution levels.

For our main analysis, we focus on researchers listed among the first two or the last two authors on a publication. In the life-sciences, authors are typically listed in order of those who contributed the most to the work and writing, while those at the end usually represent the principal investigator(s). This restriction helps control for varying norms around coauthorship across locations, effectively capping the number of credited authors at four. In our robustness section, we re-estimate our main specification using alternative author inclusion criteria, considering both a restricted sample of first and last authors only and a broader sample including all listed authors.

Our focus on the first two and last two authors likely provides a conservative estimate of place-based effects. When we expand the analysis to include middle authors, we find that location-specific influences increase by approximately 10%. This suggests that beginning and end authors tend to be more established researchers and are relatively less affected by local institutional factors, whereas middle authors may be more susceptible to place-based influences.

C. Measuring Institutional Productivity

Identifying the home institution where scientists physically conduct their research is challenging due to the prevalence of multiple simultaneous affiliations. To address this, we infer each author's principal institution by analyzing their full publication history and identifying the most consistent affiliation for any given year. If a unique primary affiliation exists for that year, we assign it. In cases of ties or alternating patterns, we assume institutional continuity. For example, if an author is affiliated with institution A in 2008 and 2010 but not in 2009, we designate institution A as the principal affiliation for all years. If an author lists two affiliations (A and B) in the same year, we pick A if A was the primary affiliation in the previous year; otherwise, we select the one that appears as the primary affiliation in the subsequent year. Finally, if a tie remains, we randomly assign one, affecting around 5% of author-years.

Having assigned each author to a unique institution for each year, we can aggregate productivity to broader geographic levels (e.g. cities and countries). Our definition of cities aims to capture economic areas around the world. Thus, for areas within the US, we use Metropolitan Statistical Areas (MSAs) and, where appropriate, combine major MSAs into integrated labor-markets. For example, we define the Bay Area, CA city as the combination of San Francisco-Oakland-Hayward and San Jose-Sunnyvale-Santa Clara. Similarly, we define the Research Triangle Park, NC as the combination of Durham-Chapel Hill and Raleigh-Cary. In the rest of the paper, we refer to these classifications as "cities".

Finally, to account for differences in productivity trends across life-sciences subfield, we categorize each author-year into one of fifteen life-sciences research fields within each 20-year period in our sample. This approach helps us control for scientists selecting into fields that are advancing at different rates (e.g. genetics versus structural or cancer biology following the human genome mapping). Our dataset spans 80 years and covers subfields including microbiology,

cell signaling, genomics, physiology, and stem cells. We employ a k-means textual clustering algorithm using paper titles, MeSH terms, and abstracts from each author's publications as inputs. We apply this clustering algorithm at the author-year level separately for each 20-year period, allowing research focus and field definitions to evolve over time. Appendix Section C details the tokenized features defining each research field. We incorporate these subfield effects as an alternative regression specification in our robustness exercises and show that our main results do not change.

To illustrate our approach, consider the following two papers that contributed to the development of the CRISPR-Cas9 genome editing technique, a groundbreaking discovery that earned Emmanuelle Charpentier and Jennifer A. Doudna the 2020 Nobel Prize in Chemistry. The first paper, "*A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity*", was published in *Science* in 2012 and has been cited 13,832 times. To construct our measure of research productivity, we first allocate an effective contribution of 0.25 to Martin Jinek (first author) at UC Berkeley (in the Bay Area), Krzysztof Chylinski (second author) at Umeå University, Jennifer A. Doudna (second to last author) at UC Berkeley, and Emmanuelle Charpentier (last author) at Umeå University. In 2012, all four authors were classified as working in the subfield of DNA & Genome Stability[8]. Using the latest 5-year impact factor of *Science* (54.5) and the average citation count per year since publication ($13832/12 \approx 1152$), we reweight our measure of productivity to represent the relative influence of this paper in comparison to other works in our sample. Consequently, these four authors (and their affiliated institutions) receive a recalibrated contribution of 28.62 articles for this publication.

As another example, the paper "*DNA interrogation by the CRISPR RNA-guided endonuclease Cas9*" was published in *Nature* in 2014 and has been cited 1,647 times. We focus on Samuel Sternberg (first author) at UC Berkeley, Sy Redding (second author) at Columbia University, Eric C. Greene (second to last author) at Columbia University, and Jennifer A. Doudna (last author) at UC Berkeley. In 2014, all four authors were classified as working in the subfield of DNA & Genome Stability[9]. After adjusting for *Nature*'s impact factor (60.9) and the average citation count per year since publication ($1647/12 \approx 137$), both Samuel Sternberk and Jennifer A. Doudna (and their affiliated institutions) are given an effective contribution of 3.30 articles for this publication. To put these magnitudes of research productivity into context, Table 1 shows that the average author in an elite set of scientific papers has an annual productivity of 0.39 articles and a lifetime productivity of 1.48 articles (measured up until the end of our time frame in 2023).

[8]Based on our subfield classification, Martin Jinek starts his research in structural biology and is currently focused on genomics. Krzysztof Chylinski starts in DNA & genome stability and now works in ecology and environmental biology. Jennifer Doudna begins in cellular metabolism & bioenergetics and now focuses on cell biology. Emmanuelle Charpentier starts her career working in neuroscience & neurobiology before expanding her research to molecular and cellular biology and genetics and genomics

[9]Throughout his career, Samuel Sternberg, Sy Redding, and Eric Greene mostly stay in the field of DNA & genome stability.

II. Productivity Variation in Fundamental Science Discovery

A. Geographic Variation in Research Productivity

Figure 1 shows the aggregate share of output produced by each country since 1945[10]. The data reveal that the US has consistently dominated fundamental science production, accounting for more than half of the global output and exceeding other countries by a substantial margin.[11] Thus, given the US' large role in fundamental science research, we do not view our restriction to US-based scientists as a major limitation (in our original sample of 563,702 papers, 294,782 were authored by 331,846 US-based researchers).

Within the US, aggregate research production is highly concentrated in a few key metropolitan areas and institutions. Boston and the Bay Area, in particular, account for a disproportionate share of total output, and within these regions, Harvard and Stanford stand out as major contributors to global science. Tables A4 and A5 list out the top 50 institutions by average annual output for US-based institutions and international intstitutions, respectively. Figures 2 and 3 illustrates the persistence of these places in producing a large share of fundamental research output. These patterns highlight that, at the aggregate level, a relatively small number of cities and institutions are responsible for much of the total research production.

However, when shifting focus from aggregate production to individual researcher productivity, we observe a more distributed pattern. To explore geographic variation in individual-level productivity, we average researcher productivity to various geographic levels.[12] Table 2 presents the top 20 cities by average researcher productivity in 2015-2023. Our measure of research productivity is interpreted as the "effective" number of publications for an author. Thus, the average researcher in the Boston Metropolitan area is effectively publishing twice as much as those in the Austin metropolitan area. Table 3 further highlights the top institutions in the ten most productive cities. These results show that the average researcher at Dana-Farber Cancer Institute is effectively producing three four as many publications per year than the average researcher at Lawrence Berkeley National Laboratory. While elite institutions like Harvard and Stanford dominant aggregate production, we observe that many institutions house highly productive researchers, including

[10]The large spike around the 1970s in share of research output from "remaining" countries is the publication of *Cleavage of structural proteins during the assembly of the head of bacteriophage T4* in *Nature* whose sole-author was Swiss scientist Ulrich K. Laemmli in 1970. Laemmli's historic paper, with over 300,000 citations, documents his discovery of SDS polyacrylamide gel electrophoresis— one of the most widely used and important techniques in modern biology

[11]While this fact highlights the pivotal role of the US in producing fundamental science research, also note the emergence of new entrants and the decline of others. As depicted in Figure 1, the decline in the US' share of production from 70% in the early 1990s to 50% today is partially the consequence of China's growing presence in fundamental science production. Since the early 2000s, when China first entered the set of large producers of science, China has made remarkable strides in producing fundamental science research— even surpassing the UK as the second largest producer in recent years. Also noteworthy is the shrinking of the United Kingdom's role from the post-WWII era when it produced about 25% of global science to the most recent period where its share is less than 10%.

[12]When calculating average productivity across author-years within a given geography, we exclude institutions with fewer than 100 authors. For US institutions, we also exclude places with less than 10 movers. These adjustments help remove outlier institutions that may be miscoded or lack sufficient representation.

universities, research organizations, government agencies, and biotechnology companies.

Appendix Tables A6 and A7 provide the list of the top 50 institutions by average researcher productivity for US and non-US institutions, respectively. These results reinforce the distinction between aggregate research output—dominated by a small number of institutions—and researcher-level productivity, which is distributed across a diverse set of institutions.

III. Empirical Strategy

A. Econometric Model

We present an empirical framework using a "movers" design to decompose the observed variation in fundamental research log productivity into place- and person-based components. This approach closely follows prior work using the same identification strategy (Finkelstein et al., 2016; Chetty and Hendren, 2018; Lerner et al., 2024; Keys et al., 2023). Place-based components capture local institutional factors such as firm-level R&D expenditures, while person-based components capture individual characteristics such as educational background and intrinsic motivation.

We model an individual scientist's log productivity, y_{it}, at institution j in year t with a two-way fixed effects specification:

$$y_{it} = \alpha_i + \gamma_j + \tau_t + \epsilon_{it} \tag{1}$$

where α_i is an individual fixed effect, γ_j is an institution fixed effect, and τ_t is a calendar year fixed effect. This model can be expanded to capture observable researcher characteristics by including $x_{it}\beta$ where the researcher-level averages would be $\alpha_i + x_{it}\beta$ rather than α_i. For expositional convenience, we describe the motivating theory with the simplified model, but in the certain regression specifications, we incorporate subfield fixed effects which would fall under $x_{it}\beta$.

To decompose the variation in average log researcher productivity across institutions into place- and person-based components, let $\overline{y_{jt}}$ denote the average of y_{it} across researchers working at institution j in year t, and $\overline{y_j}$ as the average of $\overline{y_{jt}}$ across time. Similarly, $\overline{\alpha_{jt}}$ and $\overline{\alpha_j}$ represent researcher-level averages. The difference in average log productivity between any two institutions j and j' is given by:

$$\overline{y_j} - \overline{y_{j'}} = (\gamma_j - \gamma_{j'}) + (\overline{\alpha_j} - \overline{\alpha_{j'}}).$$

We define shares attributable to the institution and researchers as:

$$S_{institution}(j, j') = \frac{\gamma_j - \gamma_{j'}}{\overline{y_j} - \overline{y_{j'}}}$$

$$S_{researcher}(j, j') = \frac{\overline{\alpha_j} - \overline{\alpha_{j'}}}{\overline{y_j} - \overline{y_{j'}}}.$$

A key identification challenge is that the institution fixed effects, γ_j, are only identified if the sample includes scientist-movers who relocate between institutions. Without movers, we cannot disentangle place effects from individual heterogeneity.

For a causal interpretation of these place effects, we require several assumptions. First, unobserved changes in a scientist's log productivity potential must not be correlated with with the log productivity difference between their origin and destination institutions. This assumption would be violated if "rising stars" systematically relocate from low- to high-productivity institutions. Violations of this kind will appear in the pre-trends of our event study. Second, we assume additive separability of α_i and γ_j in log productivity, implying that person and place characteristics are multiplicative in levels. In other words, a move to a higher-productivity institution has a larger absolute impact on lower-productivity researchers. Finally, we assume that the estimated place effects (γ_j) for movers generalize to non-movers. If this assumption is violated, then our estimates would only apply to the set of researchers who relocate rather than the entire population of researchers.

B. Scientist Movers

We consider a scientist to be a "mover" if they are affiliated with different institutions in consecutive years. Since this geographic panel is constructed from publication history, there is a concern that publication lags could affect the accuracy of identifying author move times. As a validation exercise, we plot the share of a mover's fundamental science coauthors from both their origin and destination institutions over time. In theory, the number of coauthors from the destination institution should not increase prior to the move. In reality, we find that movers begin accumulating coauthors from their destination institution around three years before their first publication officially affiliated with the new institution. Therefore, we define the move year as three years before the first observed publication with the destination institution. Figure 5 illustrates the effect of this adjustment, showing that the share of coauthors from the destination institution only increases after our newly defined move year.

Among US-based researchers, we are able to identify 31,627 scientist-movers. On average, these movers relocate 1.4 times during our timeframe. Due to potential noise in the accuracy of OpenAlex's author disambiguation algorithm and our affiliation imputation method, our empirical analysis focuses on a set of 19,086 researchers who only move once. Table 1 provides summary statistics for movers and non-movers in our sample. On average, movers tend to come from slightly smaller cities but produce more output on average than non-movers. One-time movers represent about 60% of the mover sample and are generally representative of the broader mover population.

C. Event Study Estimation

We implement an event study approach to visually observe the dynamic evolution of place effects on scientist's log productivity and assess the robustness of our assumptions through the pre-trends. We first scale each mover's log productivity to account for the differing productivity levels of origin and destination institutions. This is crucial because not all movers have the same origin and destination location and, without scaling, moves from j to j' would be canceled out by moves from j' to j. Specifically, we define the mover i's "move size" as the difference in average log productivity between institution $d(i)$ (destination) and institution $o(i)$ (origin), excluding mover i

themselves:

$$\delta_i = \overline{y_{d(-i)}} - \overline{y_{o(-i)}}$$

Here, $\overline{y_{d(-i)}}$ is constructed by averaging across all author-years (including non-movers but excluding the mover i themselves) at the destination institution and $\overline{y_{o(-i)}}$ is computed in the same manner for the origin institution. Since we restrict to researchers with one move, each mover is associated a single value of δ_i. Moreover, institution averages are computed across all author-years, which means that researchers having the same destination-origin institution pairing will have the same δ_i[13]

Following (Bronnenberg et al., 2012), we then scale log productivity for mover i as

$$y_{it}^{scaled} = \frac{y_{it} - \overline{y_{o(i)}}}{\delta_i}.$$

Under this scaling, $y_{it}^{scaled} = 0$ when the researcher's log productivity equals their origin institution's average log productivity and $y_{it}^{scaled} = 1$ when it equals their destination institution's average log productivity. Thus, in a simple event study regression

$$y_{it}^{scaled} = \theta_{r(i,t)} + \epsilon_{it},$$

the coefficients $\theta_{r(i,t)}$ would represent the average fraction of the move size δ_i closed by relative year $r(i,t)$. Here, we define relative year $r(i,t)$ as $t - t_i^*$ if mover i relocates in year t_i^*. Thus, $r(i,t) = -1$ represents the last year at the origin institution and $r(i,t) = 0$ represents the first year at the destination location.

In practice, because δ_i is often close to 0 as shown in Figure A1, we do not estimate the simple scaled model directly. Instead, we rearrange our equation to avoid dividing by δ_i as follows:

$$y_{it} = \overline{y_{o(i)}} + \delta_i \theta_{r(i,t)} + \epsilon_{ijft}$$

When taking this equation to our data, we include a set of fixed effect controls and the sample analogue of δ_i to obtain our baseline event study specification of:

$$y_{ijft} = \widetilde{\alpha}_i + \tau_{ft} + \rho_{r(i,t)} + \theta_{r(i,t)}\hat{\delta}_i + \epsilon_{ijft} \tag{2}$$

where τ_{ft} are field-by-year fixed effects and $\rho_{r(i,t)}$ are relative year fixed-effects. We combine $\overline{y_{o(i)}}$ and α_i into a single individual-level fixed effect, $\widetilde{\alpha}_i$. This specification also allows for a common productivity trend around the move (captured by $\rho_{r(i,t)}$) across all movers, irrespective of move size and direction. This is important if we think that there are specific reasons correlated with one's productivity that lead a researcher to relocate when they do.

The parameters of interest from Equation 3 are the relative year coefficients, $\theta_{r(i,t)}$, which we normalize to zero in the year before the move $r(i,t) = -1$. These coefficients measure the

[13]Note that δ_i is a time-invariant measure, calculated as the difference in the average outcome across all time-periods between the destination and origin locations. If the place averages, $\overline{y_{d(-i)}}$ and $\overline{y_{o(-i)}}$, were calculated for every time period to create a time-varying δ_{it}, this would likely introduce endogeneity into the model. Such a time-varying measure could capture contemporaneous shocks or trends in the origin and destination locations that are correlated with both the timing of the move and the individual's outcome, rather than solely reflecting the stable difference in place effects.

percentage change in log productivity in the years around the move scaled by the size and direction of the move, δ_i. We expect the log productivity of one's destination to not be predictive of log productivity before a move, but to be predictive of log productivity after a move. Thus, $\theta_{r(i,t)}$ should be statistically indistinguishable from zero for $r(i,t) < 0$ (no pre-trends) and non-zero following a move. The magnitude of the jump in $\theta_{r(i,t)}$ after a move captures how much place-specific factors influence log productivity on average. In other words, $\theta_{r(i,t)} = 1$ indicates that the mover's log productivity has converged completely to the average at the destination location and would imply that log productivity is completely influenced by place-specific factors. On the other hand, $\theta_{r(i,t)} = 0$ indicates no convergence of the mover's log productivity to that at the destination and would imply no role for place-specific factors. We report robust standard errors clustered at the institution level.

IV. Estimates of Institution and Researcher Effects

A. Baseline Event Study Results

We motivate the event study estimates from the movers analysis by first showing the identifying variation behind our movers specification. Appendix Figure A1 presents the distribution of our key dependent variable, $\hat{\delta}_i$, the difference in average log productivity in a mover's destination and origin institution. By construction, $\hat{\delta}_i = 0$ indicates equal average log productivity at the origin and destination. The standard deviation is 0.77, with a significant number of large moves in either direction. This distribution is centered around -0.087 and is relatively symmetric, indicating an approximately equal number of moves from low-to-high productivity institutions ($\hat{\delta}_i > 0$) as there are moves from high-to-low productivity institutions ($\hat{\delta}_i < 0$). To contextualize a move size of -0.087, the average "wandering scholar" in our sample is moving from Baylor College of Medicine to the University of Chicago. In Table A6, we've listed out the top 50 institutions in the US based on average researcher productivity (in levels, not logs). For illustration purposes, this table has been restricted to institutions with more than 100 researchers and 10 movers between 2015 and 2023 to provide a clearer sense of the institutional variation driving our estimates. However, when running our regression specifications we include all institutions.

As an initial exploration of the impact of moving to a new institution on a scientist's log productivity, we plot the post-move change in scientist log productivity (on the y-axis) against the size of the move (on the x-axis) in Figure 6 . We define a scientist's change in log productivity as their average post-move log productivity minus their average pre-move log productivity. The slope of the fitted line indicates the extent to which moving to higher-productivity institution influences a scientist's log productivity. If place fully explained log productivity differences, the slope would 1. Conversely, if individual factors fully explained the variation, the slope would be 0. Figure 6 shows a slope of 0.72 suggesting that 72% of the variation in scientist log productivity can be attributed to institutional factors, while the remaining 28% is due to individual differences. Moreover, the linear and symmetric nature of the relationship above and below a move size of zero supports our assumption of additive separability between α_i and γ_j in equation 1.

We present event study results from equation 3 in Figure 7a for the 1945–2023 period. The

figure plots estimates of $\theta_{r(i,t)}$ with 95% confidence intervals, normalizing $r(i,t) = -1$ to zero. Standard errors are clustered at the institution level to account for intra-institutional correlation in research log productivity.

The figure shows no significant pre-trend in $\theta_{r(i,t)}$ in the ten years before relocation, supporting our identifying assumption that unobserved person-based productivity factors do not drive the decision to move. Log productivity remains stable until about two years after the move, at which point it exhibits a discontinuous increase to approximately 0.5. This effect remains relatively stable in the following year, peaking at around 0.6. These findings suggest that institutional factors account for roughly 50–60% of the variation in research productivity over time. This estimate is slightly lower than the 72% static estimate from Figure 6, reflecting the dynamic nature of place-specific effects while controlling for observable researcher.

Figure 7b shows the same event-study analysis for movers in the more recent 1995-2023 period. The estimated place effect is of a similar magnitude, indicating that the rise of digital communication and easier long-distance collaboration has not materially reduced the influence of the local institutional environment on productivity.

B. Decomposition Estimates

We leverage the two-way fixed effects model in 3 to perform the additive decomposition exercise. Using our sample of movers and non-movers, we estimate institution specific effects ($\hat{\gamma_j}$). These estimates $\hat{\gamma}_j$ are consistent as long as the previously stated identifying assumptions hold[14].

Table 4 presents our additive decomposition analysis, systematically comparing productivity across institutional groups. The first row is the difference in average researcher productivity between the two groups of institutions: $\hat{y} - \hat{y_{R'}}$. The second row is the difference due to researcher characteristics, $\hat{\alpha_R} - \hat{\alpha_{R'}}$ and the third row is the remaining difference due to institution effects $\hat{\gamma_R} - \hat{\gamma_{R'}}$. The last two rows convert the researcher and institution components into relative shares of the overall difference. Columns in Table 4 represent comparisons between institutions categorized by productivity percentiles, ranked according to their average productivity metrics. Our findings consistently demonstrate that approximately 55% of the observed variation in researcher productivity between institutional groups derives from differential place characteristics, with the remaining 45% attributable to differences in researcher-specific factors. The magnitudes are consistent with the event-study analysis, which suggested a place-based share of 50-60%. Note, that the estimates differ because the additive decomposition exercise is analyzing difference between two groups of institutions rather than averaging the place-based share across all movers and also includes both movers and non-movers in the estimation. Moreover, the stability of the institution share across different partitions is consistent with the linear relationship shown in Figure 8.

Table 5 presents an alternative decomposition approach, examining the proportion of cross-sectional institutional variance in research productivity that can be attributed to researcher versus institutional effects. We quantify these components using the following metrics:

1. The share of cross-institution variance that would be eliminated if all institution effects

[14]To reduce noise in our estimates, we exclude institutions with less than 100 researchers

were equalized: $S^{var}_{institution} = 1 - \frac{Var(\bar{\alpha}_j)}{Var(\bar{y}_j)}$

2. The share of variance that would be eliminated if all researcher effects were equalized: $S^{var}_{researcher} = 1 - \frac{Var(\bar{y}_j)}{Var(\bar{y}_j)}$

Our analysis reveals that equalizing institutional effects would eliminate over 53% of the observed variance, while equalizing researcher effects would eliminate 49% of the variance. These findings underscore the substantial and approximately equal contributions of both institutional environments and researcher characteristics in determining productivity differentials across geographic contexts.

C. Extensions

To understand the underlying mechanisms of how place influences research productivity, we examine whether our effects vary according to the type of move and the characteristics of the researcher.

The first mechanism we explore is the role of move direction. We categorize moves as positive ($\hat{\delta}_i > 0$) if a researcher moves to a higher productivity institution and negative ($\hat{\delta}_i < 0$) if a researcher moves to a lower productivity institution. Higher productivity institutions likely have more resources and better facilities that enable scientists to work at the frontier of the field with more advanced research techniques. Assuming researchers incorporate these benefits into their own research when moving to higher productivity places but don't unlearn these skills when moving to lower productivity places, then we would expect to see larger place effects for positive moves. We illustrate the estimated place effects by move direction in Figure A2. These plots confirm our theory as we see a difference in place effects between scientists moving to higher- versus lower-productivity institutions. The estimated place effect for positive moves is larger and goes from 0.6 to 0.8, whereas the place effect for negative moves is between 0.3 to 0.5.

Given the complementary strand of research on the role of "star" researchers in driving innovation, we also examine the influence of stars in explaining our place effect estimates (Azoulay et al., 2011, 2019; **?**). To do this, we define "stars" as researchers in the top 95th percentile of their institution's productivity distribution and modify our definition of δ_i to represent the difference between the average productivity of stars in the destination and origin institutions. Appendix Figure A1 illustrates the distribution of this revised move size, while Figure 8a presents the estimated place effect using this new definition. The results indicate that stars significantly contribute to the productivity of movers, with the estimated place effect being around 0.4. In other words, stars and everything associated with them, are driving roughly two-thirds of the place effect we see in our baseline results. Similar to (Azoulay et al., 2019), we also explore the impact of coauthor productivity when a star relocates in Figure **??**. While the estimates are noisy, we see that the point estimates range from -0.3 to -0.6, implying that a coauthor's productivity decreases by 30 to 60% when their coauthor leaves and is considered a star researcher.

Motivated by prior research on the life-cycle productivity of researchers and agglomeration effects (Dietz and Bozeman, 2005; Jones, 2010, 2009; Levin and Stephan, 1991; Myers, 2020; Carlino and Kerr, 2015; Kerr and Robert-Nicoud, 2020; Moretti, 2021), we also examine heterogeneity in place effects by mover age and city size. Figures A3 and A4 show similar magnitudes of place

effects, regardless of whether researchers move before or after the age of 40 or whether they relocate to larger or smaller cities. This suggests that neither age nor city size significantly impacts the role of place effects in changing researcher productivity.

To better understand the heterogeneity in place-effects, we follow the standard practice in movers designs and examine correlates of these place-based effects. We project our estimated institution-level fixed effects onto various characteristics obtained from the National Science Foundation's annual Higher Education Research and Development (HERD) survey. Appendix Figure A5 presents results from these bivariate regressions, revealing significant positive correlations between place-specific productivity effects and several institutional factors, including total R&D expenditure, success in attracting competitive research grants, and investment in large-scale scientific instrumentation. These correlations suggest that resource-intensive aspects of the institutional environment may play an important role in shaping scientific productivity.

D. Fundamental Science and Commercial Relevance

So far, we have not established whether fundamental science research has direct implications for commercial activity, as opposed to being knowledge that is fundamental but that is quite distant from inventions. Building on work by (Bryan et al., 2020), (Marx and Fuegi, 2020), we measure the propensity of published research to be cited in patents. This measure of "papers in patents" productivity is similar to our measure of research productivity, except that the paper-specific weights adjust for the number of times the paper has been cited within the body text of patents. Patents cite research papers on the front page of the application ("front-page citations") and in the body of the text ("body citations"). We focus on body citations because prior research has shown that these are more likely to capture an inventor's knowledge and better represent the diffusion of knowledge, as they are included by the inventors themselves (in contrast, front-page citations are typically added by patent lawyers (Bryan et al., 2020)). As before, we rescale these weights to sum to 1 to ensure that the sum of our "papers in patents" measure equals the total number of papers in our sample.

We plot the city-level relationship between fundamental science research productivity and "papers in patents" productivity in Figure 4. The slope is effectively 1, which allows us to infer that cities that produce more fundamental science research are also more likely to produce research papers that are cited in patents. This underscores the commercial and welfare implications of fundamental science research.[15]

[15]We also examined the sensitivity of our results to using front page vs. patent body citations – regardless of the chosen measure, we see in Appendix A3 there is a strong correlation between producing fundamental science research and having it cited in patents at the country, city, and institution-level.

E. Robustness

E.1. Subfield Controls

To address concerns about selection bias related to productivity differences across scientific domains, we extend our baseline event study specification to incorporate field-specific fixed effects. This allows us to account for the possibility that researchers might preferentially select into subfields experiencing rising productivity or benefiting from breakthrough discoveries. By controlling for these field-level trends, we can more confidently isolate the causal effects we seek to measure in our primary analysis. The results of this alternative specification, presented in Figure A6, demonstrate that our core findings remain robust to the inclusion of these field fixed effects. This consistency across specifications provides compelling evidence that our observed effects represent genuine phenomena rather than artifacts of differential selection or field-specific productivity trajectories.

E.2. Alternative Event Study Specification

An alternative to estimating equation 3 is to not include the relative year fixed effects:

$$y_{ijft} = \widetilde{\alpha}_i + \tau_{ft} + \theta_{r(i,t)}\hat{\delta}_i + \epsilon_{ijft} \tag{3}$$

The $\rho_{r(i,t)}$ allowed for a common trend in productivity for all movers regardless of the size and direction of their move. Not including these covariates has little effect on our results as shown in Appendix Figure A7.

E.3. Alternative Author Samples

To assess the robustness of our findings to author selection, we replicate our analysis for (1) first and last authors and (2) all authors. For first and last authors, the estimated institutional share of productivity variation in this sample is slightly smaller, around 40% (Appendix Figure A8, suggesting that place-based effects may play a marginally smaller role among researchers in leading authorship positions. Conversely, when we expand the sample to include all listed authors, the institutional share increases to approximately 60% (Appendix Figure A9), implying that institutional effects are more pronounced when considering the broader set of contributors to scientific output. These findings suggest that while institution-driven productivity differences persist across author roles, their magnitude varies depending on whether the focus is on lead researchers or the entire research team.

V. Conclusion

This study provides new evidence on the critical role of place in influencing the productivity of fundamental science research. By tracking scientists who move between institutions, we estimate that 50-60% of the variation in researcher productivity is attributable to place-specific factors.

These effects have not diminished over time, even as across-institution collaborations have become easier. In short, our findings underscore the significant role of institutional factors in fostering scientific output in the life sciences.

Our results contribute to the broader research agenda on the production function of innovation. While we focus on one key determinant—the location where research happens—this work builds on a rich literature examining other factors that shape research output, including funding mechanisms, team composition, competition, and career incentives(Iaria et al., 2018; Azoulay et al., 2019; Bloom et al., 2020; Myers, 2020; Akcigit et al., 2021; Arora et al., 2021; Hill and Stein, 2023, 2024; Acemoglu, 2023). Our findings complement research on patenting, an adjacent but downstream outcome of fundamental research, which has explored how public funding, university policies, and social factors shape technological development(Azoulay et al., 2013, 2019; Myers and Lanahan, 2022; Gross and Sampat, 2023; Moretti et al., 2023; Jacob and Lefgren, 2011; Jensen and Thursby, 2001; Thursby et al., 2007; Lerner et al., 2024; Whittington, 2009; Koning et al., 2020; Bell et al., 2019; Williams, 2017; Bryan and Williams, 2021). If the forces that influence patenting also shape fundamental research—as they are often undertaken by the same individuals and likely complement one another—then the role of place in scientific productivity may be even larger than we estimated.

Our work also opens several avenues for future research. A key question is understanding what factors influence the mobility of scientists. Prior studies have pointed to the importance of funding availability, personal circumstances, and institutional support systems. Understanding these drivers could inform policies to incentivize beneficial moves. Another area is causally identifying the institutional characteristics that make a place effective at fostering research. We provide preliminary descriptive evidence of some factors that are positively correlated with highly productive places. Unpacking this black box and exploring how different institutions structure their research environments and the impact of these structures on productivity can inform best practices for maximizing the effectiveness of research funding.

In conclusion, fundamental science is the cornerstone of long-term medical innovation, and our study highlights that where science is done significantly influences how much and how impactful research is produced. While continuing to support the existing hubs of innovation is likely efficient in the short run, there is also a tremendous opportunity in enabling a broader range of institutions to become more productive by attracting talent and adopting best practices. A more geographically and institutionally dispersed network of star researchers—supported by adequate resources—could accelerate scientific discovery and its translation into societal benefits, while also ensuring that the progress and its rewards are more broadly shared. The challenge for policymakers and research leaders is to build initiatives that harness the power of place, thereby creating a more inclusive and productive scientific enterprise.

References

Acemoglu, D. (2023, May). Distorted Innovation: Does the Market Get the Direction of Technology Right? *AEA Papers and Proceedings 113*, 1–28.

Akcigit, U., D. Hanley, and N. Serrano-Velarde (2021, January). Back to Basics: Basic Research Spillovers, Innovation Policy, and Growth. *The Review of Economic Studies 88*(1), 1–43.

Arora, A., S. Belenzon, and L. Sheer (2021, March). Knowledge Spillovers and Corporate Investment in Scientific Research. *American Economic Review 111*(3), 871–898.

Azoulay, P., C. Fons-Rosen, and J. S. Graff Zivin (2019, August). Does Science Advance One Funeral at a Time? *American Economic Review 109*(8), 2889–2920.

Azoulay, P., E. Fuchs, A. P. Goldstein, and M. Kearney (2019, January). Funding Breakthrough Research: Promises and Challenges of the "ARPA Model". *Innovation Policy and the Economy 19*, 69–96. Publisher: The University of Chicago Press.

Azoulay, P., J. Graff Zivin, and B. N. Sampat (2011, January). The Diffusion of Scientific Knowledge Across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2011). Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics 42*(3), 527–554. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-2171.2011.00140.x.

Azoulay, P., J. S. Graff Zivin, and G. Manso (2013, January). National Institutes of Health Peer Review: Challenges and Avenues for Reform. *Innovation Policy and the Economy 13*, 1–22. Publisher: The University of Chicago Press.

Azoulay, P., W. H. Greenblatt, and M. L. Heggeness (2021, November). Long-term effects from early exposure to research: Evidence from the NIH "Yellow Berets". *Research Policy 50*(9), 104332.

Azoulay, P. and D. Li (2022, March). Scientific Grant Funding. In *Innovation and Public Policy*, National Bureau of Economic Research Conference Report, pp. 117–150. University of Chicago Press.

Bell, A., R. Chetty, X. Jaravel, N. Petkova, and J. Van Reenen (2019, May). Who Becomes an Inventor in America? The Importance of Exposure to Innovation*. *The Quarterly Journal of Economics 134*(2), 647–713.

Bloom, N., C. I. Jones, J. Van Reenen, and M. Webb (2020, April). Are Ideas Getting Harder to Find? *American Economic Review 110*(4), 1104–1144.

Bronnenberg, B. J., J.-P. H. Dubé, and M. Gentzkow (2012, May). The Evolution of Brand Preferences: Evidence from Consumer Migration. *American Economic Review 102*(6), 2472–2508.

Bryan, K. A., Y. Ozcan, and B. Sampat (2020, May). In-text patent citations: A user's guide. *Research Policy 49*(4), 103946.

Bryan, K. A. and H. L. Williams (2021, January). Chapter 13 - Innovation: market failures and public policies. In K. Ho, A. Hortaçsu, and A. Lizzeri (Eds.), *Handbook of Industrial Organization*, Volume 5 of *Handbook of Industrial Organization, Volume 5*, pp. 281–388. Elsevier.

Budish, E., B. N. Roin, and H. Williams (2015, July). Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials. *American Economic Review 105*(7), 2044–2085.

Carlino, G. and W. R. Kerr (2015). Agglomeration and innovation. *Handbook of regional and urban economics 5*, 349–404. ISBN: 1574-0080 Publisher: Elsevier.

Chetty, R. and N. Hendren (2018, August). The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates*. *The Quarterly Journal of Economics 133*(3), 1163–1228.

Clarivate (2023). 2022 Journal Impact Factor, Journal Citation Reports. Technical report.

Dietz, J. S. and B. Bozeman (2005, April). Academic careers, patents, and productivity: industry experience as scientific and technical human capital. *Research Policy 34*(3), 349–367.

Ellison, S. F., I. Cockburn, Z. Griliches, and J. Hausman (1997). Characteristics of Demand for Pharmaceutical Products: An Examination of Four Cephalosporins. *The RAND Journal of Economics 28*(3), 426–446. Publisher: [RAND Corporation, Wiley].

Finkelstein, A., M. Gentzkow, and H. Williams (2016, November). Sources of Geographic Variation in Health Care: Evidence From Patient Migration*. *The Quarterly Journal of Economics 131*(4), 1681–1726.

Furman, J. L. and S. Stern (2011, August). Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *American Economic Review 101*(5), 1933–1963.

Gross, D. P. and B. N. Sampat (2023, November). The World War II crisis innovation model: What was it, and where does it apply? *Research Policy 52*(9), 104845.

Hill, R. and C. Stein (2023). Scooped! Estimating Rewards for Priority in Science.

Hill, R. and C. Stein (2024). Race to the Bottom: Competition and Quality in Science.

Iaria, A., C. Schwarz, and F. Waldinger (2018, May). Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science*. *The Quarterly Journal of Economics 133*(2), 927–991.

Jacob, B. A. and L. Lefgren (2011, October). The impact of research grant funding on scientific productivity. *Journal of Public Economics 95*(9), 1168–1177.

Jensen, R. and M. Thursby (2001, March). Proofs and Prototypes for Sale: The Licensing of University Inventions. *American Economic Review 91*(1), 240–259.

Jones, B. F. (2009, January). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *The Review of Economic Studies 76*(1), 283–317.

Jones, B. F. (2010, February). Age and Great Invention. *The Review of Economics and Statistics 92*(1), 1–14.

Kerr, W. R. and F. Robert-Nicoud (2020, August). Tech Clusters. *Journal of Economic Perspectives 34*(3), 50–76.

Kesselheim, A. S., R. M. Cook-Deegan, D. E. Winickoff, and M. M. Mello (2013, August). Gene Patenting—The Supreme Court Finally Speaks. *The New England journal of medicine 369*(9), 869–875.

Keys, B. J., N. Mahoney, and H. Yang (2023, January). What Determines Consumer Financial Distress? Place- and Person-Based Factors. *The Review of Financial Studies 36*(1), 42–69.

Koning, R., S. Samila, and J.-P. Ferguson (2020, May). Inventor Gender and the Direction of Invention. *AEA Papers and Proceedings 110*, 250–254.

Lerner, J., H. J. Manley, C. Stein, and H. L. Williams (2024, January). The Wandering Scholars: Understanding the Heterogeneity of University Commercialization.

Levin, S. G. and P. E. Stephan (1991). Research Productivity Over the Life Cycle: Evidence for Academic Scientists. *The American Economic Review 81*(1), 114–132. Publisher: American Economic Association.

Marx, M. and A. Fuegi (2020). Reliance on science: Worldwide front-page patent citations to scientific articles. *Strategic Management Journal 41*(9), 1572–1594. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.3145.

Moretti, E. (2021, October). The Effect of High-Tech Clusters on the Productivity of Top Inventors. *American Economic Review 111*(10), 3328–3375.

Moretti, E., C. Steinwender, and J. Van Reenen (2023, February). The Intellectual Spoils of War? Defense R&D, Productivity, and International Spillovers. *The Review of Economics and Statistics*, 1–46.

Myers, K. (2020, October). The Elasticity of Science. *American Economic Journal: Applied Economics 12*(4), 103–134.

Myers, K. and L. Lanahan (2022). Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy. *American Economic Review 112*(7), 239202423.

Nelson, R. R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy 67*(3), 297–306. Publisher: University of Chicago Press.

Priem, J., H. Piwowar, and R. Orr (2022, June). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. arXiv:2205.01833 [cs].

Sung, N. S., W. F. Crowley, Jr, M. Genel, P. Salber, L. Sandy, L. M. Sherwood, S. B. Johnson, V. Catanese, H. Tilson, K. Getz, E. L. Larson, D. Scheinberg, E. A. Reece, H. Slavkin, A. Dobs, J. Grebb, R. A. Martinez, A. Korn, and D. Rimoin (2003, March). Central Challenges Facing the National Clinical Research Enterprise. *JAMA 289*(10), 1278–1287.

Thursby, M., J. Thursby, and S. Gupta-Mukherjee (2007, August). Are there real effects of licensing on academic research? A life cycle view. *Journal of Economic Behavior & Organization 63*(4), 577–598.

Whittington, K. B. (2009, August). 6. Patterns of Male and Female Scientific Dissemination in Public and Private Science. In *6. Patterns of Male and Female Scientific Dissemination in Public and Private Science*, pp. 195–228. University of Chicago Press.

Williams, H. L. (2017, August). How Do Patents Affect Research Investments? *Annual Review of Economics 9*(Volume 9, 2017), 441–469. Publisher: Annual Reviews.

Zerhouni, E. A. (2005, September). US Biomedical ResearchBasic, Translational, and Clinical Sciences. *JAMA 294*(11), 1352–1358.

Zucker, L. G., M. R. Darby, and M. B. Brewer (1998). Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises. *The American Economic Review 88*(1), 290–306. Publisher: American Economic Association.

Table 1: Summary Statistics

	Mean	Std. Dev.
A. Individual-Level		
Everyone (N = 182,921)		
Years Actively Publishing	1.79	2.02
Average Team Size	3.62	0.66
Annual City Size	7352.16	7395.00
Annual Productivity	0.39	2.08
Lifetime Productivity	0.79	4.50
Number of Moves	0.13	0.33
Non-Movers (N = 151,294)		
Years Actively Publishing	1.48	1.50
Average Team Size	3.63	0.67
Annual City Size	7333.28	7558.65
Annual Productivity	0.38	2.16
Lifetime Productivity	0.59	3.49
Movers (N = 31,627)		
Years Actively Publishing	3.90	3.41
Average Team Size	3.57	0.56
Annual City Size	7481.27	6159.27
Annual Productivity	0.48	1.39
Lifetime Productivity	2.11	8.56
One-time Movers (N = 19,086)		
Years Actively Publishing	4.11	3.56
Average Team Size	3.56	0.55
Annual City Size	7489.07	6139.04
Annual Productivity	0.49	1.32
Lifetime Productivity	2.23	9.07
B. City-Level		
Number of Institutions	173.04	177.53
Annual City Size	7792.41	8124.42
Annual Productivity	0.38	2.03
Lifetime Productivity	46.07	469.51

Notes: Table shows summary statistics for our sample of authors in the time period 1945-2023. Panel A presents the individual-level statistics and Panel B presents the city-level statistics.

Table 2: Cities with Most Productive Researchers on Average, 2015-2023

Rank	City	Average Author Productivity
1	Boston-Cambridge-Newton, MA-NH	1.57
2	Bay Area, CA	1.56
3	Trenton, NJ	1.42
4	Dallas-Fort Worth-Arlington, TX	1.34
5	New York-Newark-Jersey City, NY-NJ-PA	1.21
6	San Diego-Carlsbad, CA	1.04
7	New Haven-Milford, CT	1.01
8	Seattle-Tacoma-Bellevue, WA	0.99
9	Worcester, MA-CT	0.98
10	St. Louis, MO-IL	0.97
11	Ithaca, NY	0.84
12	Houston-The Woodlands-Sugar Land, TX	0.83
13	Austin-Round Rock, TX	0.79
14	Los Angeles-Long Beach-Anaheim, CA	0.76
15	Memphis, TN-MS-AR	0.74
16	Santa Cruz-Watsonville, CA	0.73
17	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.73
18	Baltimore-Columbia-Towson, MD	0.67
19	Chicago-Naperville-Elgin, IL-IN-WI	0.60
20	Rochester, MN	0.58

Notes: Table shows the top 20 cities by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the total number of researchers in each location. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal* between 2015 and 2023.

Table 3: Leading Institutions in the Ten Most Productive Cities, 2015-2023

Rank	Institution	Average Researcher Productivity
I. Boston-Cambridge-Newton, MA-NH		
1.	Broad Institute	3.39
2.	Dana-Farber Cancer Institute	2.73
3.	Massachusetts Institute of Technology	2.29
4.	Harvard University	1.59
5.	Mass General Brigham	1.53
II. Bay Area, CA		
1.	Stanford University	1.87
2.	Gladstone Institutes	1.74
3.	University of California, Berkeley	1.32
4.	University of California, San Francisco	1.30
5.	Lawrence Berkeley National Laboratory	0.67
III. Trenton, NJ		
1.	Princeton University	1.42
IV. Dallas-Fort Worth-Arlington, TX		
1.	The University of Texas Southwestern Medical Center	1.34
V. New York-Newark-Jersey City, NY-NJ-PA		
1.	Memorial Sloan Kettering Cancer Center	2.22
2.	Novartis (United States)	1.77
3.	Icahn School of Medicine at Mount Sinai	1.69
4.	Cold Spring Harbor Laboratory	1.42
5.	Rockefeller University	1.34
VI. San Diego-Carlsbad, CA		
1.	Salk Institute for Biological Studies	1.72
2.	University of California, San Diego	0.99
3.	Scripps Research Institute	0.91
4.	Sanford Burnham Prebys Medical Discovery Institute	0.53
VII. New Haven-Milford, CT		
1.	Yale University	1.01
VIII. Seattle-Tacoma-Bellevue, WA		
1.	Allen Institute	2.37
2.	Fred Hutch Cancer Center	1.42
3.	University of Washington	0.82
IX. Worcester, MA-CT		
1.	University of Massachusetts Chan Medical School	0.98
X. St. Louis, MO-IL		
1.	Washington University in St. Louis	1.06
2.	Saint Louis University	0.40

Notes: Table shows the leading institutions in the ten most productive cities by average researcher productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers in each geography. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal* between 2015 and 2023.

Table 4: Additive Decomposition of Researcher Productivity

	Top & Bottom			
	5%	**10%**	**25%**	**50%**
Difference in Researcher Productivity				
Overall	3.54	2.82	1.86	1.11
Researchers	1.66	1.25	0.83	0.54
Institutions	1.87	1.57	1.03	0.58
Share of Difference Attributable to:				
Researchers	0.47	0.44	0.45	0.48
Institutions	0.53	0.56	0.55	0.52

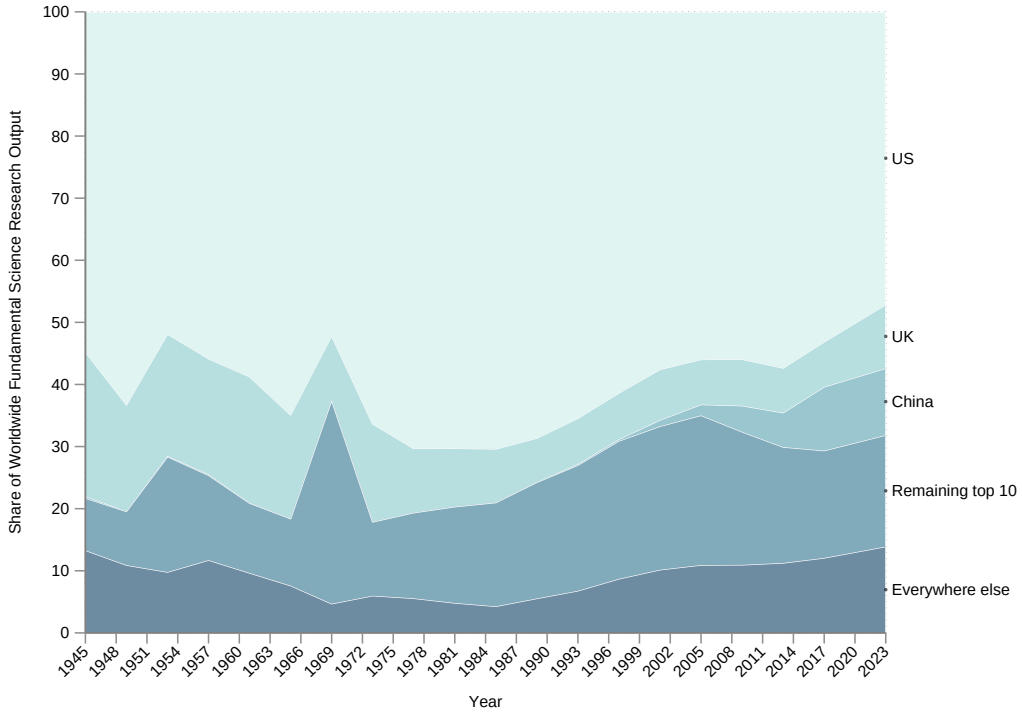
Notes: This table decomposes the difference in researcher productivity across top and bottom percentile groups. The first panel presents the absolute difference in productivity, while the second panel shows the share of this difference attributable to individual researchers versus institutions. Sample includes all authors (movers and non-movers) from 1945-2023.

Table 5: Variance Decomposition of Research Productivity

Cross-Institution Variance of Average Productivity	
Productivity	0.67
Institution effects	0.34
Researcher effects	0.31
Correlation of Average Institution and Researcher effects	0.02
Share of Variance would be Reduced if:	
Researcher effects were made equal	0.49
Institution effects were made equal	0.53

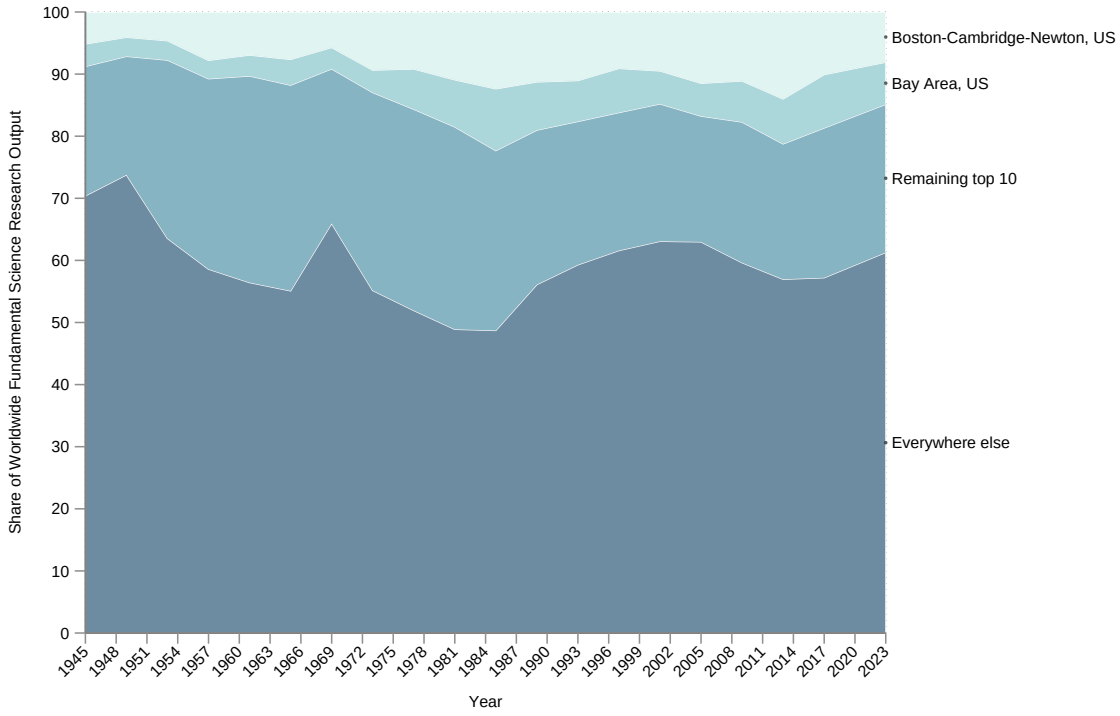
Notes: This table presents the variance decomposition of research productivity across institutions. The first panel reports the overall variance and its components, while the second panel illustrates the share of variance that would be reduced if researcher or institution effects were made equal. Sample includes all authors (movers and non-movers) from 1945-2023.

Figure 1: Country-Level Trends in Fundamental Science Production, 1945-2023



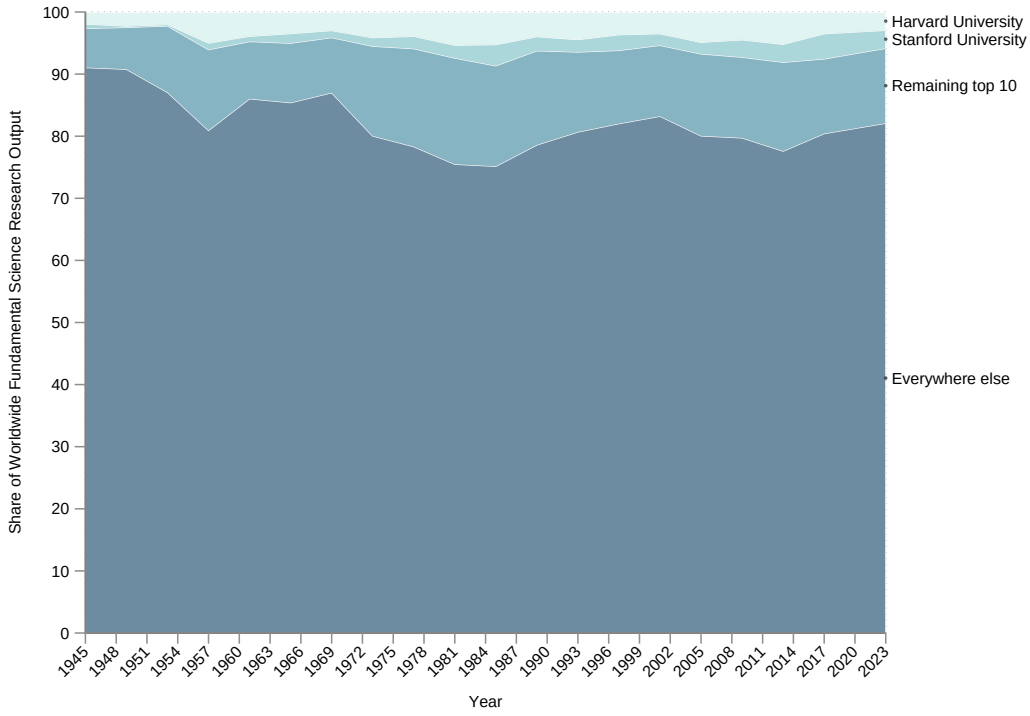
Notes: Figure shows the country-level trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal).* Our measure of productivity has been adjusted for number of authors, citation count, and the journal's latest 5-year impact factor. We only count a paper toward a scientist's productivity if they are either the first or last author listed on a publication. Shares are shown separately for the top two producers (as well as China), the remaining top 10 producers, and everywhere else.

Figure 2: City-Level Trends in Fundamental Science Production, 1945-2023



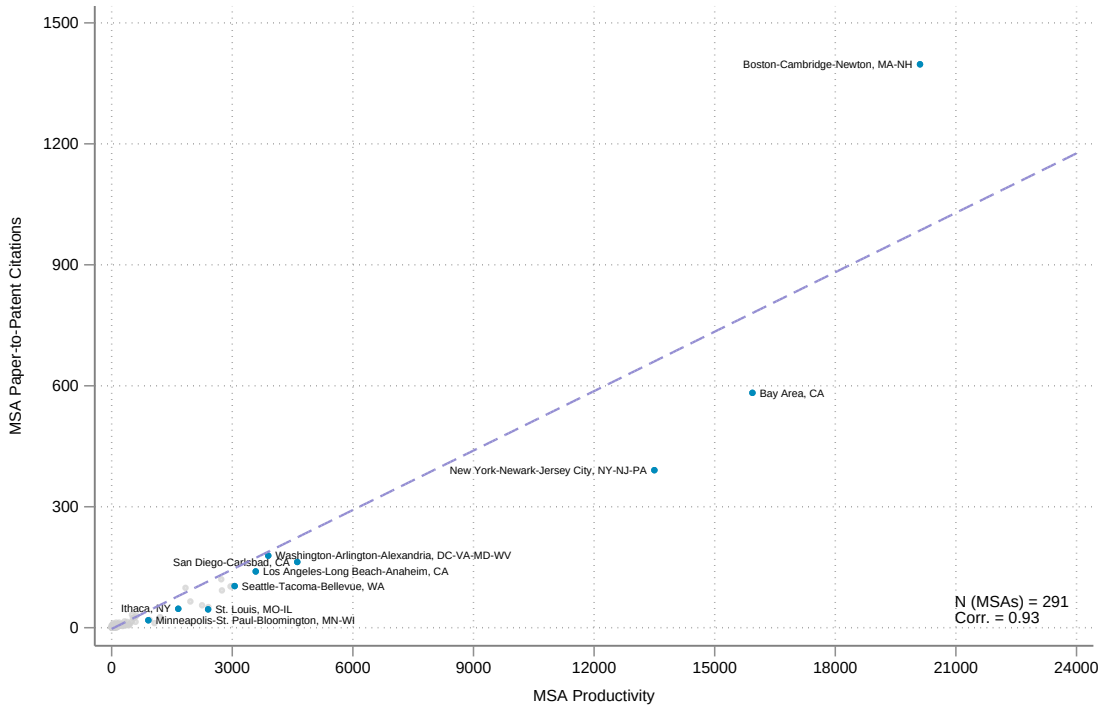
Notes: Figure shows the city-level trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal).* Our measure of productivity has been adjusted for number of authors, citation count, and the journal's latest 5-year impact factor. We only count a paper toward a scientist's productivity if they are either the first or last author listed on a publication. Shares are shown separately for the top two producers (as well as China), the remaining top 10 producers, and everywhere else.

Figure 3: Institution-Level Trends in Fundamental Science Production, 1945-2023



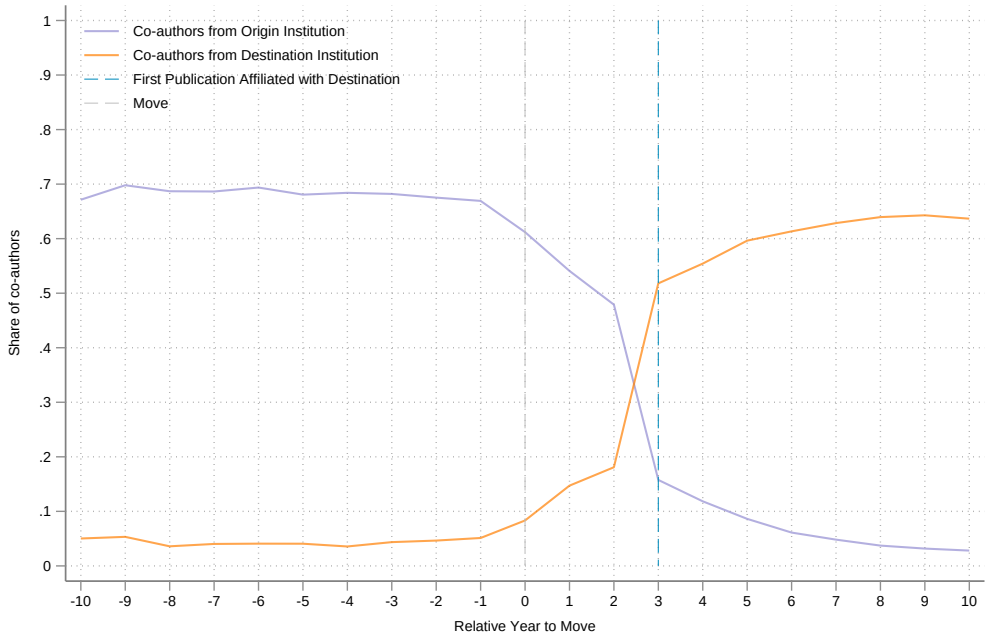
Notes: Figure shows the institution-level trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal).* Our measure of productivity has been adjusted for number of authors, citation count, and the journal's latest 5-year impact factor. We only count a paper toward a scientist's productivity if they are either the first or last author listed on a publication. Shares are shown separately for the top two producers (as well as China), the remaining top 10 producers, and everywhere else.

Figure 4: Fundamental Science Research Cited in Patents, 1945-2023



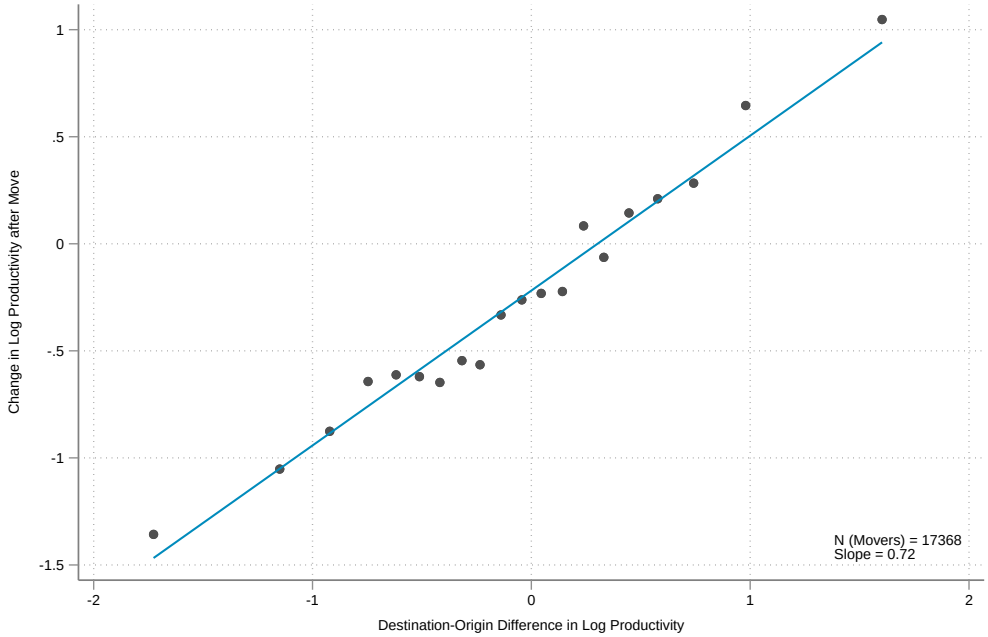
Notes: Figure shows a scatter plot of the number of papers in patents citations against the citation-adjusted measure of fundamental science research produced at the MSA-level. Around 15.5% of papers in our sample have ever been cited in a patent. There are N = 291 MSAs in our sample.

Figure 5: Share of Fundamental Life Science Coauthors from Origin vs. Destination Institution



Note: Figure plots the share of coauthors that are affiliated with the mover's origin and destination institution over time. Move year is mechanically defined as three year prior to the first publication at the mover's destination.

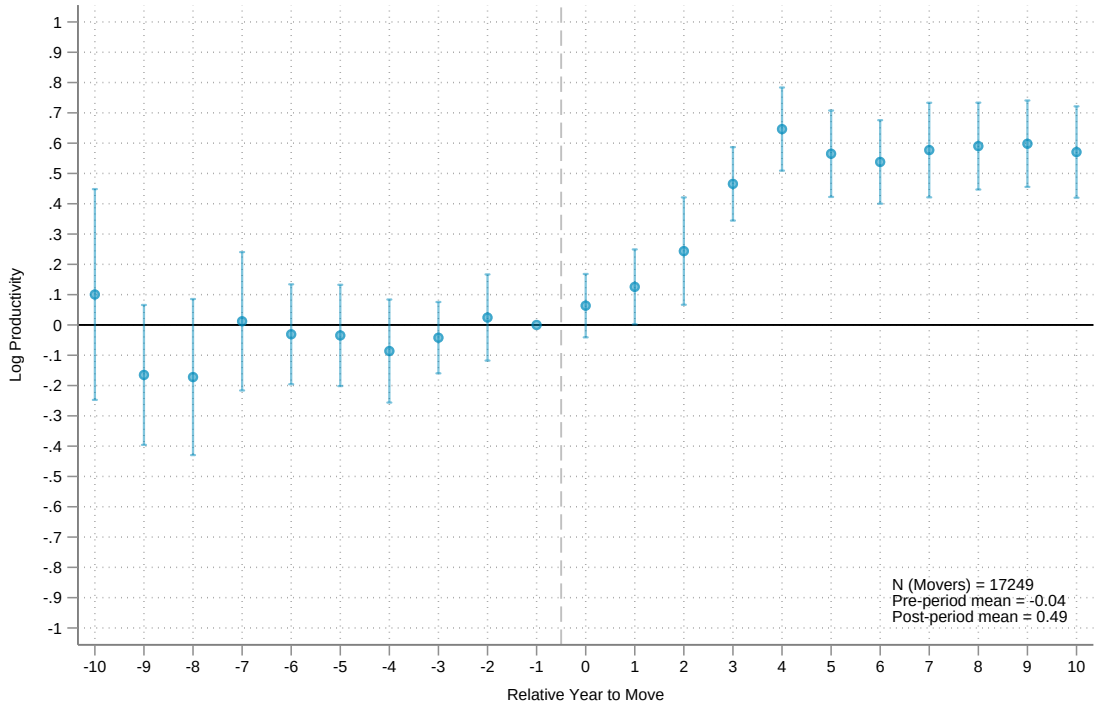
Figure 6: Change in Log Productivity by Move Size



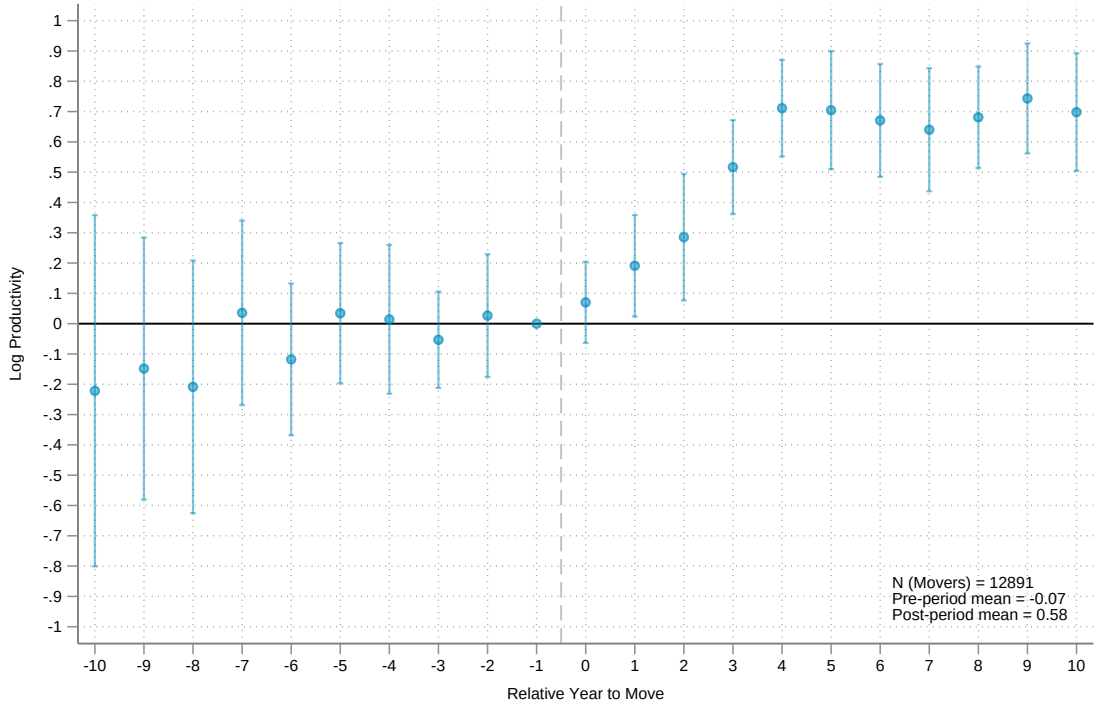
Notes: Figure shows the change in log productivity relative to a mover's productivity move size. Move size is defined as the difference in average productivity between a mover's origin and destination institution.

Figure 7: Productivity Change in Response to Move

(a) 1945-2023



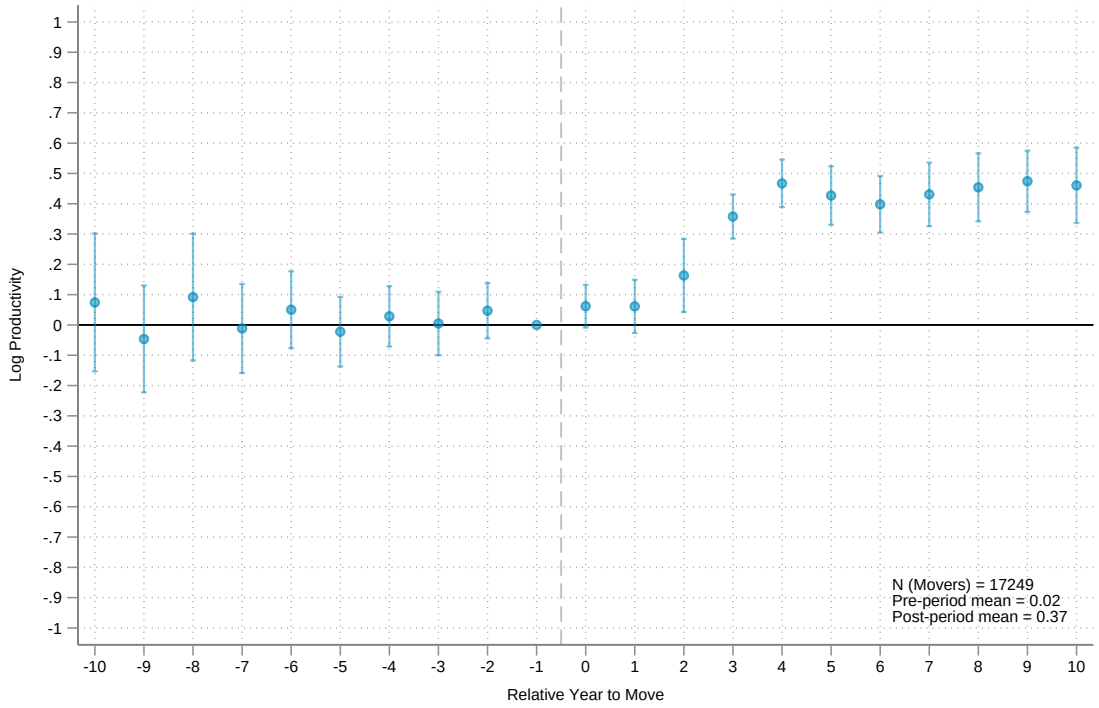
(b) 1995-2023



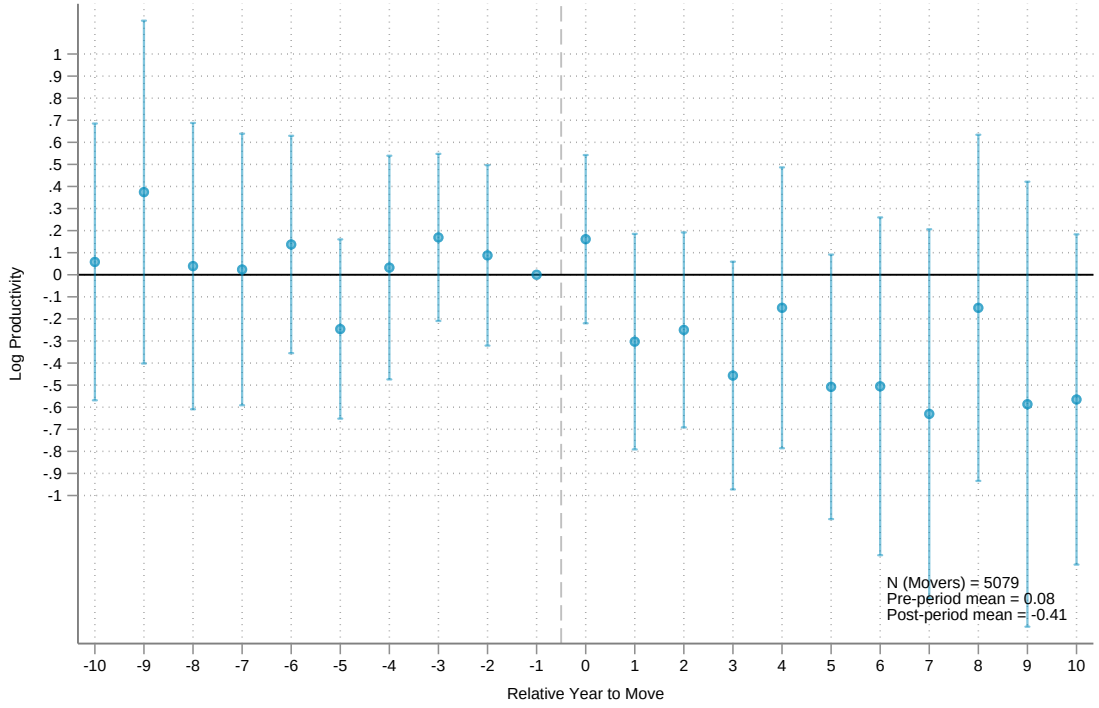
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023. Panel (a) shows the results for movers publishing anytime between 1945-2023 and panel (b) shows the results for movers publishing between 1995-2023.

Figure 8: Impact of Star Researchers

(a) Move Size defined by Star Researcher Productivity



(b) Role of Star Leaving on Coauthor Productivity



Appendix

A. Additional Exhibits

Appendix Table A1: Journal Sample Sizes

Sample	Journal	% Fundamental	# Fundamental
CNS	Science	37.11	39,470
	Nature	64.42	64,093
	Cell	93.92	16,167
Leading Life-Sciences Journals	Neuron	93.16	8,465
	Nature Genetics	90.59	5,775
	Nature Medicine	57.11	4,625
	Nature Biotechnology	43.16	3,447
	Nature Neuroscience	87.08	4,252
	Nature Cell Biology	84.97	3,182
	Nature Chemical Biology	78.45	2,235
	Cell Stem Cell	89.24	1,642
Other Life-Sciences Journals	PLOS One	81.45	210,349
	Journal of Bio Chem	95.09	171,724
	Oncogene	97.93	15,606
	The FASEB Journal	69.08	12,722

Notes: Table shows the share of fundamental science research published in each journal. Our full sample for each journal consists of all articles and clinical trials that appear in OpenAlex between 1945 and 2023.

Appendix Table A2: Measure Weighting Robustness

	Country-Level	City-Level	Institution-Level
Research Productivity			
Unweighted vs. Impact Factor Weighted	0.9932	0.9471	0.9262
Unweighted vs. Impact Factor & Citation Weighted	0.9951	0.8993	0.8602
Papers-in-Patents Citations			
All Cites vs. Front Page Cites	0.9996	0.9961	0.9896
All Cites vs. Body Cites	0.9992	0.9920	0.9784

Notes: Table shows the correlation between different measures of the fundamental science research share produced at the country, city, and institution level for articles published in a set of top 15 life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal).*

Appendix Table A3: Correlation of Productivity and Papers-in-Patents Citation

	State-Level	City-Level	Institution-Level
Research Productivity vs. Papers-in-Patents Citations			
Impact Factor & Citation Weighted vs. All Cites	0.9950	0.9915	0.9589
Impact Factor & Citation Weighted vs. Front Page Cites	0.9939	0.9922	0.9669
Impact Factor & Citation Weighted vs. Body Cites	0.9935	0.9848	0.9294

Notes: Table shows the correlation between our main measure of fundamental science productivity and different versions of papers-in-patents citation counts at the state, city, and institution level for articles published in a set of top 15 life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal).*

Appendix Table A4: Top 50 US Institutions by Average Annual Output, 2015-2023

Rank	Institution	Annual Output
1	Stanford University	793.74
2	Harvard University	778.70
3	Massachusetts Institute of Technology	484.23
4	University of California, San Francisco	432.04
5	Mass General Brigham	346.02
6	National Institutes of Health	318.66
7	Johns Hopkins University	289.70
8	Memorial Sloan Kettering Cancer Center	283.18
9	University of California, San Diego	276.67
10	Columbia University	266.24
11	Yale University	258.23
12	Dana-Farber Cancer Institute	239.30
13	University of Pennsylvania	233.01
14	University of California, Berkeley	224.55
15	NYU Langone Health	214.54
16	Icahn School of Medicine at Mount Sinai	202.24
17	Washington University in St. Louis	200.49
18	The University of Texas Southwestern Medical Center	197.00
19	Cornell University	184.38
20	University of Washington	181.41
21	The University of Texas MD Anderson Cancer Center	165.19
22	University of California, Los Angeles	141.04
23	University of Michigan–Ann Arbor	124.30
24	Rockefeller University	122.77
25	Duke University	122.51
26	University of North Carolina at Chapel Hill	115.57
27	University of Chicago	111.71
28	Princeton University	105.87
29	California Institute of Technology	99.71
30	Baylor College of Medicine	96.35
31	University of Minnesota	93.16
32	Boston Children's Hospital	87.37
33	Salk Institute for Biological Studies	83.80
34	University of Pittsburgh	81.93
35	Northwestern University	73.90
36	University of Southern California	71.82
37	Scripps Research Institute	71.63
38	Mayo Clinic	66.62
39	St. Jude Children's Research Hospital	64.64
40	Broad Institute	63.57
41	University of Massachusetts Chan Medical School	63.56
42	Vanderbilt University	57.78
43	Janelia Research Campus	54.96
44	The University of Texas at Austin	53.47
45	Boston University	53.11
46	University of Wisconsin–Madison	48.49
47	Cold Spring Harbor Laboratory	48.17
48	Fred Hutch Cancer Center	46.96
49	University of California, Davis	45.01
50	Allen Institute	44.53

Notes: Table shows the top 50 institutions by average annual output in fundamental science research in life sciences journals. Research output is measured as the total number of publications between 2015 and 2023. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal.*

Appendix Table A5: Top 50 International Institutions by Average Annual Output, 2015-2023

Rank	Institution	Country	Annual Output
1	Chinese Academy of Sciences	China	413.57
2	Max Planck Society	Germany	259.73
3	University of Cambridge	United Kingdom	254.99
4	University of Oxford	United Kingdom	240.80
5	Helmholtz Association of German Research Centres	Germany	160.63
6	Karolinska Institutet	Sweden	153.93
7	Weizmann Institute of Science	Israel	144.34
8	Peking University	China	127.36
9	Wellcome Sanger Institute	United Kingdom	118.57
10	University of Queensland	Australia	113.20
11	University College London	United Kingdom	110.54
12	Heidelberg University	Germany	93.83
13	Imperial College London	United Kingdom	91.95
14	Leibniz Association	Germany	91.50
15	University of Copenhagen	Denmark	87.67
16	University of Toronto	Canada	85.25
17	Medical Research Council	United Kingdom	84.19
18	Tsinghua University	China	79.49
19	University of Hong Kong	Hong Kong	75.88
20	École Polytechnique Fédérale de Lausanne	Switzerland	75.74
21	ETH Zurich	Switzerland	73.97
22	Centre National de la Recherche Scientifique	France	72.27
23	University of Edinburgh	United Kingdom	71.85
24	The Netherlands Cancer Institute	Netherlands	66.85
25	The University of Tokyo	Japan	65.13
26	King's College London	United Kingdom	64.87
27	Royal Netherlands Academy of Arts and Sciences	Netherlands	61.85
28	Technical University of Denmark	Denmark	57.66
29	UNSW Sydney	Australia	54.65
30	Institut Pasteur	France	53.99
31	Kyoto University	Japan	52.57
32	Inserm	France	52.19
33	KU Leuven	Belgium	52.14
34	Chongqing Medical University	China	50.40
35	The Francis Crick Institute	United Kingdom	49.87
36	Seoul National University	South Korea	49.31
37	Keio University	Japan	48.86
38	Barcelona Institute for Science and Technology	Spain	48.67
39	Sun Yat-sen University	China	47.11
40	Charité - Universitätsmedizin Berlin	Germany	46.07
41	University of Zurich	Switzerland	45.84
42	RIKEN	Japan	45.23
43	McGill University	Canada	44.28
44	University of Helsinki	Finland	43.05
45	Zhejiang University	China	42.56
46	University of Freiburg	Germany	41.89
47	Lund University	Sweden	41.35
48	University of Geneva	Switzerland	40.19
49	Ghent University	Belgium	40.03
50	Ludwig-Maximilians-Universität München	Germany	39.35

Notes: Table shows the top 50 international institutions by average annual output in fundamental science research in life sciences journals. Research output is measured as the total number of publications between 2015 and 2023. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal.*

Appendix Table A6: Top 50 US Institutions by Average Researcher Productivity, 2015-2023

Rank	Institution	Researcher Productivity
1	Broad Institute	3.39
2	Dana-Farber Cancer Institute	2.73
3	Allen Institute	2.37
4	Massachusetts Institute of Technology	2.29
5	Memorial Sloan Kettering Cancer Center	2.22
6	Stanford University	1.87
7	Novartis (United States)	1.77
8	Gladstone Institutes	1.74
9	Salk Institute for Biological Studies	1.72
10	Icahn School of Medicine at Mount Sinai	1.69
11	Harvard University	1.59
12	Mass General Brighaml	1.53
13	The University of Texas MD Anderson Cancer Center	1.49
14	California Institute of Technology	1.43
15	Fred Hutch Cancer Center	1.42
16	Princeton University	1.42
17	Cold Spring Harbor Laboratory	1.42
18	Rockefeller University	1.34
19	The University of Texas Southwestern Medical Center	1.34
20	University of California, Berkeley	1.32
21	University of California, San Francisco	1.30
22	NYU Langone Health	1.29
23	Janelia Research Campus	1.27
24	Boston Children's Hospital	1.20
25	Jackson Laboratory	1.19
26	Columbia University	1.15
27	City Of Hope National Medical Center	1.14
28	University of Pennsylvania	1.08
29	St. Jude Children's Research Hospital	1.08
30	Washington University in St. Louis	1.06
31	University of Chicago	1.01
32	Yale University	1.01
33	University of California, San Diego	0.99
34	University of Massachusetts Chan Medical School	0.98
35	Cedars-Sinai Medical Center	0.96
36	Scripps Research Institute	0.91
37	MSD (United States)	0.85
38	Beth Israel Deaconess Medical Center	0.84
39	Cornell University	0.84
40	University of Washington	0.82
41	The University of Texas at Austin	0.79
42	Carnegie Mellon University	0.79
43	Johns Hopkins University	0.78
44	The University of Texas Medical Branch at Galveston	0.77
45	Baylor College of Medicine	0.74
46	University of Southern California	0.74
47	University of California, Santa Cruz	0.73
48	National Institutes of Health	0.69
49	Lawrence Berkeley National Laboratory	0.67
50	Albert Einstein College of Medicine	0.67

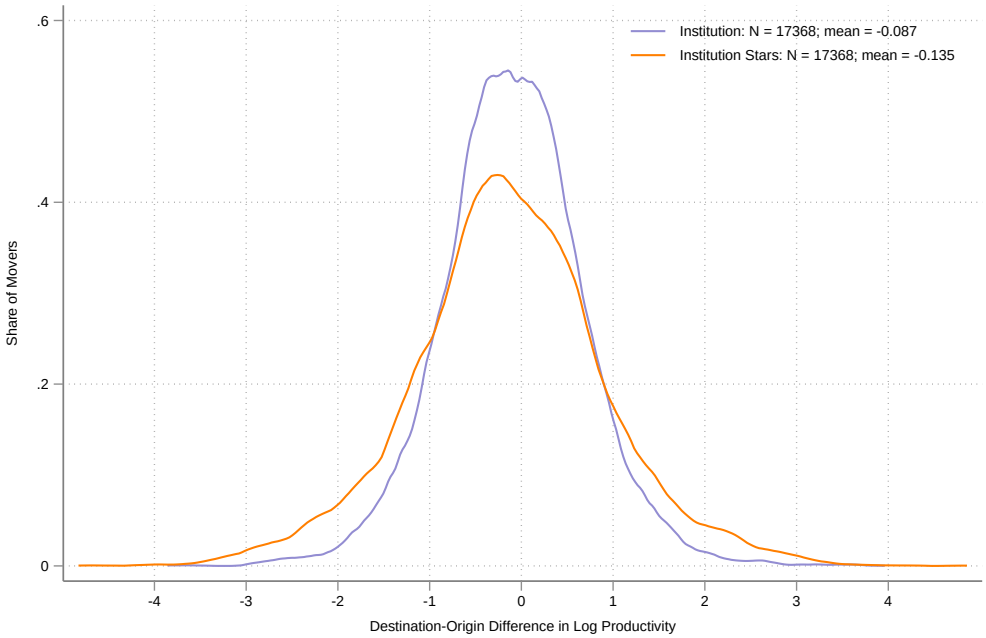
Notes: Table shows the top 50 institutions by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers at each institution. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal* between 2015 and 2023.

Appendix Table A7: Top 50 International Institutions by Average Researcher Productivity, 2015 - 2023

Rank	Institution	Country	Researcher Productivity
1	Wellcome Sanger Institute	United Kingdom	2.68
2	University of Trento	Italy	2.27
3	Royal Netherlands Academy of Arts and Sciences	Netherlands	2.23
4	The Netherlands Cancer Institute	Netherlands	1.94
5	Weizmann Institute of Science	Israel	1.63
6	Austrian Academy of Sciences	Austria	1.62
7	Barcelona Institute for Science and Technology	Spain	1.55
8	Chongqing Medical University	China	1.51
9	Cancer Research UK	United Kingdom	1.48
10	Princess Margaret Cancer Centre	Canada	1.48
11	Technical University of Denmark	Denmark	1.47
12	École Polytechnique Fédérale de Lausanne	Switzerland	1.35
13	Friedrich Miescher Institute	Switzerland	1.29
14	University of Hong Kong	Hong Kong	1.27
15	Institut Pasteur	France	1.13
16	University of Cambridge	United Kingdom	1.12
17	Biomedical Research Council	Singapore	1.12
18	University of Iceland	Iceland	1.10
19	KTH Royal Institute of Technology	Sweden	1.09
20	Medical Research Council	United Kingdom	1.07
21	University of Oxford	United Kingdom	1.05
22	The Francis Crick Institute	United Kingdom	1.01
23	Novartis (Switzerland)	Switzerland	0.99
24	Tsinghua University	China	0.97
25	Institute of Cancer Research	United Kingdom	0.97
26	University of Queensland	Australia	0.93
27	University Medical Center Freiburg	Germany	0.90
28	Leibniz Association	Germany	0.89
29	UNSW Sydney	Australia	0.89
30	Max Planck Society	Germany	0.87
31	Hospital for Sick Children	Canada	0.86
32	Helmholtz Association of German Research Centres	Germany	0.86
33	Peking University	China	0.82
34	Institut Curie	France	0.80
35	Stockholm University	Sweden	0.78
36	University Health Network	Canada	0.78
37	King's College London	United Kingdom	0.77
38	Karolinska Institutet	Sweden	0.77
39	Technion – Israel Institute of Technology	Israel	0.76
40	ETH Zurich	Switzerland	0.74
41	Chinese Center for Disease Control and Prevention	China	0.73
42	University of Edinburgh	United Kingdom	0.72
43	Shandong University	China	0.69
44	Walter and Eliza Hall Institute of Medical Research	Australia	0.68
45	Keio University	Japan	0.68
46	Garvan Institute of Medical Research	Australia	0.67
47	Agency for Science, Technology and Research	Singapore	0.67
48	Korea Advanced Institute of Science and Technology	South Korea	0.66
49	University of Dundee	United Kingdom	0.66
50	Roche (Switzerland)	Switzerland	0.66

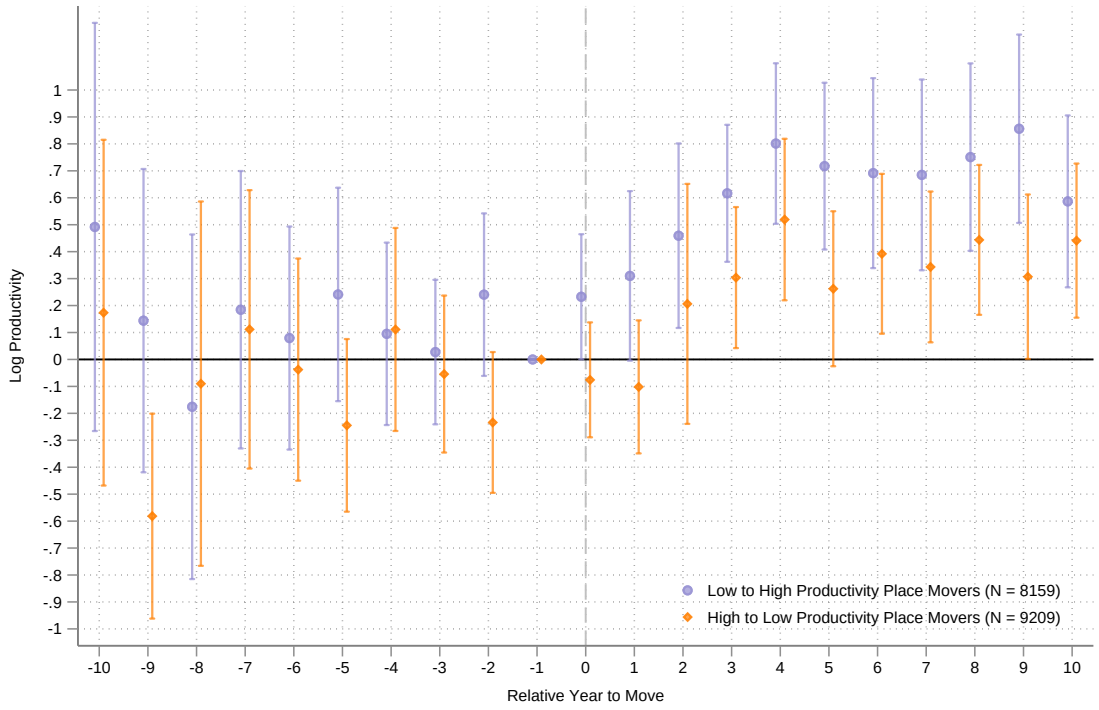
Notes: Table shows the top 50 international institutions by average researcher productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers at each institution. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal* between 2015 and 2023.

Appendix Figure A1: Distribution of δ_i



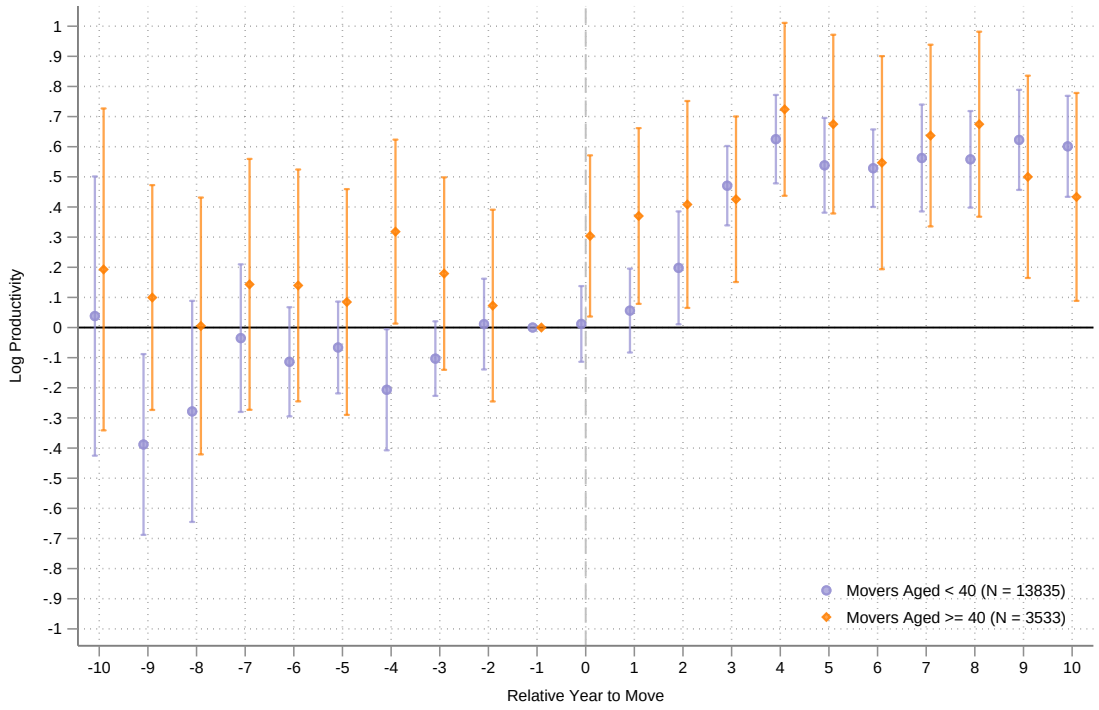
Notes: Figure shows the distribution of δ_i computed at various levels for scientist-movers. Here, δ_i measures the size of the move and is constructed as the average difference in productivity between the destination and origin location. We show the distributions for δ_i at institution-level in purple and for stars at the institution-level in orange.

Appendix Figure A2: Productivity Change based on Move Size



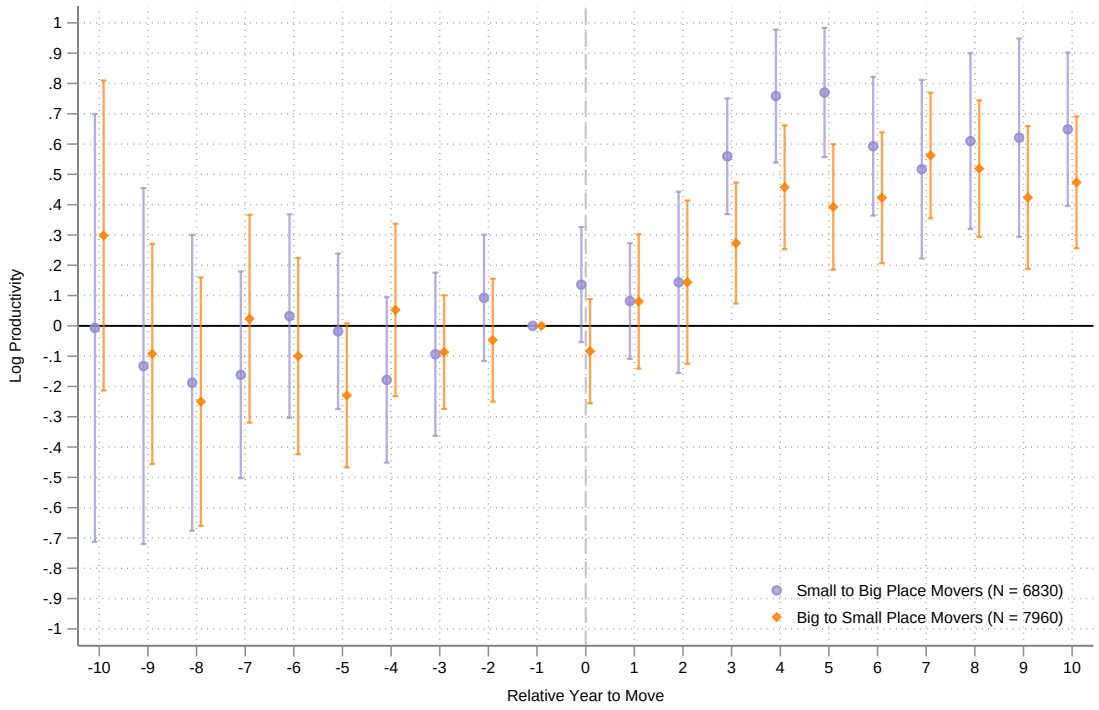
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3 for movers based on move size for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

Appendix Figure A3: Productivity Change based on Age



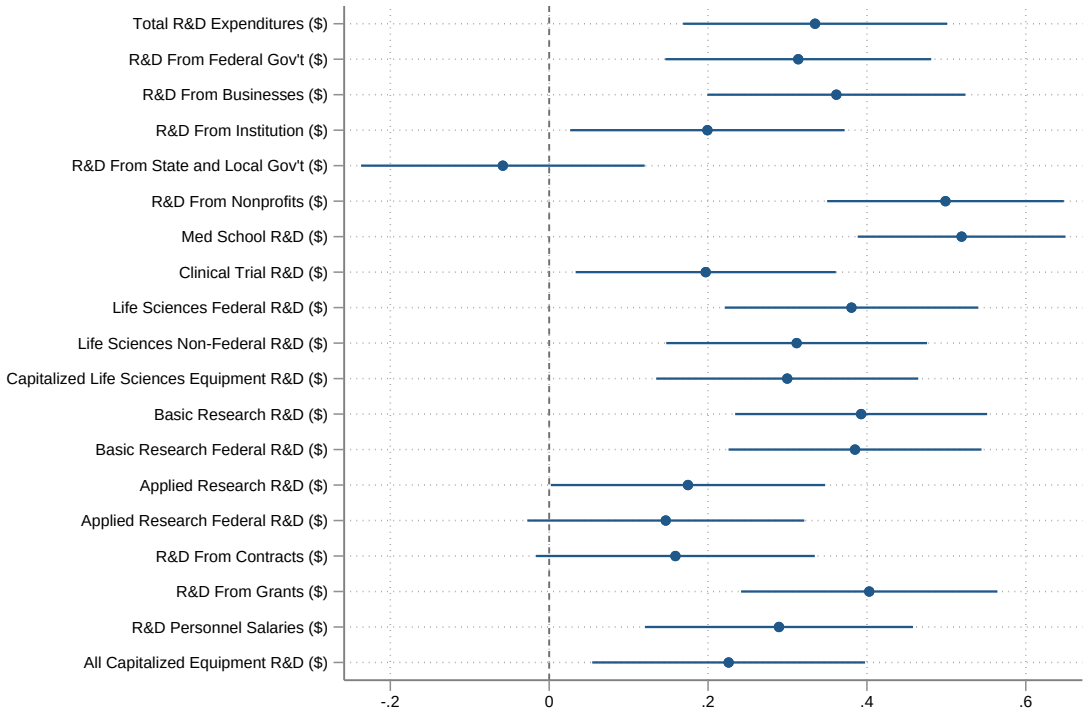
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3 for movers split by age for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

Appendix Figure A4: Productivity Change based on City Size



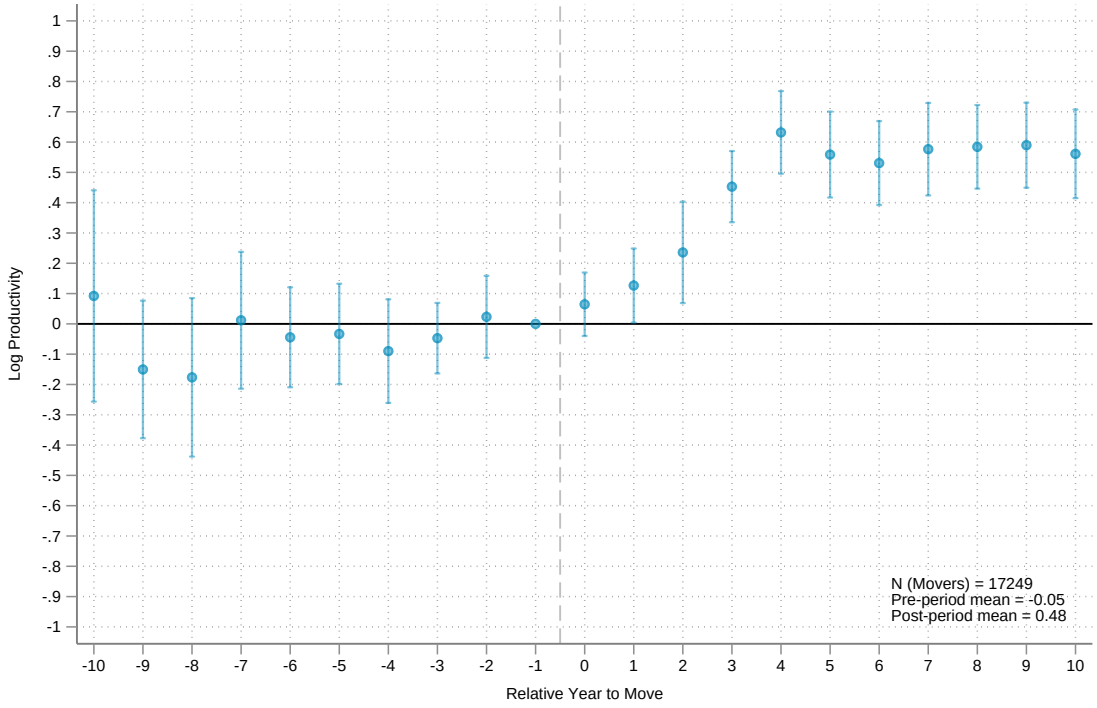
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3 for movers based on city size changes for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

Appendix Figure A5: Correlates of Institution Effects



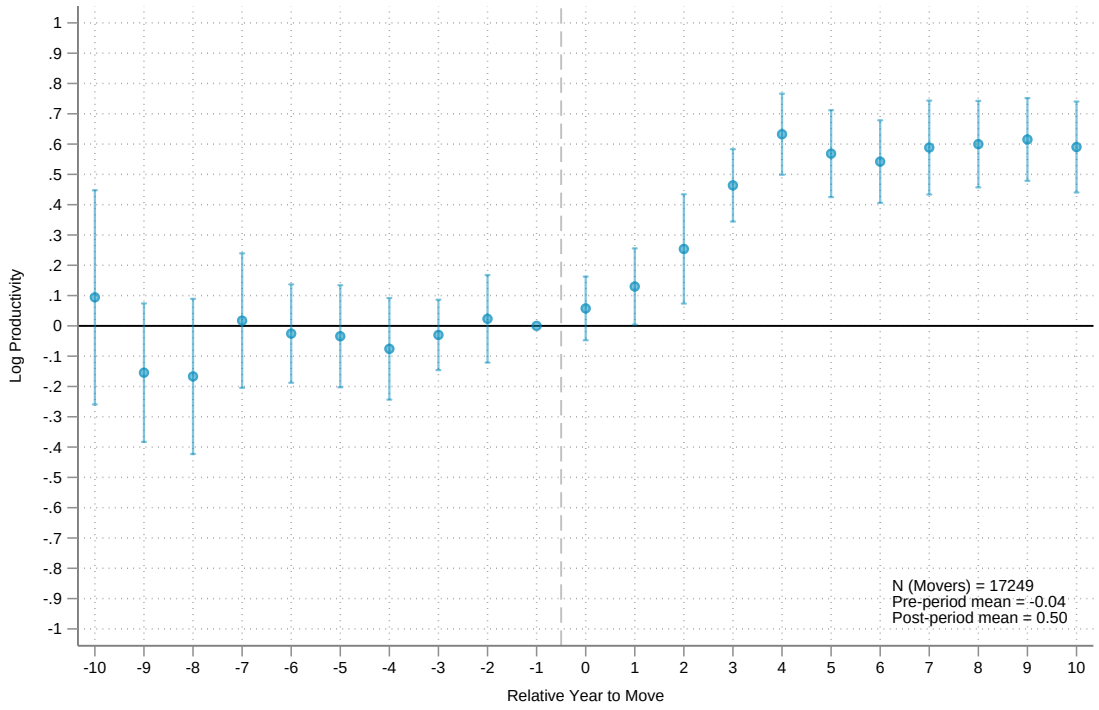
Notes: Figure shows results from regressing the estimated institution effects on various place-based factors from the NSF Higher Education Research and Development (HERD) Survey.

Appendix Figure A6: Specification including Field Fixed Effects



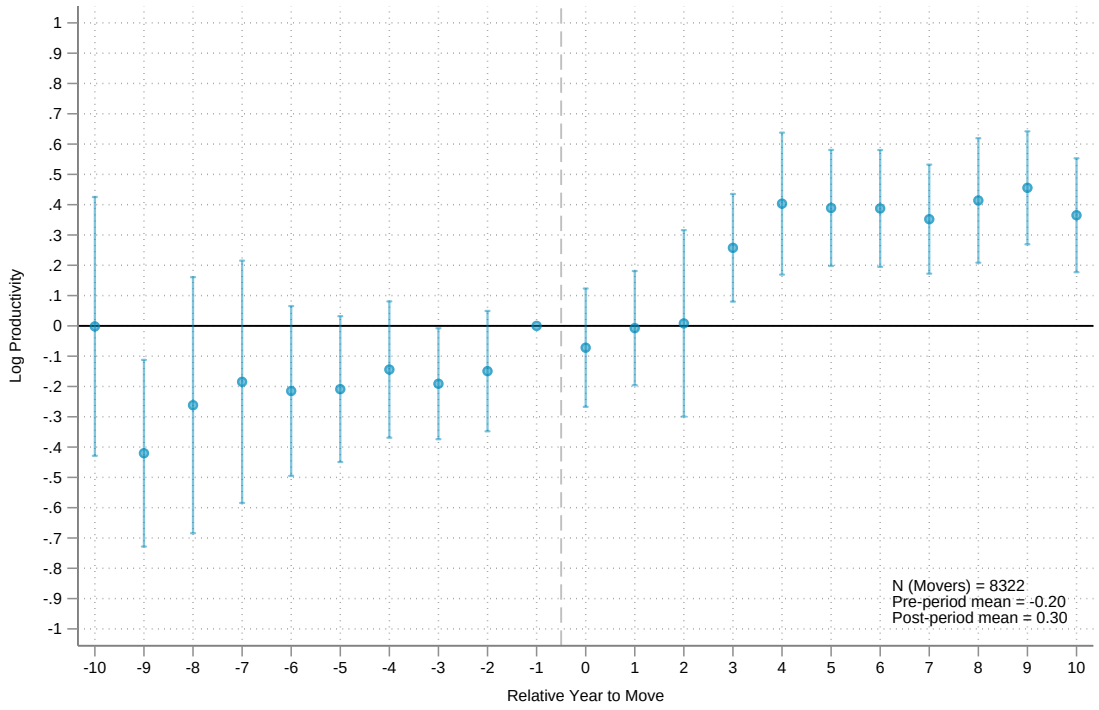
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3 with the inclusion of subfield fixed effects for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

Appendix Figure A7: Specification without Relative Year Fixed Effects



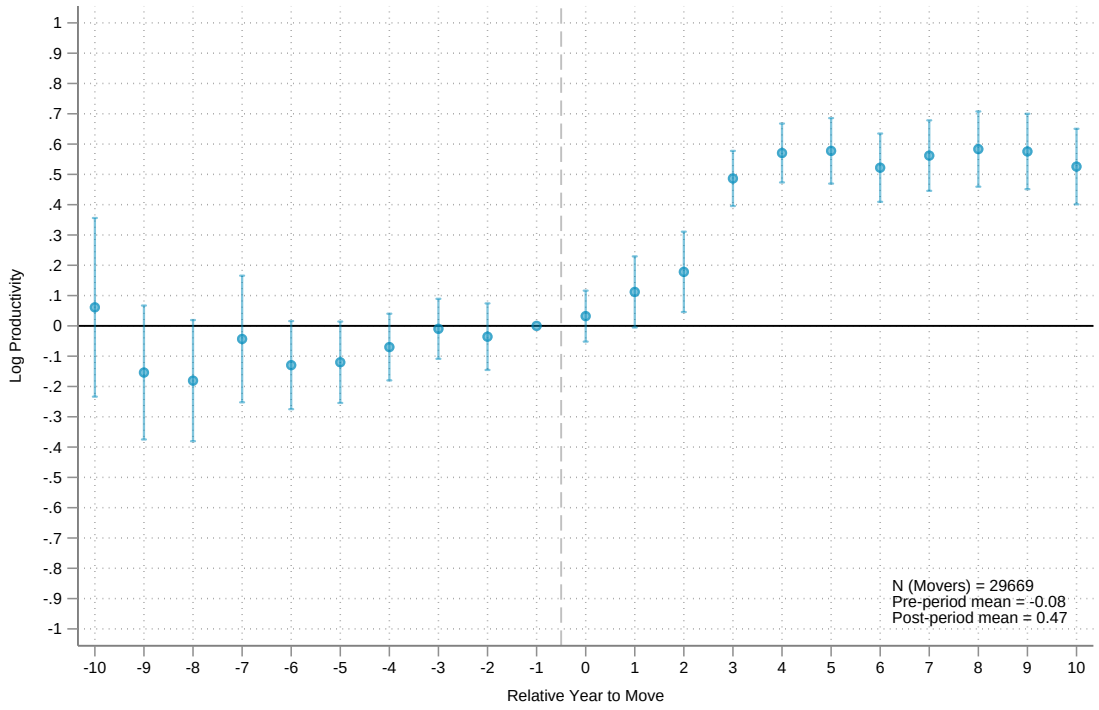
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3 without the inclusion of relative year fixed effects for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year and author fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

Appendix Figure A8: First and Last Authors



Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3 for the sample of first and last authors. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

Appendix Figure A9: All Authors



Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 3 for the sample of all authors publishing in the top 15 journals. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

B. Productivity Weighting

Each observation in our data represents an author publishing paper i in journal j. We implement a dual-level weighting approach that accounts for both the journal's overall prestige and the specific paper's influence when evaluating an author's research productivity.

Letting j denote journals and i denote papers, define

- N be the raw number of papers in our sample
- N_j be the raw number of papers in journal j
- IF_j be the latest 5-year impact factor of journal j
- C_{ij} be the average annual citation count for paper i in journal j
- A_i be the number of authors for paper i

We first adjust the distribution of journal representation in our sample according to their impact fact. This redistribution results in a sample that shifts weight toward journals with higher impact factors, regardless of their original number of papers. This helps adjust for the fact that higher impact journals may publish less often than lower impact journals. For instance, PLOS One and Journal of Biological Chemistry have raw journal sample sizes that each represent 31% to 37% of our full sample but have relative small impact factors of 3.8 and 4.8, respectively. In contrast, the two highest impact factor journals in our sample, Nature Medicine and Nature, have impact factors over 60 but are only represented 0.8% and 10% in our raw sample. Thus, for journal j we reweight it's paper count W_j as

$$W_j = \frac{IF_j}{\sum_k IF_k} \sum_k N_k$$

where the summations are over the k journals in our sample. Note, $\sum_k W_k = \sum_k N_k = N$.

To incorporate a paper's citation count, we first calculate a relative citation weight RCW_{ij} which is the paper's average annual citation count relative to the total average annual citation count in our sample.

$$RCW_{ij} = \frac{C_{ij}}{\sum_i \sum_j C_{ij}}$$

This measure captures the relative impact of each paper in the life-sciences as determined by annual citation performance.

In order to account for the fact that there may be journal-specific factors that influence citation rates, we create a journal-specific relative citation weight for each paper that quantifies each paper's citation impact within the context of its specific journal. A higher journal-specific relative citation weight indicates that a paper is cited more frequently relative to other papers in the same journal. We calculate this as

$$JRCW_{ij} = \frac{RCW_{ij}}{\sum_i RCW_{ij}}$$

where the denominator is the sum of RCW_{ij} across all papers i in journal j. Thus, for each journal the sum of $JRCW_{ij}$ across all papers in journal j is 1. By creating this intra-journal citation weight, we're effectively placing more weight on papers that may not have received high citation counts because it was published within a low-impact journal but was relatively influential within that low-impact journal.

Finally, we divide these paper-level weights by the number of authors A_i on a paper. Our final impact-

factor and citation weighted productivity measure is equal to

$$W_j \cdot JRCW_{ij} \cdot \frac{1}{A_i}$$

C. Constructing Life-Sciences Subfields

We implement K-means++, which improves on the original k-means algorithm by choosing a smarter initial centroid than k-means. This tends to produce more well-separated and balanced clusters than the original k-means algorithm. For each author-year we combine all the text (mesh terms, title, abstract) from their papers.

Year Bin: 1945–1964

1. Neuroscience: brain, physiology, metabolism, pharmacology, liver, inhibition, tissue, animal, acid, drug, chemistry, blood, pattern, action, technique
2. Muscle Physiology & Biomechanics: physiology, muscle, metabolism, function, pharmacology, liver, cell, blood, pig, mechanism, action, inhibition, influence, temperature, vitro
3. Comparative Physiology: liver, tissue, animal, compound, metabolism, muscle, solution, concentration, action, substance, made, strain, urine, specie, water
4. Biochemistry: acid, metabolism, amino, synthesis, liver, formation, compound, pharmacology, tissue, cent, extract, isolated, solution, protein, derivative
5. Microbiology: metabolism, synthesis, formation, liver, bacteria, tissue, nucleotide, plant, glucose, mechanism, compound, oxidation, vitro, pharmacology, vitamin
6. Cell Biology: cell, metabolism, culture, tissue, acid, pharmacology, growth, liver, synthesis, medium, blood, vitro, animal, structure, demonstrated
7. Hematology: blood, serum, plasma, pharmacology, cell, metabolism, chemistry, acid, concentration, animal, rabbit, mechanism, liver, action, three
8. Molecular Biology: protein, metabolism, synthesis, chemistry, acid, liver, serum, cell, plasma, fraction, pharmacology, amino, solution, component, formation
9. Endocrinology: hormone, pharmacology, growth, metabolism, action, physiology, injection, tissue, animal, urine, administration, chemistry, blood, influence, substance
10. Structural Biology: chemistry, acid, structure, metabolism, chromatography, pharmacology, tissue, synthesis, component, content, nucleotide, plant, property, serum, isolation
11. Pharmacology: pharmacology, action, drug, metabolism, acid, inhibition, agent, growth, influence, derivative, compound, induced, vitamin, experimental, liver
12. Virology: virus, strain, cell, culture, acid, plant, isolated, pharmacology, tissue, agent, induced, chemistry, animal, protein, metabolism
13. Radiation Biology: radiation, pharmacology, cell, induced, light, biological, produced, solution, acid, metabolism, action, tissue, body, given, plant
14. Enzymology: enzyme, liver, metabolism, synthesis, acid, chemistry, phosphate, formation, tissue, pharmacology, inhibition, action, property, protein, extract
15. Physiology: factor, blood, serum, pharmacology, physiology, metabolism, liver, chemistry, vitamin, growth, plasma, certain, animal, acid, protein

Year Bin: 1965–1984

1. Biophysics & Membrane Biology: concentration, blood, rate, membrane, binding, liver, brain, complex, chain, muscle, structure, cytochrome, factor, site, high
2. Metabolism & Biochemistry: metabolism, biosynthesis, pharmacology, enzymology, brain, liver, calcium, cell, blood, synthesis, muscle, protein, glucose, erythrocyte, sodium
3. Pharmacology & Drug Action: receptor, binding, metabolism, cell, membrane, affinity, brain, site, physiology, hormone, drug, pharmacology, protein, complex, specific
4. Molecular & Microbial Genetics: dna, sequence, genetics, polymerase, cell, metabolism, synthesis, fragment, gene, protein, coli, biosynthesis, site, genome, virus
5. Virology & Viral Genetics: virus, cell, immunology, genetics, antibody, antigen, disease, dna, rna, strain, particle, protein, culture, line, isolated
6. Enzymology & Protein Function: enzyme, substrate, purified, molecular, weight, acid, liver, subunit, gel, complex, protein, concentration, site, dehydrogenase, rate
7. Neurophysiology & Electrophysiology: physiology, muscle, metabolism, cell, brain, pharmacology, potential, calcium, membrane, stimulation, cytology, mechanism, sodium, blood, function
8. Genetics: chromosome, genetics, gene, cell, dna, mouse, region, physiology, sequence, cytology, long, metabolism, lymphocyte, genetic, line
9. Structural Biology: acid, amino, residue, peptide, sequence, chain, protein, structure, enzyme, isolated, synthesis, cell, chromatography, product, molecular
10. Genomics & Gene Expression: gene, genetics, sequence, dna, region, expression, transcription, nucleotide, mrna, rna, cell, cloned, mutation, protein, mouse
11. Immunology: immunology, antigen, antibody, lymphocyte, cell, serum, mouse, surface, specific, vitro, genetics, virus, pharmacology, blood, disease
12. Cell Signaling: protein, kinase, membrane, cell, binding, molecular, subunit, gel, synthesis, weight, purified, phosphorylation, acid, peptide, fraction
13. Neuropharmacology: drug, pharmacology, metabolism, action, induced, cell, inhibition, brain, physiology, behavior, receptor, blood, vitro, calcium, muscle
14. Cell Biology: cell, culture, line, cytology, growth, membrane, mouse, synthesis, medium, protein, cultured, surface, vitro, tumor, fibroblast
15. Molecular Biology: rna, mrna, sequence, polymerase, synthesis, metabolism, dna, biosynthesis, cell, transcription, nucleotide, gene, genetics, protein, vitro

Year Bin: 1985–2004

1. Structural Biology & Biophysics: binding, domain, peptide, structure, residue, protein, site, interaction, acid, complex, sequence, region, mutant, affinity, cell
2. Cancer Biology & Oncology: cell, expression, protein, apoptosis, growth, activation, receptor, line, tumor, factor, gene, antibody, kinase, death, cancer
3. Cell Signaling & Signal Transduction: kinase, phosphorylation, cell, protein, activation, tyrosine, receptor, signaling, phosphorylated, pathway, domain, insulin, factor, growth, site

4. Molecular & Regulatory Genetics: promoter, transcription, gene, cell, expression, element, factor, site, protein, binding, region, transcriptional, sequence, dna, nuclear
5. Membrane & Cellular Dynamics: protein, cell, sequence, acid, domain, cdna, membrane, amino, antibody, full, binding, gene, complex, region, mutant
6. Cellular Metabolism & Bioenergetics: cell, protein, metabolism, membrane, acid, immunology, factor, complex, expression, chemistry, drug, muscle, transport, rna, disease
7. Receptor Biology & Pharmacology: receptor, cell, binding, protein, ligand, metabolism, domain, affinity, activation, agonist, expression, signaling, mutant, physiology, factor
8. Neuroscience & Neurobiology: physiology, neuron, metabolism, cell, cytology, neuronal, brain, genetics, protein, receptor, neural, synaptic, apoptosis, pharmacology, transcription_factors
9. Genetics & Genomic Disorders: genetics, gene, mutation, chromosome, syndrome, metabolism, cancer, disease, carcinoma, protein, transcription_factors, locus, mouse, dna_binding_proteins, expression
10. Ion Channel Biology & Electrophysiology: channel, ca2, current, cell, calcium, subunit, membrane, protein, receptor, potassium, ion, intracellular, oocyte, activation, neuron
11. Immunology & Molecular Immunogenetics: alpha, beta, subunit, cell, receptor, protein, binding, chain, gene, sequence, acid, mrna, expression, complex, peptide
12. Genomics & Functional Genomics: gene, sequence, chromosome, genome, expression, cell, genetics, protein, region, exon, mrna, dna, rna, mutation, mouse
13. Tumor Suppressor Biology & Cancer Genetics: p53, cell, apoptosis, tumor, gene, protein, dna, expression, suppressor, cancer, genetics, mutant, arrest, mutation, metabolism
14. DNA Damage & Genome Stability: dna, replication, repair, protein, polymerase, binding, complex, cell, strand, site, sequence, genetics, gene, base, structure
15. Enzymology & Catalytic Mechanisms: enzyme, substrate, acid, residue, protein, site, mutant, active, catalytic, purified, structure, amino, sequence, subunit, reductase

Year Bin: 2005–2024

1. Structural Biology: domain, binding, structure, protein, residue, site, complex, interaction, peptide, enzyme, structural, substrate, receptor, crystal, chemistry
2. Genomics: genome, genetic, variant, gene, locus, population, snp, genetics, genomewide, polymorphism, sequence, disease, variation, chromosome, trait
3. Ecology & Environmental Biology: specie, plant, different, area, water, community, physiology, population, high, condition, rate, pattern, performance, concentration, temperature
4. Genetics & Genomic Medicine: mutation, gene, genetics, sequencing, mutant, cancer, cell, syndrome, disease, protein, identified, variant, genetic, tumor, family
5. Molecular Biology: gene, expression, cell, transcription, promoter, rna, genetics, protein, transcriptional, factor, pathway, identified, plant, regulation, expressed
6. Virology: virus, infection, vaccine, antibody, viral, cell, immune, infected, strain, disease, respiratory, immunology, host, protein, antigen
7. Cancer Biology: cancer, tumor, cell, breast, expression, metastasis, lung, growth, carcinoma, gene, therapy, survival, progression, protein, tissue

8. Cell Biology: protein, cell, membrane, complex, interaction, domain, function, expression, mutant, metabolism, kinase, binding, phosphorylation, full, acid
9. Epidemiology: woman, child, health, mortality, disease, outcome, hiv, factor, prevalence, care, cohort, rate, higher, participant, score
10. Neuroscience: neuron, brain, neuronal, cortex, cell, neural, physiology, receptor, memory, network, behavior, mouse, expression, disease, function
11. Metabolism & Endocrinology: liver, expression, cell, fatty, disease, injury, diet, gene, lipid, metabolism, acid, tissue, insulin, serum, metabolic
12. Stem Cell & Regenerative Biology: cell, expression, stem, protein, proliferation, apoptosis, tumor, line, receptor, activation, cancer, signaling, gene, growth, pathway
13. Immunology: cell, receptor, expression, metabolism, signaling, activation, pathway, drug, macrophage, protein, inflammation, disease, function, mechanism, inhibitor
14. Physiology & Systems Biology: muscle, cell, expression, protein, gene, mitochondrial, function, metabolism, tissue, insulin, signaling, activation, mass, physiology, mouse
15. DNA & Genome Stability: dna, repair, cell, damage, replication, protein, gene, binding, complex, genome, sequence, site, chromatin, polymerase, genetics