NBER WORKING PAPER SERIES

WHERE DISCOVERY HAPPENS:
RESEARCH INSTITUTIONS AND FUNDAMENTAL KNOWLEDGE IN THE LIFE-SCIENCES

Amitabh Chandra
Connie Xu

Working Paper 33996
http://www.nber.org/papers/w33996

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2025

We thank, without implicating, Alan Garber, George Daley, Kyle Myers, Bitsy Perlman, Bhaven Sampat and Amy Wagers for helpful discussions and comments. This research was funded by the Harvard Business School. We thank seminar participants at Harvard Medical School in April 2024, the NBER Entrepreneurship meeting in December 2024, and USC in February 2025 for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33996

Where Discovery Happens: Research Institutions and Fundamental Knowledge in the Life-Sciences

Amitabh Chandra and Connie Xu
NBER Working Paper No. 33996
July 2025
JEL No. H5, I2, O3

ABSTRACT

Fundamental knowledge in the life sciences has consequential implications for medicine and subsequent medical innovations. Using publications in leading life science journals to measure fundamental knowledge, we document large agglomerations in the institutions where it is discovered and a robust correlation between knowledge and subsequent citations in patents. We assess whether the institution where research is produced affects the output of scientists by using a scientist-mover design, which compares annual research output before and after a move for the same scientist. Between 50 60% of a scientist's research output is attributable to the institution where they work, and two thirds of this effect is driven by the presence of star researchers. The magnitude of these effects has not decreased in more recent time periods, in the wake of technologies that make cross-institution collaborations easier, nor is it larger for moves to larger agglomerations, nor concentrated in particular scientific fields. We discuss the implications of these findings for research allocations in science and scientists' leaving one institution for another.

Amitabh Chandra
Harvard University
Department of Economics
and NBER
amitabh_chandra@harvard.edu

Connie Xu
Harvard University
conniexu@g.harvard.edu

Fundamental breakthroughs in life sciences, such as the discovery of DNA, genes, mRNA, antibodies, and CRISPR, have enabled new medicines, improved clinical practice, and deepened our understanding of biology and physiology, resulting in even more knowledge being discovered (Zerhouni, 2005). These discoveries originate primarily from universities and research hospitals, raising the question of whether some institutions are more productive than others in discovering fundamental scientific knowledge. This answer is especially salient because the market is likely to under-provide funding for fundamental knowledge, despite its foundational nature, and because funding for such science may be inadequate or capricious. Understanding the forces that produce these discoveries has implications for welfare, understanding how resources should be allocated to institutions versus individual scientists, and how scientists should evaluate moving to other institutions.

The institution where research is performed can increase scientific output by providing access to inputs such as talented students, better equipment, experienced mentors, and richer networks for collaboration (Furman and Stern, 2011). Moreover, if proximity to other researchers facilitates knowledge spillovers and collaborations, there may be benefits from larger research ecosystems, which also implies that the discovery of knowledge will be concentrated (Zucker et al., 1998). Assessing the causal effect of an institution on research output is challenging because more productive scientists might self-select into locations based on personal or professional preferences— such as where they trained, dual-career opportunities, or amenities— making some places appear more productive because of who chooses to work there. Such a sorting means that agglomerations in discovery could reflect differences in institutional output or an agglomeration of productive scientists. Our paper overcomes these identification challenges by disentangling institution effects from research selection and contributes to the extensive literature documenting the central role of institutions in fundamental scientific discoveries (Nelson, 1959; Allison and Long, 1990; Zucker et al., 1998; Belenzon and Schankerman, 2013; Azoulay et al., 2011; Hvide and Jones, 2018).

We present new facts on the concentration of where fundamental discoveries in the life-sciences occur. We offer a way to define such discoveries using data on publications that is transparent and can be replicated by others, and rely on publication data from *OpenAlex*, a comprehensive open catalog of global research (Priem et al., 2022). For this descriptive exercise, our sample consists of over 560,000 life-science papers (published 1945-2023) across 15 journals, collectively cited over 65 million times.[1] We measure research output as the frequency with which a scientist is listed as the first 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).

To determine whether the association between where scientists work and their output is causal, we implement a movers design, where we estimate an event study specification of an individual's research output around job moves, exploiting variation in the "move size", defined as the difference in the average (per-scientist) research output of the origin and destination institutions.[2] If institutional factors explain 100% of a scientist's output, then moving to an

---

[1] These data highlight 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 would not be true for industries that rely on trade secrets over patents or where fundamental knowledge is not published in journals.

[2] We refer to papers per scientist as "output" and not "productivity" because we are not measuring other inputs that affect productivity, like facilities, resources, administrative support, colleagues and students.

institution with (say) a 20% higher average output should increase a scientist's output by 20%. Conversely, if research output is entirely an individual-level phenomenon, then relocation to a more or less productive institution will have no effect on a scientist's output. Since the movers design links changes in an individual scientist's output to changes in the output of their location, we can control for time-invariant individual characteristics and overall time trends and test whether more productive scientists tend to move to more productive institutions, which would confound the identification of true place effects. For this work, we track the location of over 180,000 authors and their research output, including the change in output of those moving (over 19,000 authors), 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). As we discuss below, the average move in our sample is to a destination institution that is about 9 percent less productive than the institution of origin, with large variation in both directions, around this average. Empirically, we find that the event-study approach gives results that are very similar to those from a simpler two-way fixed-effects model, which also uses movers to estimate institution and individual effects.

Our approach for the movers design is closely related to Lerner et al. (2024), who pioneered using a 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 eventually influences patents, clinical practice, or therapeutics, these impacts often emerge over decades through complex non-linear pathways (Myers and Lanahan, 2022). If this is true, Lerner et al.'s estimates, which primarily capture direct and immediate connections, may be conservative because they do not measure the full influence of university environments on scientific output writ large. Our main results are also similar to those of (Allison and Long, 1990) who use a different empirical approach to study a smaller sample of movers from a broader range of fields.

For our analysis, we need to define fundamental science research, measure it in a systematic way, and attribute it to the institution where it was done. We adopt the fundamental research definition of the National Academy of Medicine.[3] This definition includes 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 discoveries 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

---

[3] 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).

(Nelson, 1959; Akcigit et al., 2021; Azoulay and Li, 2022; Myers and Lanahan, 2022).

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. In particular, the gap in output between the US and its closest competitor in recent times, China, is more than five times. Three countries (the US, China and the UK) account for 70 percent of the world's research output. Within the US, Boston and the Bay Area produce a disproportionate share– over 15 percent– of the world's fundamental life-science research. A handful of institutions, mostly in the US, account for an enormous share of global discovery, with Harvard and Stanford accounting for over 8 percent of the world's output, and underscoring the extent to which a handful of institutions, mostly in the US, produce more fundamental science than the total of many countries. 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), which reinforces the role of fundamental research in driving subsequent innovation in the life sciences. A corollary of this facts is that disrupting this allocation will slow knowledge throughout the world (one scientist's output is another scientist's input) with damaging effects on commercial prospects.

It is important to recognize that geographic agglomerations reflect two forces: the number of researchers multiplied by their average output, and not just the latter. Although cities and institutions with many researchers house some of the most productive researchers, our analysis of geographic variation in output-per-author shows that highly productive scientists are distributed across a diverse set of institutions, including universities, government labs, and biotechnology firms. This motivates understanding whether institutions with high average output causally increase the output of scientists that move to them, or whether a move to an institution with relatively lower output per scientist reduces the mover's subsequent output.

The causal estimates from the movers design show that location is highly consequential for scientific output. Approximately $50-60\%$ of the geographic variation in researcher output 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, talent, and connections. The effect is asymmetric: researchers who relocate to more productive environments experience a slightly larger increase in their output than the decline in output when a researcher locates to a less productive institution.

We discuss the implications of these results for science policy. Funders, like the National Institutes of Health (NIH) and philanthropists, face the choice between spreading funds broadly across many institutions, versus concentrating resources in a few institutions where the average scientist is highly productive. Our results suggest that treating all institutions as equal recipients would be inefficient if the goal is to maximize knowledge production (it may, of course, meet other distributional objectives). Even across the limited set of the 50 institutions with the most productive scientists, there are fourfold differences in research output, and our estimates imply that over 1/2 of these differences are causal. Said differently, if research funds are an input to discovery, as noted by a large and robust empirical literature (Hall and Lerner, 2010; Howell, 2017; Babina et al., 2023; Tham et al., 2024), and funders are choosing between two equally productive scientists, one at an institution whose average research output is twice the other's, then funders could get more than 50% more research by prioritizing a scientist at the more productive institution. At the same time, we find that research output is driven by an institution's output and not the size of its local research cluster, suggesting that agglomeration effects are weaker for scientific discovery, even though they may be important for inventiveness (e.g. Moretti (2021); Azoulay et al. (2012)). This would happen if fundamental scientific discovery depended on highly localized strengths, such as research facilities, organization, institutional leadership, collaborative culture, and mentorship, more than knowledge spillovers from one institution to another. Consequently, highly productive institutions in smaller locations that are not co-located with other large institutions (e.g., Dartmouth in Hanover, NH, or the University of Michigan in Ann Arbor, MI) are just as efficient at producing fundamental science as equally productive institutions in larger research hubs like New York City. These findings align closely with previous research from Gruber et al. (2022), who document diminishing returns to adding more researchers to dense science clusters, and Guzman et al. (2023), who highlight that place-based policies succeed most when they leverage existing institutional strengths and talent pools.

These results raise the question of how an institution may improve its productivity. We find that more than two-thirds of the estimated institutional effect is driven by the presence of star researchers, defined as scientists at the top 5% of their institution's output distribution (we also provide some suggestive evidence that the output of a star researcher's coauthors at the institution of origin falls after they move). This prominence of "stars" aligns with previous research that top scientists disproportionately drive knowledge creation, which identifies a unique opportunity to build research excellence. For example, Waldinger (2016) 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, success in attracting research grants, and large instrumentation; we emphasize that these correlations are not causal but help with validating the institutional effects that we estimate. Understanding the role of financial and non-financial inputs remains an important area for future research.

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.[4]

---

[4] The NAM definition is focused on the life-sciences and *excludes* 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.

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 systematically measuring fundamental science research.

Our dataset includes journal articles published between 1945[5] 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 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*. Authors who have published in these journals are also the ones whose research output we study in the movers design, so having a broader list of journals is more desirable because it generates more movers than a more selected list.[6] This selection process yields a core sample of 563,620 publications, authored by 1,433,962 scientists. The proportion of fundamental science articles in these journals, measured using our definition, ranges from 37% (Science) to 98% (Oncogene), with remaining articles related to the 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 catalog of the global research system developed by the nonprofit *OurResearch* (Priem et al., 2022).[7] *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*.[8] 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* to identify the geography of research. It employs intentional efforts to disambiguate authors and standardize institutional affiliations. Using the author information from *MAG*, *Crossref*,

---

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

[6] We determined this list by asking 5 leading scientists at Harvard Medical School, including the Dean, to list the 10 leading fundamental science in the life-sciences. 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 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.

[7] We are indebted to Bhaven Sampat for suggesting this resource; in earlier drafts we used PubMed. All errors in implementation are our own.

[8] 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.

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 output. 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.[9]

B. Measuring Scientist's Research Output

We move from publications to the individual scientists' research output 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 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 the 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. If anything, 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 all authors, we find that location-specific influences increase by approximately 10%. This suggests that beginning and end authors tend

---

[9] 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 output age profile using our definition matches that found in previous research (Jones, 2009)

to be more established researchers and are relatively less affected by local institutional factors, whereas middle authors may be more affected by place-based influences.

C. Measuring Institutional Output

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.[10] 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 of the institutions to the author, affecting around 5% of author-years.

Having assigned each author to a unique institution for each year, we aggregate output to broader geographic levels (e.g., cities and countries). 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 Research Triangle Park, NC, as a 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 output 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 (examples include cell signaling and stem cells). This approach helps us control for scientists selecting into fields that are advancing at different rates and that may affect the probability of acceptance at a scientific journal (e.g. genetics versus structural or cancer biology following the human genome mapping). To define these fields, 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 the definitions of the field to evolve over time. Appendix Section C details the tokenized features that define 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*",

---

[10] We omit affiliations that identify a funder (e.g. the Wellcome Trust, Howard Hughes Medical Institute (HHMI), or the Broad Institute). We count HHMI only when the entry specifies the Janelia Research Campus, and the Broad Institute only when it is the scientist's primary employer rather than an affiliated faculty member. Hospital affiliations whose names include a university (e.g. NYU Langone, Stanford Hospital) are recoded to the university, whereas academically affiliated but separately branded hospitals (e.g. Mass General Brigham) remain distinct.

was published in *Science* in 2012 and has been cited 13,832 times. To construct our measure of research output, 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.[11] 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 output 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 re-calibrated 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.[12] 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 output into context, Table 3 shows that the average author in an elite set of scientific papers has an annual output of 0.39 articles and a lifetime output of 1.48 articles (measured until the end of our time frame in 2023).

II. Fundamental Science Discovery

A. Geographic Variation in Research Output

Figure 1 shows the share of research output produced by each country since 1945. 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.[13] 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

---

[11] 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

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

[13] 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

remarkable strides in producing fundamental science research– surpassing the UK as the second largest producer in recent years. Also notable is the shrinking role of the United Kingdom 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%.

Within the US, aggregate research production is highly concentrated in a few key metropolitan areas and institutions (to be clear, aggregate production reflects the number of scientists and their individual output). 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 the top 50 institutions by average annual output for US-based institutions and international institutions, respectively. Figures 2 and 3 illustrate the persistence of these places in producing a large share of the 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 from aggregate production to individual researcher output, we observe a more distributed pattern. Table 1 presents the top 20 cities by average researcher output in 2015-2023. Our measure of research output 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.[14] Table 2 further highlights the top institutions in the ten most productive cities. These results show that the average researcher at Dana-Farber Cancer Institute effectively produced four times as many publications per year than the average researcher at Lawrence Berkeley National Laboratory. While elite institutions like Harvard and Stanford dominate worldwide scientific production because of their size, we observe that many smaller institutions house highly productive researchers, including universities, research organizations, government agencies, and biotechnology companies.

Appendix Tables A6 and A7 provide the list of the top 50 institutions by average researcher output for US and non-US institutions, respectively. These tables reinforce the distinction between aggregate research output that is dominated by a small number of institutions and researcher-level output, which is distributed across a diverse set of institutions. A number of international institutions such as the Wellcome Sanger Institute, several Dutch hospitals and universities, and the Weizmann Institute of Science in Israel have per-capita productivity that is comparable to the leading US institutions. These institutions are relatively smaller than US institutions, which is why they do not account for a large share of world output, but their scientists are on average just as productive as those at Dana-Farber, the Allen Institute, MIT, Sloan Kettering and Stanford.

B. Fundamental Science and Commercial Relevance

So far, we have not established whether fundamental science research has direct implications for commercial activity, versus being knowledge that is fundamental but that is not correlated with inventions. Building on work by (Bryan et al., 2020), (Marx and Fuegi, 2020), we measure the propensity of fundamental science research to be cited in patents. This measure of "papers

---

[14] When calculating average output 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 institutions that may be miscoded or lack sufficient representation.

in patents" output is similar to our measure of research output, 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 output and "papers in patents" output 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. At the same time, our inquiry does not rely on this fact for fundamental science research is sufficiently foundational that it may have no commercial relevance.[15]

III. Research output of Institutions and Individuals

A. Empirical Strategy

We present an empirical framework using a "movers" design to decompose the observed variation in fundamental research log output 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 output, y_{ijt}, at institution j in year t with a two-way fixed effects specification:

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

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 certain regression specifications, we incorporate subfield fixed effects which would fall under $x_{it}\beta$.

A key identification challenge is that the fixed effects for the institution, γ_j, are only identified if the sample includes the scientists who move between institutions. Without movers, we cannot disentangle place effects from individual heterogeneity. A causal interpretation of these place effects requires several assumptions. First, unobserved changes in a scientist's log output poten-

---

[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.

tial must not be correlated with the log output difference between their origin and destination institutions. This assumption would be violated if "rising stars" systematically relocate from low- to high-output institutions. Violations of this kind will appear in the pre-trends of our event study so examining these is critical. Second, we assume additive separability of α_i and γ_j in log output, implying that person and place characteristics are multiplicative in levels. In other words, a move to a higher-output institution has a larger absolute impact on lower-output 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.

If these assumptions are satisfied, then it is possible decompose the variation in average log researcher output 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 the averages at the researcher level. The difference in average log output between two institutions j and j' is given by:

$$\overline{y}_j - \overline{y}_{j'} = (\gamma_j - \gamma_{j'}) + (\overline{\alpha}_j - \overline{\alpha}_{j'})$$

This permits us to report shares attributable to the institution and researchers as:

$$S_{institution}(j, j') = \frac{\gamma_j - \gamma_{j'}}{\overline{y}_j - \overline{y}_{j'}}$$

$$S_{researcher}(j, j') = \frac{\overline{\alpha}_j - \overline{\alpha}_{j'}}{\overline{y}_j - \overline{y}_{j'}}.$$

We report these results and shares in the two panels of Table 4.

B. Event Study Estimation

We implement an event study approach to visually observe the dynamic evolution of place effects on scientist's log output and assess the robustness of our assumptions through the pre-trends. We first scale each mover's log output to account for the differing output 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 output 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 with 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.[16]

Following (Bronnenberg et al., 2012), we scale log output 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 output equals their origin institution's average log output and $y_{it}^{scaled} = 1$ when it equals their destination institution's average log output. 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 analog 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 output trend unrelated to move size, (captured by $\rho_{r(i,t)}$) across all movers, regardless of the size and direction of the move. These effects capture trends in research output that are common to all movers – for example, a decline in output in the years after a move, or a burst of output before one.

The key parameter of interest in equation 2 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 percentage change in log output in the years around the move scaled by the size and direction of the move, δ_i. We expect the log output of one's destination to not be predictive of log output before a move, but to be predictive of log output 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

---

[16] 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.

magnitude of the jump in $\theta_{r(i,t)}$ after a move captures how much place-specific factors influence log output on average. In other words, $\theta_{r(i,t)} = 1$ indicates that the mover's log output has converged completely to the average at the destination location and would imply that log output is completely influenced by place-specific factors. On the other hand, $\theta_{r(i,t)} = 0$ indicates no convergence of the mover's log output to that at the destination and would imply no role for place-specific factors. We report robust standard errors clustered at the institution level.[17]

IV. Estimating Institution and Researcher Effects

A. Characteristics of Scientist Movers

Using the sample of authors whose research was described in the above, we now focus on the sample of authors who moved between institutions (this is the principal reason that we considered a broad set of journals because this enlarges the pool of movers). We define 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.

Table 3 provides summary statistics for movers and non-movers in our sample. Among US-based researchers, we are able to identify 37,809 scientist-movers, and 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 563,620 papers, 294,782 were authored by 331,846 US-based researchers). On average, these movers relocated 1.4 times during our time frame. Due to potential noise in the accuracy of OpenAlex's author disambiguation algorithm and our affiliation imputation method, we focus on a set of 25,845 researchers who only move once. On average, movers tend to come from slightly smaller cities, but produce more output on average than non-movers. The first-time movers represent about 60% of the moving sample and are representative of the larger moving population.

---

[17] Note that if there is classical measurement error in δ_i, as there likely is for smaller institutions, then this would bias our estimates towards zero. As a robustness check, we estimate our main specification on a set of "larger institutions" defined as those with at least 100 authors and 10 movers and find that the estimated place effects increase to 60-70%.

B. Estimates from Two Way Fixed Effects

The two-way fixed effect model in 1 provides the simplest set of estimates, and is a simple way to perform a decomposition exercise where we estimate institution-specific effects ($\hat{\gamma}_j$) using our sample of movers and non-movers. These estimates are consistent as long as the identifying assumptions noted above are met.

Panel A of Table 4 examines the proportion of cross-sectional institutional variance in research output that can be attributed to researcher versus institutional effects, and we quantify these two components as (1) the share of cross-institution variance that would be eliminated if all institution effects 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: $S^{var}_{researcher} = 1 - \frac{Var(\bar{\gamma}_j)}{Var(\bar{y}_j)}$. This analysis reveals that equalizing institutional effects would eliminate over 73% of the observed variance, while equalizing researcher effects would eliminate 49% of the variance.

With these estimates, we can also examine whether the relative roles of institutions and individuals vary across the distribution of institutional output (Panel B of Table 4). This would happen if the most or least productive institutions have a disproportionate role for individuals or institutional effects. The columns of this table offer different ways to define top and bottom institutions: the top and bottom 5%, 10%, 25% and 50% (we should observe the largest differences between the top and bottom 5%). The first row is the difference in average researcher output between the two groups of institutions: $\hat{y} - \hat{y_{R'}}$. The second row is the difference due to the characteristics of the researcher, $\hat{\alpha_R} - \hat{\alpha_{R'}}$ and the third row is the remaining difference due to the effects of the institution $\hat{\gamma_R} - \hat{\gamma_{R'}}$. The last two rows convert the components of the researcher and the institution into relative shares of the overall difference. Our findings show that approximately 54% of the observed variation in researcher output between institutional groups derives from differential place characteristics, with the remaining 46% attributable to differences in researcher-specific factors. The magnitudes are consistent with the analysis of the event study, which suggested a place-based share of 50-60%. Note that the estimates will differ because the additive decomposition exercise is analyzing the 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's share across different partitions is consistent with the linear relationship shown in Figure 6.

Next, we turn to understanding whether the assumptions underlying the two-way model are supported by a more robust event study.

C. Estimates from Baseline Event Study

We motivate the event study analysis by first showing the identifying variation behind our movers specification. Appendix Figure A1 presents the distribution of our key independent variable, $\hat{\delta}_i$, the difference in average log output in a mover's destination and origin institution. By construction, $\hat{\delta}_i = 0$ indicates equal average log output 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 -0.091 (meaning that the typical move is to an institution whose output is approximately 9 percent lower than the institution of origin) and relatively symmetric, indicating an approximately equal number of moves from low-to-high-output institutions ($\hat{\delta}_i > 0$) as there are moves from high-to-low-output institutions ($\hat{\delta}_i < 0$).

It is natural to ask why scientists move. The most direct evidence is in (Lerner et al., 2024) who conduct a survey of moving scientists, which reveals that the modal reason for moving was better salary, followed by better resources and quality of life. It is possible for institutions with lower per capita output to offer higher salaries, resources, and quality of life, and we were reassured that Appendix Figure A1 shows moves in both directions (so not only to more productive institutions), and to institutions of similar research output as the institute of origin.

To contextualize a move size of -0.091, the average "wandering scholar" in our sample is moving from the University of Chicago to Baylor College of Medicine, and in Table A6, we list the top 50 institutions in the US based on average researcher output (in levels, not logs).[18]. For illustration, 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 variation driving our estimates. However, when estimating regressions, we include all institutions.

As an initial exploration of the impact of moving to a new institution on a scientist's log output, we plot the post-move change in scientist log output (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 output as their average post-move log output minus their average pre-move log output. The slope of the fitted line indicates the extent to which moving to a higher-output institution influences a scientist's log output. If place fully explained log output differences, the slope would 1. Conversely, if individual factors fully explained the variation, the slope would be 0. Figure 6 shows a slope of 0.73 suggesting that 73% of the variation in scientist log output can be attributed to institutional factors, while the remaining 27% 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, which means that moves to relatively more productive institutions are quantitatively similar, but with the opposite sign, to moves to relatively less productive institutions.

We present the results of the event study from equation 2 in Figure 7a for the period 1945-1923. 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 output. 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 output factors do not drive the decision to move. Log output 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 output over time[19]. This estimate is slightly lower than the 72% static estimate from Figure 6, reflecting the dynamic nature of place-specific effects and controlling for unobservable researcher attributes through individual fixed effects.

---

[18] We define move size as the difference in the average log output of two institutions, which is not the same as the differences of log average differences. Table A6 shows the average research output in levels for an institution; so logging these values produces the log of an average and will not be the same as move size

[19] We run the same specification using research output without any citation weighting as an alternative outcome and get similar results.

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 output. Moreover, as a way of examining the potential effect of measurement in δ_i, if we restrict the sample to institutions with at least 100 authors and 10 movers between 2015-2023, then we get essentially the same result.

D. Heterogeneity by Size of Move and Researcher Characteristics

To understand the underlying mechanisms of how place influences research output, 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 output institution and negative ($\hat{\delta}_i < 0$) if a researcher moves to a lower output institution. Higher output 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 output places but don't unlearn these skills when moving to lower output places, then we would expect to see larger place effects for positive moves. We separate the estimated effect of the place by the direction of movement in Figure A2. These plots confirm our theory as we see a difference in place effects between scientists moving to higher- versus lower-output 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 "stars" in discovery, we examine the influence of star researchers in explaining our place effect estimates (Azoulay et al., 2012, 2019; Waldinger, 2016). We define "stars" as researchers in the top 95th percentile of their institution's output distribution and modify our definition of δ_i to represent the difference between the average output 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 output of movers, with the estimated place effect around 0.4. In other words, star scientists and everything associated with them (which can include research laboratories and students), explain 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 output when a star relocates in Figure 8b. While the estimates are noisy, we see that the point estimates range from -0.3 to -0.8, implying that a coauthor's output decreases by 30 to 80% when their coauthor leaves and is considered a star researcher.

Motivated by prior research on the life-cycle output 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 as measured by the number of scientists in these cities excluding their own institution. This suggests that neither age nor size of the scientific agglomeration significantly impacts the role of place effects in changing researcher output.

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 output 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 output.

E. Robustness

E.1. Subfields

To address concerns about selection bias related to output 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 benefit from breakthrough discoveries in these fields through knowledge spillovers. 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 effects. This consistency across specifications provides compelling evidence that our observed effects represent genuine phenomena rather than artifacts of differential selection or field-specific output trajectories.

E.2. Alternative Samples

To assess the robustness of our findings to the selection of authors, we replicate our analysis for (1) only the first and last authors and (2) all authors. For first and last authors, the estimated institutional share of output 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. In contrast, 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 output differences persist in author roles, their magnitude varies depending on whether the focus is on the lead researchers or the entire research team.

E.3. Alternative Specifications

An alternative specification to our main estimating equation 2 is to not include the set of relative year fixed effects, $\rho_{r(i,t)}$. Including these fixed effects allows for a common trend in research output across all movers regardless of size and direction of the move. We show in Appendix Figure A7 that not including these covariates has little effect on our results.

V. 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. We note the importance of large research clusters and particular institutions in producing a disproportionate share of the world's science, but note substantially more variation in the geographic dispersion of average research output. By tracking the research output of scientists who move between institutions, we estimate that 50 to 60% of the variation in researcher output is attributable to place-specific factors, which means a significant portion, at least 40% is the consequence of more productive institutions selecting more productive researchers, as opposed to making them more productive. These effects have not diminished over time, even as across-institution collaborations have become easier, and underscore the significant role of institutional factors in fostering scientific output in the life sciences. Consequently, making public or philanthropic funding less generous at institutions that have high per-scientist output will directly reduce the production of knowledge, especially commercially relevant knowledge. Almost two-thirds of institutional effects are predicted by the presence of star researchers, which will include the infrastructure they require and students that they attract that increase the output of other scientists. This is consistent with the work of (Waldinger, 2016) who noted that the dismissal of famous scientists in Nazi Germany led to persistent declines in research output. Similarly, (Azoulay et al., 2019) notes that the death of a star researcher causes large declines in the research output of their collaborators.

Our results contribute to the broader research agenda on the production function of innovation. A rich literature has examined other factors that shape research output, including funding mechanisms, team composition, competition, and career incentives (Lach and Schankerman, 2008; Azoulay et al., 2013; Tabakovic and Wollmann, 2019; Azoulay et al., 2019; Bloom et al., 2020; Myers, 2020; Arora et al., 2021; Acemoglu, 2023; Babina et al., 2023; Hill and Stein, 2023, 2024; Tham et al., 2024). We contribute to this enterprise by focusing on one key determinant, the location where research happens. Our work also complements research on sources of patenting, which is a downstream outcome of fundamental science research (Jaffe, 1989; Trajtenberg et al., 1997; Jensen and Thursby, 2001; Thursby et al., 2007; Belenzon and Schankerman, 2013; Williams, 2017; Bell et al., 2019; Koning et al., 2020; Akcigit et al., 2021; Tartari and Stern, 2021; Hausman, 2022; Moretti et al., 2023; Lerner et al., 2024). If the forces that influence patenting also shape fundamental research– as they are often undertaken by the same individuals and likely to be complementary activities– then the role of place in for broader medical innovation may be even larger than we estimated.

In the context of life-science discovery, the institution of a researcher has a consequential effect on their output. Consequently, if funders want to maximize the impact of their resources on the fundamental scientist discovery, institution-level resources (such as a gift that is not associated with a particular scientist) should be allocated to the most productive institutions. If funders face the choice between funding two scientists with the same output, then funding the scientist at the more productive institution will generate more research. The potential effect sizes are large: Appendix Tables A6 and A7 show a 3x-4x variation in average researcher output across institutions, even among the list of most productive institutions, which is a truncated portion of the true distribution of institutional output. Our estimates predict that over half of this variation is causal, which has implications for scientists who want to move from one institution to another, especially from high- to low-output institutions in return for a more favorable offer. In our analysis, we studied scientist movers in the United States, but if our results generalize to moves in other countries, there will be consequential reductions in knowledge creation if US-based researchers leave for ex-US institutions with lower average output per scientist (the potential effect of this can be seen by comparing the average output of leading US institutions in Appendix Table A6 to Table A7)

At the same time, institutions may want to increase their output. Our results suggest that agglomeration is not a prerequisite for high research output (which echoes a point made in (Ellison et al., 1997) about agglomerations in manufacturing). In other words, agglomerations do not make researchers more productive and reflect the agglomeration of productive scientists. Second, factors associated with having star researchers explain this effect; the departure of stars also reduces the output of scientists in the institution that they left. Investing in the resources that attract star researchers may be one way to increase the output of institutions. However, we emphasize that such a reallocation would possibly require large resources (especially in light of Myers (2020), who notes that large sums of money are required to get scientists to switch fields, let alone institutions).

These results suggest several avenues for future research. A key question is causally identifying the institutional characteristics that make some institutions more productive than others. In our exploration, we note that institutions with higher R&D expenditures were more productive, but this was not a causal statement; rather, it was a simple way of validating the institution effects. Understanding the link between funding and research output, as in Myers and Lanahan (2022), and exploring how different institutions structure their research environments can inform best practices to maximize the effectiveness of research funding. Finally, if institutional environments causally affect scientists' output, as our evidence indicates, then initiatives that focus on young investigators could yield high returns if focused on enabling talent at the most productive institutions (Azoulay et al., 2011, 2019, 2021). Understanding these relationships would be a fruitful way to increase fundamental science discovery, whose implications for welfare and science are enormous.

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.

Allison, P. D. and J. S. Long (1990). Departmental Effects on Scientific Productivity. *American Sociological Review 55*(4), 469–478. Publisher: [American Sociological Association, Sage Publications, Inc.].

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. 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.

Azoulay, P., J. S. G. Zivin, and B. N. Sampat (2012, March). The Diffusion of Scientific Knowledge across Time and Space: Evidence from Professional Transitions for the Superstars of Medicine. In *The Rate and Direction of Inventive Activity Revisited*, pp. 107–155. University of Chicago Press.

Babina, T., A. X. He, S. T. Howell, E. R. Perlman, and J. Staudt (2023, May). Cutting the Innovation Engine: How Federal Funding Shocks Affect University Patenting, Entrepreneurship, and Publications*. *The Quarterly Journal of Economics 138*(2), 895–954.

Belenzon, S. and M. Schankerman (2013, July). Spreading the Word: Geography, Policy, and Knowledge Spillovers. *The Review of Economics and Statistics 95*(3), 884–903.

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.

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.

Gruber, J., S. Johnson, and E. Moretti (2022, September). Place-Based Productivity and Costs in Science.

Guzman, J., F. Murray, S. Stern, and H. Williams (2023, July). Accelerating Innovation Ecosystems: The Promise and Challenges of Regional Innovation Engines. In *Entrepreneurship and Innovation Policy and the Economy, volume 3*. University of Chicago Press.

Hall, B. H. and J. Lerner (2010, January). Chapter 14 - The Financing of R&D and Innovation. In B. H. Hall and N. Rosenberg (Eds.), *Handbook of the Economics of Innovation*, Volume 1 of *Handbook of The Economics of Innovation, Vol. 1*, pp. 609–639. North-Holland.

Hausman, N. (2022, July). University Innovation and Local Economic Growth. *The Review of Economics and Statistics 104*(4), 718–735.

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.

Howell, S. T. (2017, April). Financing Innovation: Evidence from R&D Grants. *American Economic Review 107*(4), 1136–1164.

Hvide, H. K. and B. F. Jones (2018, July). University Innovation and the Professor's Privilege. *American Economic Review 108*(7), 1860–1898.

Jaffe, A. B. (1989). Real Effects of Academic Research. *The American Economic Review 79*(5), 957–970. Publisher: American Economic Association.

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.

Lach, S. and M. Schankerman (2008). Incentives and Invention in Universities. pp. 32.

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.

Tabakovic, H. and T. G. Wollmann (2019, October). The impact of money on science: Evidence from unexpected NCAA football outcomes. *Journal of Public Economics 178*, 104066.

Tartari, V. and S. Stern (2021, May). More than an Ivory Tower: The Impact of Research Institutions on the Quantity and Quality of Entrepreneurship.

Tham, W. Y., J. Staudt, E. R. Perlman, and S. D. Cheng (2024, February). Scientific Talent Leaks Out of Funding Gaps. arXiv:2402.07235 [econ].

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.

Trajtenberg, M., H. , Rebecca, , and A. Jaffe (1997, January). University Versus Corporate Patents: A Window On The Basicness Of Invention. *Economics of Innovation and New Technology 5*(1), 19–50. Publisher: Routledge _eprint: https://doi.org/10.1080/10438599700000006.

Waldinger, F. (2016, December). Bombs, Brains, and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge. *The Review of Economics and Statistics 98*(5), 811–831.

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: Cities with Largest Annual Output per Researcher, 2015–2023

Rank	City	Average Author Output	Share of World Output (%)
1	Boston–Cambridge–Newton, MA–NH	1.58	9.79
2	Bay Area, CA	1.56	7.67
3	Trenton, NJ	1.42	0.54
4	Dallas–Fort Worth–Arlington, TX	1.35	0.86
5	New York–Newark–Jersey City, NY–NJ–PA	1.19	6.67
6	San Diego–Carlsbad, CA	1.05	2.25
7	Seattle–Tacoma–Bellevue, WA	1.02	1.68
8	New Haven–Milford, CT	1.02	1.20
9	St. Louis, MO–IL	0.97	1.39
10	Worcester, MA–CT	0.97	0.28
11	Ithaca, NY	0.85	0.78
12	Houston–The Woodlands–Sugar Land, TX	0.82	1.43
13	Austin–Round Rock, TX	0.81	0.30
14	Los Angeles–Long Beach–Anaheim, CA	0.76	1.85
15	Memphis, TN–MS–AR	0.74	0.34
16	Santa Cruz–Watsonville, CA	0.73	0.15
17	Philadelphia–Camden–Wilmington, PA–NJ–DE–MD	0.73	1.34
18	Baltimore–Columbia–Towson, MD	0.67	1.29
19	Chicago–Naperville–Elgin, IL–IN–WI	0.58	0.98
20	Rochester, MN	0.58	0.09

Notes: Table shows the top 20 cities by average author-year output in fundamental science research in life sciences journals. Research output 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. Share of world output is calculated using total publications across 2015–2023.

Table 2: Leading Institutions in Ten Cities with Largest Fundamental Science Output, 2015-2023

Rank	**Institution**	**Average Researcher Output**
I. Boston-Cambridge-Newton, MA-NH		
1.	Dana-Farber Cancer Institutee	2.81
2.	Massachusetts Institute of Technology	2.33
3.	Harvard University	1.64
4.	Mass General Brigham	1.52
5.	Boston Children's Hospital	1.15
II. Bay Area, CA		
1.	Stanford University	1.86
2.	Gladstone Institutes	1.77
3.	University of California, Berkeley	1.33
4.	University of California, San Francisco	1.31
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	1.35
V. New York-Newark-Jersey City, NY-NJ-PA		
1.	Memorial Sloan Kettering Cancer Center	2.24
2.	Novartis (United States)	1.72
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	1.73
2.	University of California, San Diego	1.00
3.	Scripps Research Institute	0.91
4.	Sanford Burnham Prebys Medical Discovery Institute	0.51
VII. Seattle-Tacoma-Bellevue, WA		
1.	Allen Institute	2.40
2.	Fred Hutch Cancer Center	1.42
3.	University of Washington	0.85
VIII. New Haven-Milford, CT		
1.	Yale University	1.01
IX. St. Louis, MO-IL		
1.	Washington University in St. Louis	1.06
2.	Saint Louis University	0.40
X. Worcester, MA-CT		
1.	University of Massachusetts Chan Medical School	0.97

Notes: Table shows the leading institutions in the ten most productive cities by average researcher output in fundamental science research in life sciences journals. The 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 3: Movers Design: Summary Statistics

	Mean	Std. Dev.
A. Individual-Level		
Everyone (N = 296,757)		
Years Actively Publishing	1.79	2.03
Average Team Size	3.62	0.66
Annual City Size	7386.67	7404.94
Annual Productivity	0.39	2.08
Lifetime Productivity	0.79	4.59
Number of Moves	0.13	0.33
Non-Movers (N = 258,948)		
Years Actively Publishing	1.48	1.51
Average Team Size	3.63	0.67
Annual City Size	7369.74	7570.54
Annual Productivity	0.38	2.16
Lifetime Productivity	0.60	3.57
Movers (N = 37,809)		
Years Actively Publishing	3.90	3.42
Average Team Size	3.56	0.56
Annual City Size	7502.61	6150.92
Annual Productivity	0.49	1.42
Lifetime Productivity	2.13	8.70
One-time Movers (N = 25,845)		
Years Actively Publishing	4.12	3.58
Average Team Size	3.56	0.56
Annual City Size	7504.57	6127.13
Annual Productivity	0.49	1.32
Lifetime Productivity	2.24	9.21
B. City-Level		
Number of Institutions	169.46	173.15
Annual City Size	7796.07	8134.21
Annual Productivity	0.38	2.04
Lifetime Productivity	47.25	480.80

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 4: Decompositions of Researcher Output

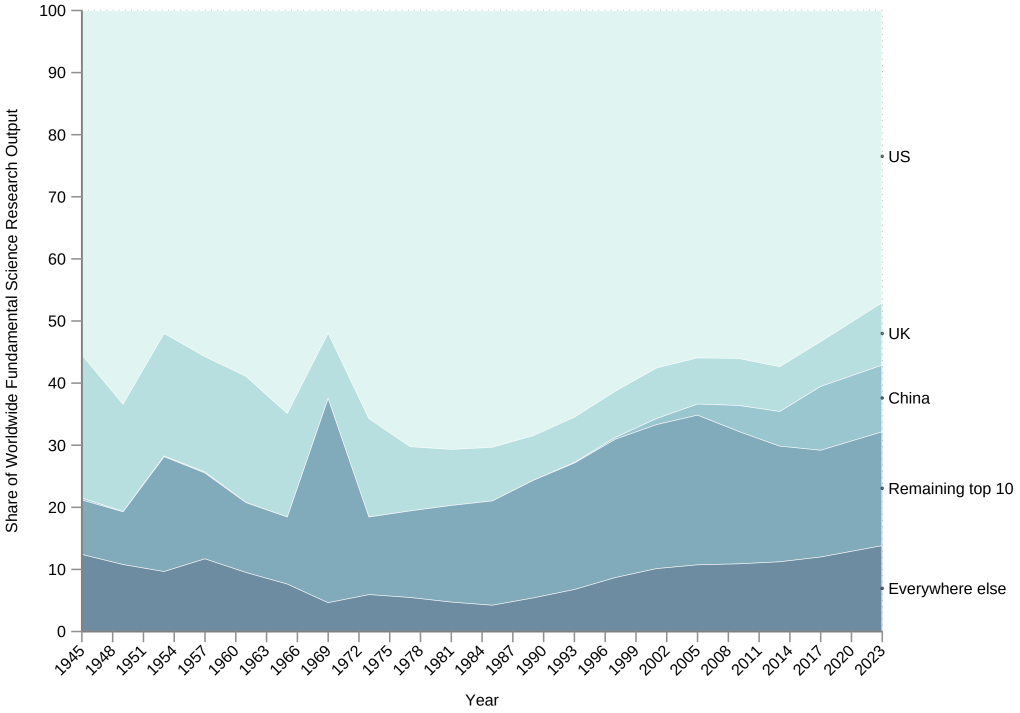
Panel A: Variance Decomposition of Average Researcher Output	
Variance of total researcher output	0.16
Attributable to institution effects	0.08
Attributable to researcher effects	0.04
Attributable to institution & researcher effects	0.28
Share of variance that would be reduced if:	
Researcher effects were made equal	0.49
Institution effects were made equal	0.73

Panel B: Additive Decomposition of Average Researcher Output

Top & Bottom	**5%**	**10%**	**25%**	**50%**
Difference in Researcher Output				
Overall	3.52	2.79	1.87	1.14
Researchers	1.66	1.33	0.84	0.53
Institutions	1.85	1.46	1.04	0.60
Share of difference attributable to				
Researchers	0.47	0.48	0.45	0.47
Institutions	0.53	0.52	0.55	0.53

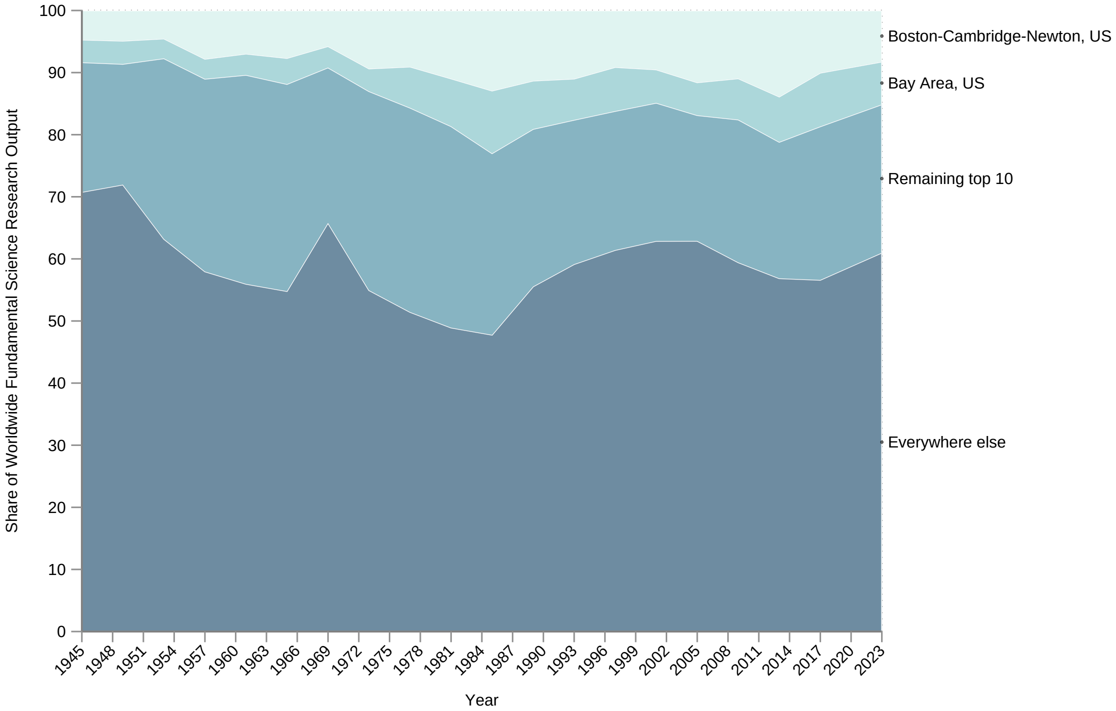
Notes: Panel A reports the variance decomposition of research output across institutions. The first part reports the overall variance and its components, while the second part illustrates the share of variance that would be reduced if researcher or institution effects were made equal. Panel B decomposes the difference in researcher productivity across top and bottom percentile groups. The first part presents the absolute difference in productivity, while the second part shows the share of this difference attributable to individual researchers versus institutions. Sample includes all authors (movers and non-movers) from 1945–2023. We reduce noise in our estimates by restricting to institutions with at least 100 authors per year (Panel A) and at least 100 authors and 10 movers ever (Panel B).

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



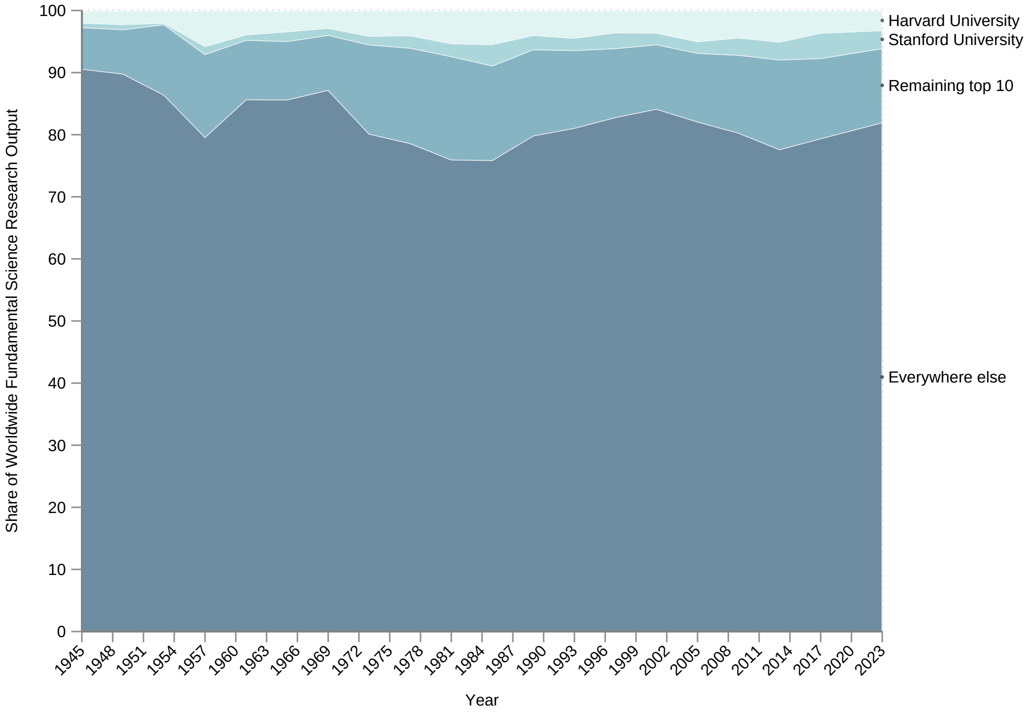
Notes: Figure shows the country-level trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, 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



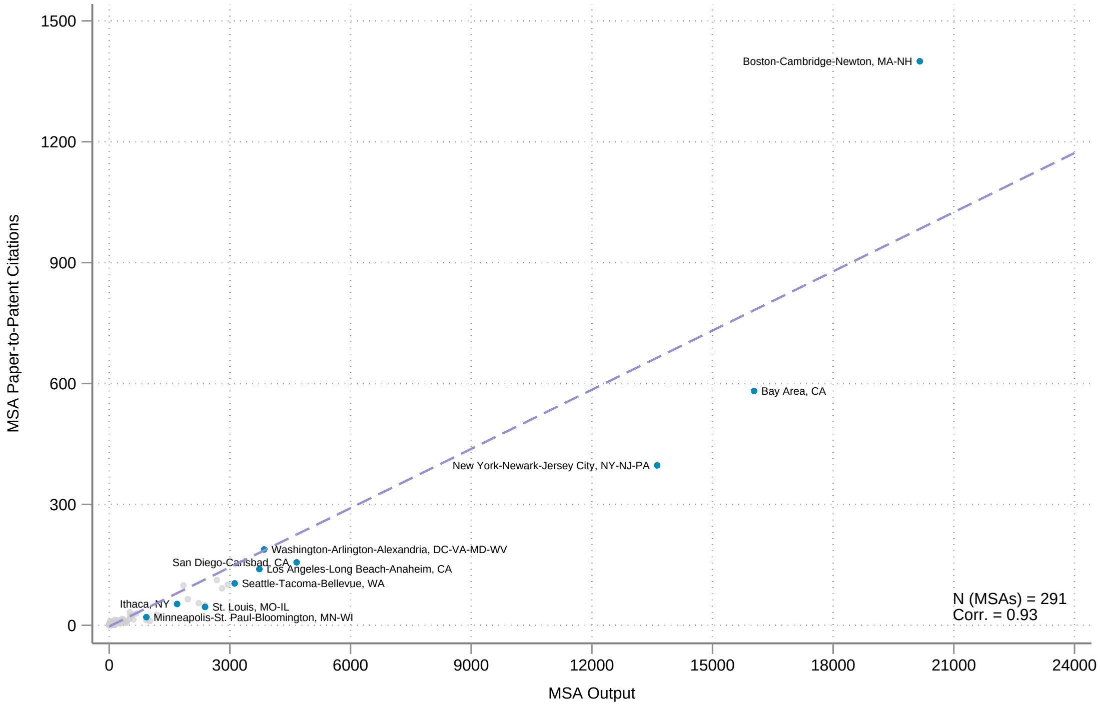
Notes: Figure shows the city-level trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, 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



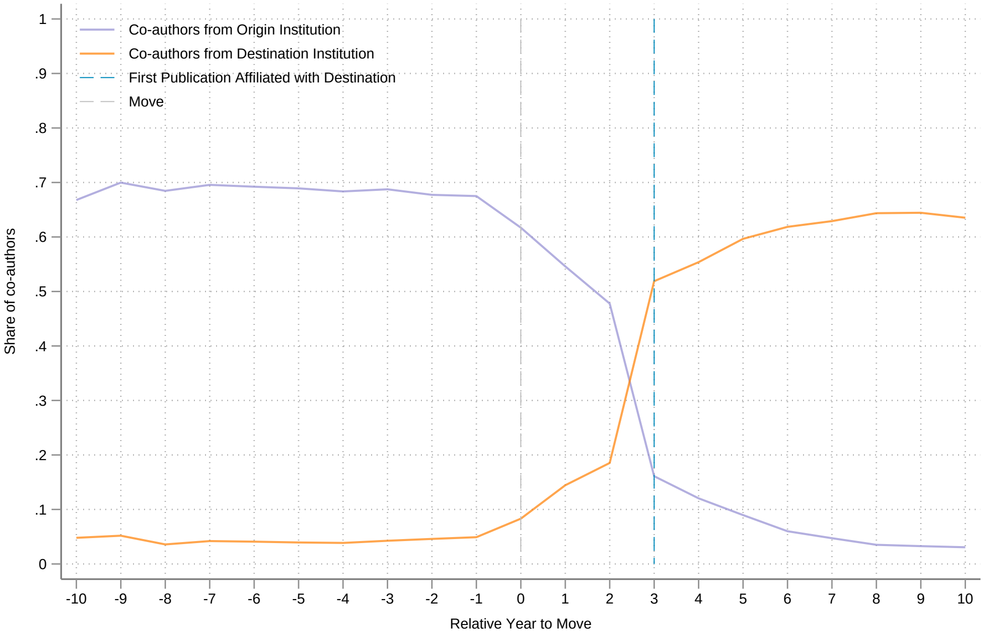
Notes: Figure shows the institution-level trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, 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



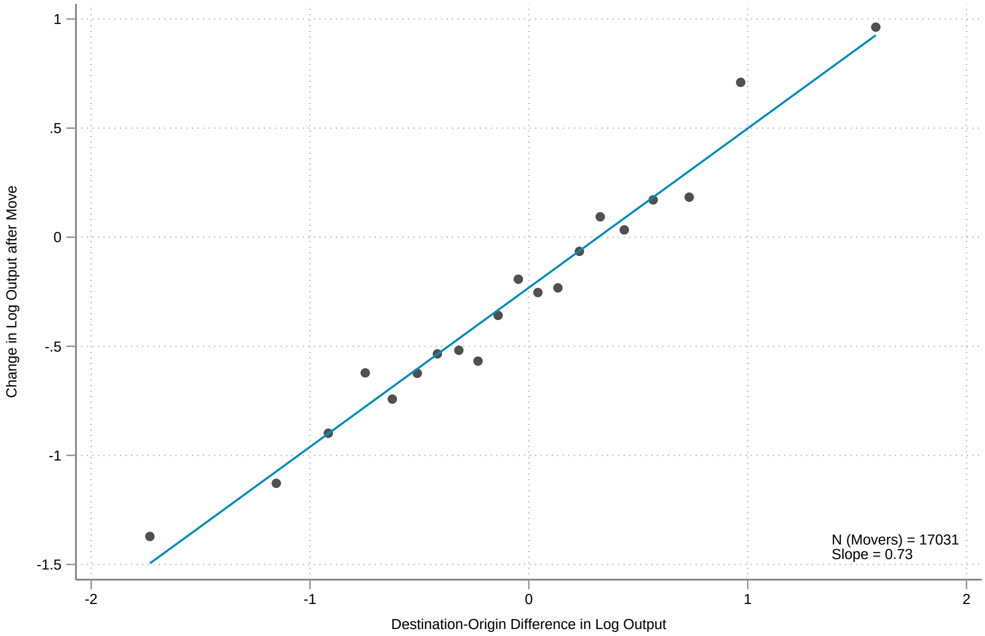
Notes: Figure shows a scatter plot of the number of papers in patents citations against the citation-adjusted measure of fundamental science research produced at the MSA-level. Around 15.5% of papers in our sample have ever been cited in a patent. There are N = 291 MSAs in our sample.

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.

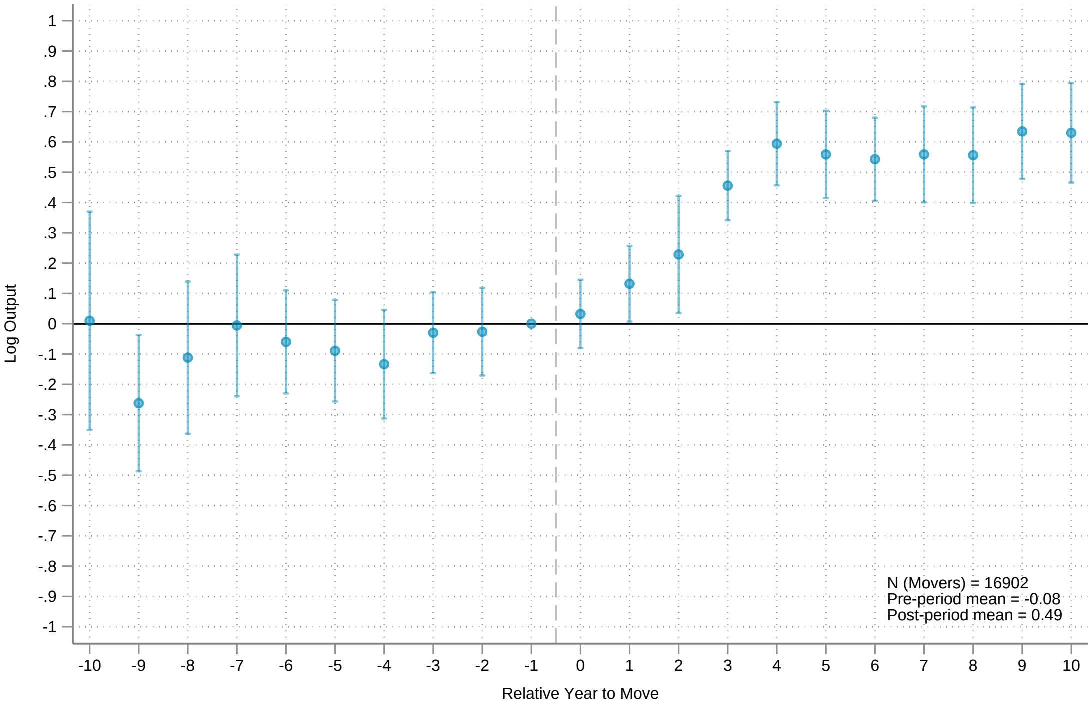
Figure 6: Change in Log Output by Move Size



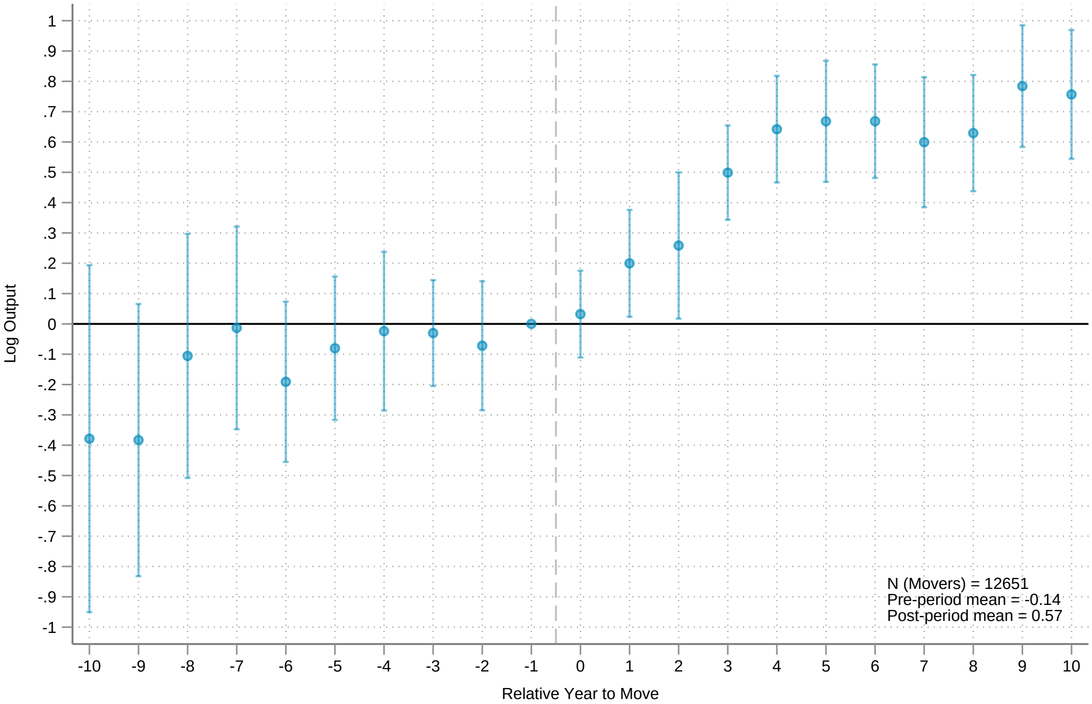
Notes: Figure shows the change in log output relative to a mover's move size. Move size is defined as the difference in average output between a mover's origin and destination institution.

Figure 7: Output Change in Response to Move

(a) 1945-2023



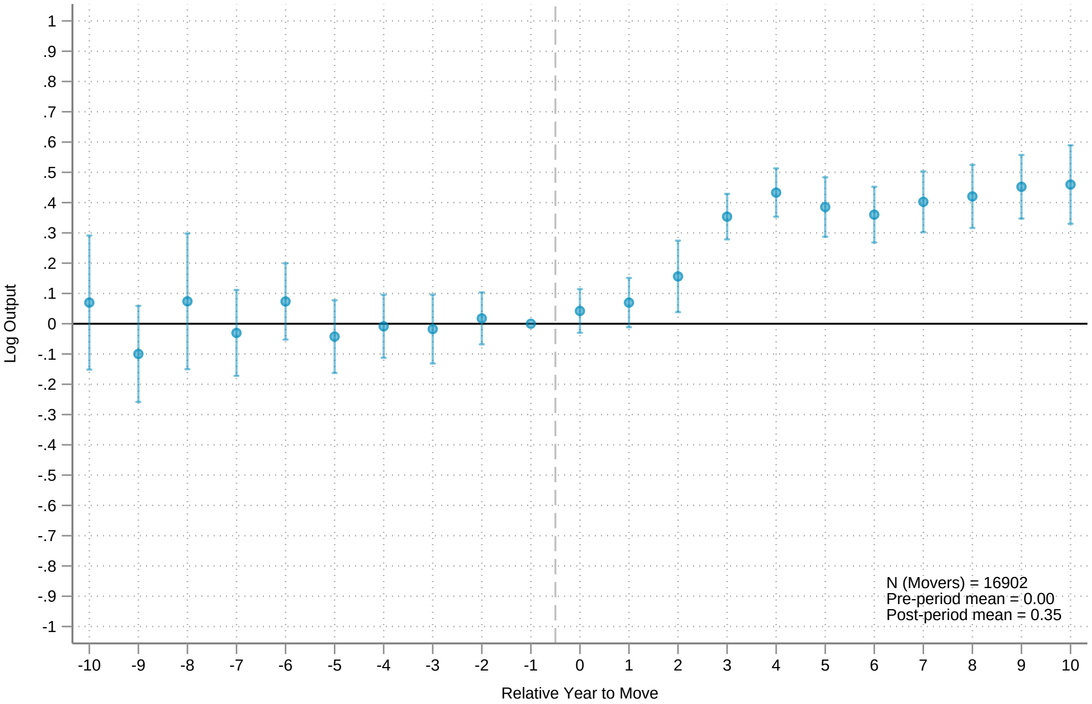
(b) 1995-2023



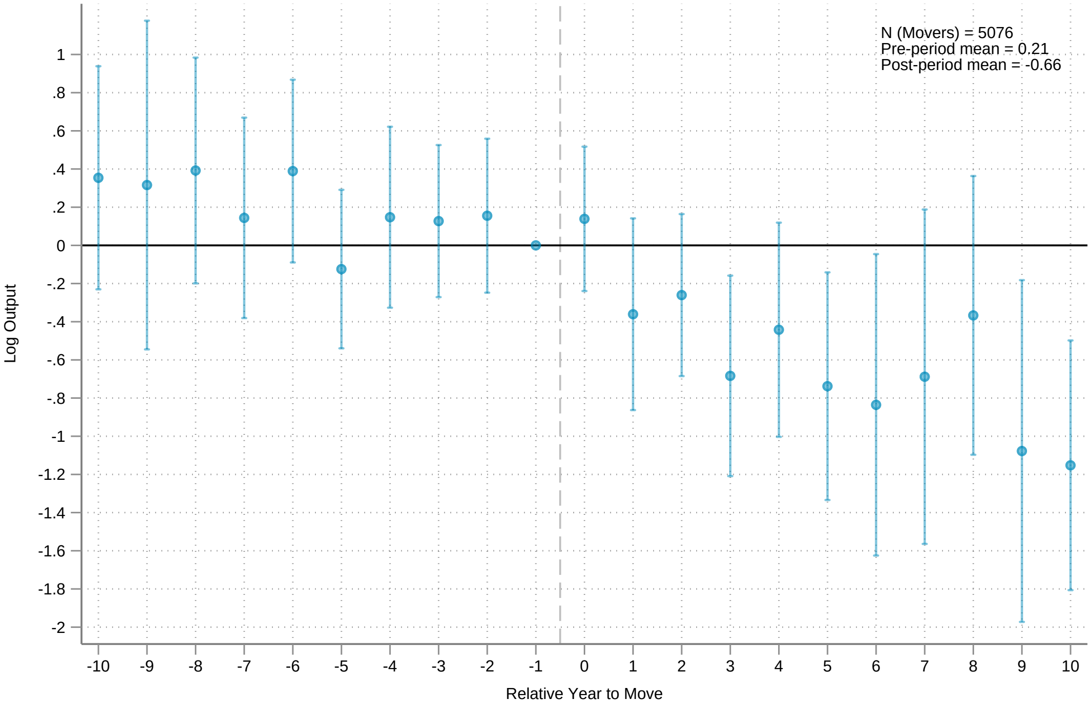
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2. The coefficient for relative year -1 is normalized to 0. The dependent variable is log output and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at 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 Output



(b) Role of Star Leaving on Coauthor Output



Notes: Panel (a) shows the coefficients $\hat{\theta}$ estimated from equation 2 using the definition of move size as defined by the productivity of stars in the location. Panel (b) reports coefficients from a canonical event–study that tracks how the log productivity of the star's coauthors evolves in the years surrounding the move. For both, 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

A. Additional Exhibits

Appendix Table A1: Journal Sample Sizes

Sample	**Journal**	**% Fundamental**	**# Fundamental**
CNS	Science	37.10	39,454
	Nature	64.36	64,034
	Cell	93.89	16,163
Leading Life-Sciences Journals	Neuron	93.14	8,464
	Nature Genetics	90.59	5,775
	Nature Medicine	57.11	4,625
	Nature Biotechnology	43.14	3,445
	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	81.44	210,334
	Journal of Bio Chem	95.07	171,693
	Oncogene	97.92	15,604
	The FASEB Journal	69.06	12,718

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 Output			
Unweighted vs. Impact Factor Weighted	0.9934	0.9495	0.9300
Unweighted vs. Impact Factor & Citation Weighted	0.9953	0.9042	0.8667
Papers-in-Patents Citations			
All Cites vs. Front Page Cites	0.9996	0.9967	0.9907
All Cites vs. Body Cites	0.9993	0.9933	0.9810

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 Output 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	0.9916	0.9598
Impact Factor & Citation Weighted vs. Front Page Cites	0.9948	0.9929	0.9682
Impact Factor & Citation Weighted vs. Body Cites	0.9939	0.9857	0.9337

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: Largest US Institutions by Average Annual Output and Share of World Output, 2015–2023

Rank	**Institution**	**Annual Output**	**Share of World Output (%)**
1	Harvard University	678.78	3.63
2	Stanford University	632.12	3.38
3	Massachusetts Institute of Technology	389.78	2.09
4	University of California, San Francisco	349.68	1.87
5	Mass General Brigham	291.98	1.56
6	National Institutes of Health	272.94	1.46
7	Memorial Sloan Kettering Cancer Center	223.44	1.20
8	Johns Hopkins University	222.75	1.19
9	Yale University	219.00	1.17
10	University of California, San Diego	218.64	1.17
11	Columbia University	217.52	1.16
12	University of Pennsylvania	193.20	1.03
13	Dana-Farber Cancer Institute	189.01	1.01
14	University of California, Berkeley	186.16	1.00
15	Washington University in St. Louis	178.82	0.96
16	University of Washington	177.00	0.95
17	Icahn School of Medicine at Mount Sinai	168.34	0.90
18	The University of Texas Southwestern Medical Center	154.93	0.83
19	NYU Langone Health	146.74	0.79
20	Cornell University	145.96	0.78
21	New York Genome Center	140.14	0.75
22	The University of Texas MD Anderson Cancer Center	130.37	0.70
23	University of California, Los Angeles	116.44	0.62
24	Mammoth Biosciences	114.05	0.08
25	Rockefeller University	112.19	0.60
26	University of Michigan–Ann Arbor	105.21	0.56
27	University of North Carolina at Chapel Hill	100.49	0.54
28	Duke University	99.16	0.53
29	Princeton University	95.87	0.51
30	University of Chicago	88.14	0.47
31	California Institute of Technology	86.62	0.46
32	VA St. Louis Health Care System	83.23	0.33
33	Baylor College of Medicine	75.96	0.41
34	Boston Children's Hospital	72.28	0.39
35	University of Minnesota	71.24	0.38
36	Northwestern University	64.11	0.34
37	University of Pittsburgh	61.65	0.33
38	Scripps Research Institute	61.23	0.33
39	St. Jude Children's Research Hospital	60.18	0.32
40	Salk Institute for Biological Studies	59.65	0.32
41	University of Southern California	55.60	0.30
42	Vanderbilt University	53.44	0.29
43	The University of Texas at Austin	52.72	0.28
44	University of Massachusetts Chan Medical School	51.83	0.28
45	Janelia Research Campus	48.73	0.26
46	La Jolla Institute for Immunology	46.75	0.25
47	Fred Hutch Cancer Center	45.69	0.24
48	Google	45.12	0.24
49	Emory University	41.78	0.22
50	Allen Institute	41.62	0.22

Notes: Table shows the leading 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. Share of world output is calculated using total output across 2015-2023. 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: Largest International Institutions by Average Annual Output and Share of World Output, 2015–2023

Rank	Institution	Country	Annual Output	Share of World Output (%)
1	Chinese Academy of Sciences	China	405.31	2.17
2	University of Oxford	United Kingdom	230.76	1.23
3	University of Cambridge	United Kingdom	223.67	1.20
4	Max Planck Society	Germany	222.95	1.19
5	Karolinska Institutet	Sweden	132.44	0.71
6	Weizmann Institute of Science	Israel	127.17	0.68
7	Helmholtz Association of German Research Centres	Germany	125.49	0.67
8	Peking University	China	113.63	0.61
9	Wellcome Sanger Institute	United Kingdom	107.49	0.58
10	DeepMind	United Kingdom	101.32	0.54
11	University College London	United Kingdom	95.11	0.51
12	Leibniz Association	Germany	93.77	0.50
13	University of Copenhagen	Denmark	85.97	0.46
14	University of Queensland	Australia	85.44	0.46
15	Institute for Integrative and Experimental Genomics	Germany	84.56	0.11
16	Heidelberg University	Germany	78.70	0.42
17	Imperial College London	United Kingdom	76.80	0.41
18	Tsinghua University	China	75.28	0.40
19	Medical Research Council	United Kingdom	73.35	0.39
20	University of Toronto	Canada	66.05	0.35
21	University of Hong Kong	Hong Kong	63.07	0.34
22	The University of Tokyo	Japan	61.08	0.33
23	ETH Zurich	Switzerland	60.17	0.32
24	École Polytechnique Fédérale de Lausanne	Switzerland	59.37	0.32
25	Centre National de la Recherche Scientifique	France	59.08	0.32
26	University of Edinburgh	United Kingdom	56.74	0.30
27	King's College London	United Kingdom	54.24	0.29
28	Institut Pasteur	France	54.02	0.29
29	The Netherlands Cancer Institute	Netherlands	48.67	0.26
30	Sun Yat-sen University	China	47.87	0.26
31	UNSW Sydney	Australia	47.61	0.25
32	Technical University of Denmark	Denmark	45.98	0.25
33	Inserm	France	44.42	0.24
34	Charité - Universitätsmedizin Berlin	Germany	44.07	0.24
35	The Francis Crick Institute	United Kingdom	43.94	0.24
36	Zhejiang University	China	43.43	0.23
37	Royal Netherlands Academy of Arts and Sciences	Netherlands	43.05	0.23
38	Barcelona Institute for Science and Technology	Spain	43.05	0.23
39	McGill University	Canada	42.66	0.23
40	KU Leuven	Belgium	42.20	0.23
41	RIKEN	Japan	41.51	0.22
42	European Molecular Biology Laboratory	Germany	41.30	0.23
43	Lund University	Sweden	40.94	0.22
44	Peking Union Medical College	China	40.84	0.22
45	LanzaTech	New Zealand	39.88	0.03
46	University of Helsinki	Finland	38.02	0.20
47	University of Zurich	Switzerland	37.78	0.20
48	Kyoto University	Japan	37.19	0.20
49	Ludwig Maximilian University of Munich	Germany	36.12	0.19
50	Westlake University	China	35.96	0.12

Notes: Table shows the leading 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. Share of world output is calculated using total output across 2015-2023. 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: US Institutions Ranked by Average Researcher Output, 2015-2023

Rank	Institution	Researcher Output
1	Dana-Farber Cancer Institute	2.81
2	Allen Institute	2.40
3	Massachusetts Institute of Technology	2.33
4	Memorial Sloan Kettering Cancer Center	2.24
5	Stanford University	1.86
6	Gladstone Institutes	1.77
7	Salk Institute for Biological Studies	1.73
8	Novartis (United States)	1.72
9	Icahn School of Medicine at Mount Sinai	1.69
10	Harvard University	1.64
11	Mass General Brigham	1.52
12	The University of Texas MD Anderson Cancer Center	1.49
13	Fred Hutch Cancer Center	1.42
14	Princeton University	1.42
15	California Institute of Technology	1.42
16	Cold Spring Harbor Laboratory	1.42
17	The University of Texas Southwestern Medical Center	1.35
18	Rockefeller University	1.34
19	University of California, Berkeley	1.33
20	University of California, San Francisco	1.31
21	Janelia Research Campus	1.26
22	Jackson Laboratory	1.20
23	Boston Children's Hospital	1.15
24	Columbia University	1.15
25	City Of Hope National Medical Center	1.13
26	NYU Langone Health	1.11
27	St. Jude Children's Research Hospital	1.08
28	University of Pennsylvania	1.06
29	Washington University in St. Louis	1.06
30	Yale University	1.02
31	University of California, San Diego	1.00
32	University of Massachusetts Chan Medical School	0.97
33	Cedars-Sinai Medical Center	0.95
34	University of Chicago	0.94
35	Scripps Research Institute	0.91
36	University of Washington	0.85
37	Cornell University	0.85
38	MSD (United States)	0.85
39	Beth Israel Deaconess Medical Center	0.84
40	The University of Texas at Austin	0.81
41	Johns Hopkins University	0.79
42	Carnegie Mellon University	0.76
43	The University of Texas Medical Branch at Galveston	0.76
44	University of California, Santa Cruz	0.73
45	University of Southern California	0.73
46	Baylor College of Medicine	0.73
47	National Institutes of Health	0.69
48	Lawrence Berkeley National Laboratory	0.67
49	Albert Einstein College of Medicine	0.66
50	University of California, Los Angeles	0.65

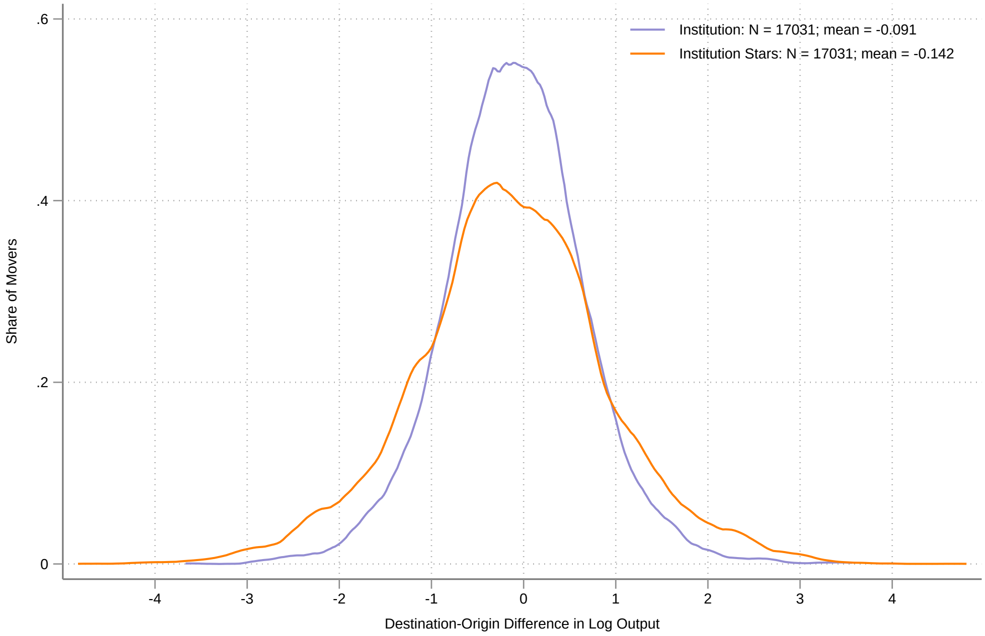
Notes: Table shows the leading 50 institutions by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers at each institution. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature 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: International Institutions Ranked by Average Researcher Output, 2015 - 2023

Rank	**Institution**	**Country**	**Researcher Output**
1	Wellcome Sanger Institute	United Kingdom	2.66
2	University of Trento	Italy	2.24
3	Royal Netherlands Academy of Arts and Sciences	Netherlands	2.24
4	Academic Medical Center	Netherlands	2.21
5	The Netherlands Cancer Institute	The Netherlands	1.96
6	Austrian Academy of Sciences	Austria	1.63
7	Weizmann Institute of Science	Israel	1.63
8	Princess Margaret Cancer Centre	Canada	1.60
9	Chongqing Medical University	China	1.57
10	Barcelona Institute for Science and Technology	Spain	1.56
11	Technical University of Denmark	Denmark	1.50
12	Cancer Research UK	United Kingdom	1.49
13	École Polytechnique Fédérale de Lausanne	Switzerland	1.36
14	KTH Royal Institute of Technology	Sweden	1.32
15	Novartis	Switzerland	1.29
16	Friedrich Miescher Institute	Switzerland	1.29
17	University of Hong Kong	Hong Kong	1.26
18	Institut Pasteur	France	1.19
19	Biomedical Research Council	Singapore	1.17
20	University of Iceland	Iceland	1.16
21	University of Cambridge	United Kingdom	1.13
22	Medical Research Council	United Kingdom	1.09
23	University of Oxford	United Kingdom	1.05
24	Roche (Switzerland)	Switzerland	1.05
25	University Medical Center Freiburg	Germany	1.02
26	The Francis Crick Institute	United Kingdom	1.00
27	University of Queensland	Australia	0.97
28	Tsinghua University	China	0.95
29	Institute of Cancer Research	United Kingdom	0.94
30	Leibniz Association	Germany	0.88
31	UNSW Sydney	Australia	0.88
32	Max Planck Society	Germany	0.87
33	Peking University	China	0.84
34	Helmholtz Association of German Research Centres	Germany	0.84
35	Hospital for Sick Children	Canada	0.82
36	Stockholm University	Sweden	0.81
37	Karolinska Institutet	Sweden	0.78
38	King’s College London	United Kingdom	0.77
39	University Health Network	Canada	0.76
40	Institut Curie	France	0.76
41	Technion – Israel Institute of Technology	Israel	0.76
42	ETH Zurich	Switzerland	0.75
43	QIMR Berghofer Medical Research Institute	Australia	0.71
44	Walter and Eliza Hall Institute of Medical Research	Australia	0.69
45	Keio University	Japan	0.68
46	Garvan Institute of Medical Research	Australia	0.67
47	Chinese Academy of Sciences	China	0.67
48	Korea Advanced Institute of Science and Technology	South Korea	0.67
49	RIKEN	Japan	0.66
50	University of Dundee	United Kingdom	0.66

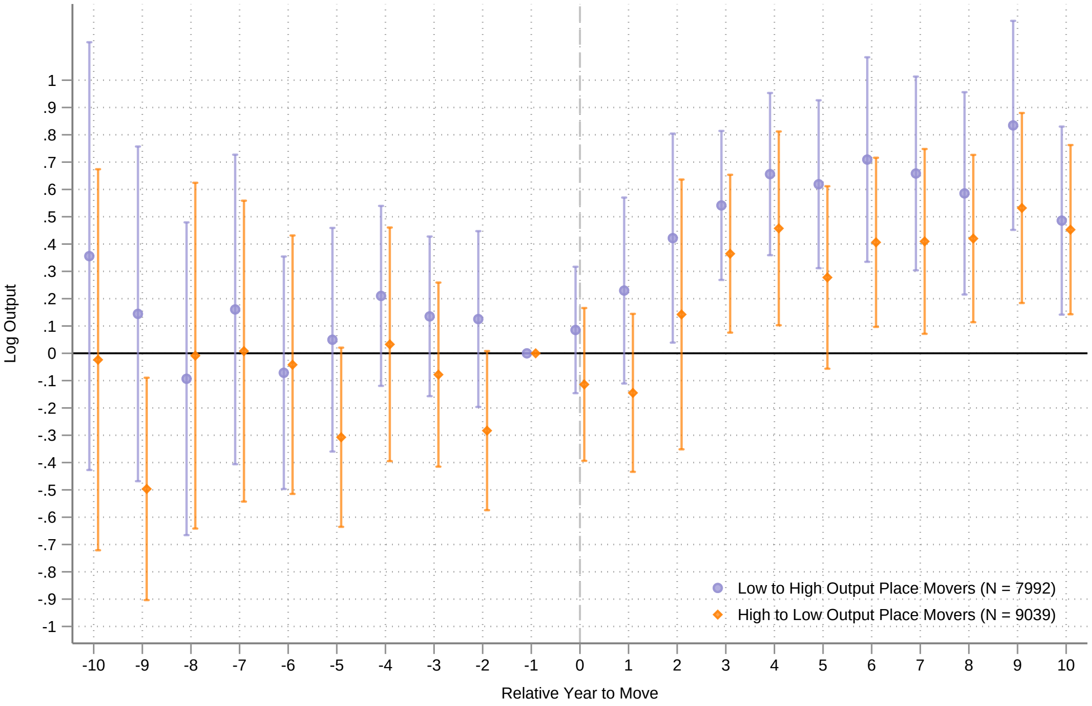
Notes: Table shows the leading 50 international institutions by average researcher productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers at each institution. Journals included are: *Cell, Nature, Science, Nature Cell 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



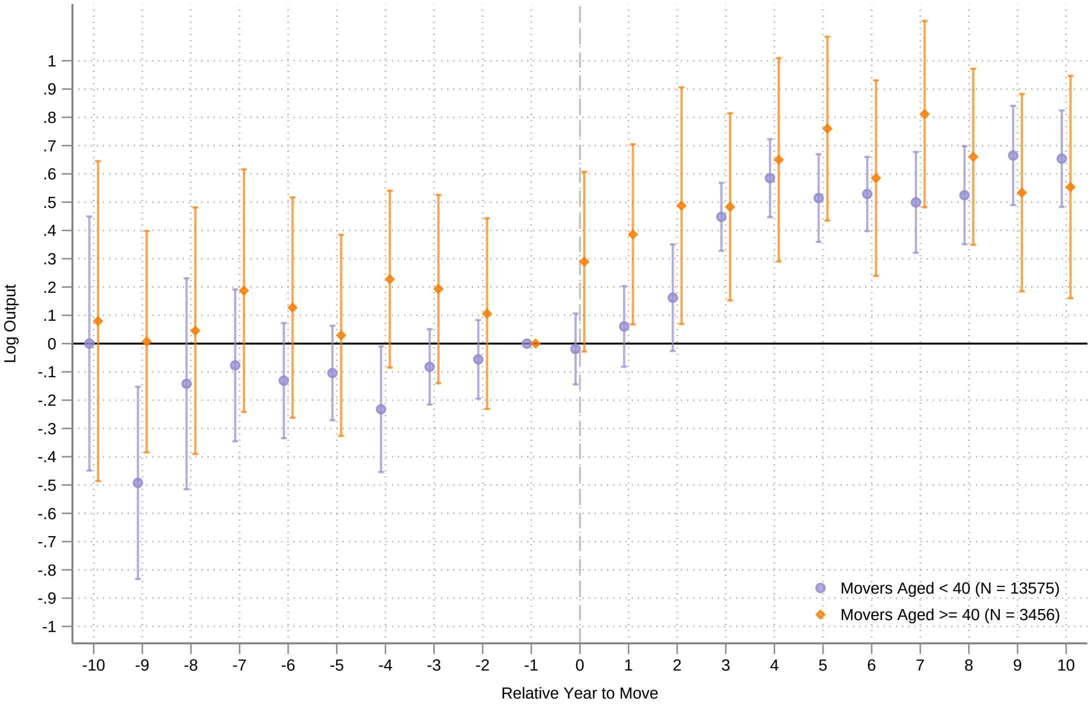
Notes: Figure shows the distribution of δ_i computed at various levels for scientist-movers. Here, δ_i measures the size of the move and is constructed as the average difference in productivity between the destination and origin location. We show the distributions for δ_i at institution-level in purple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

Appendix Figure A2: Output Change based on Move Size



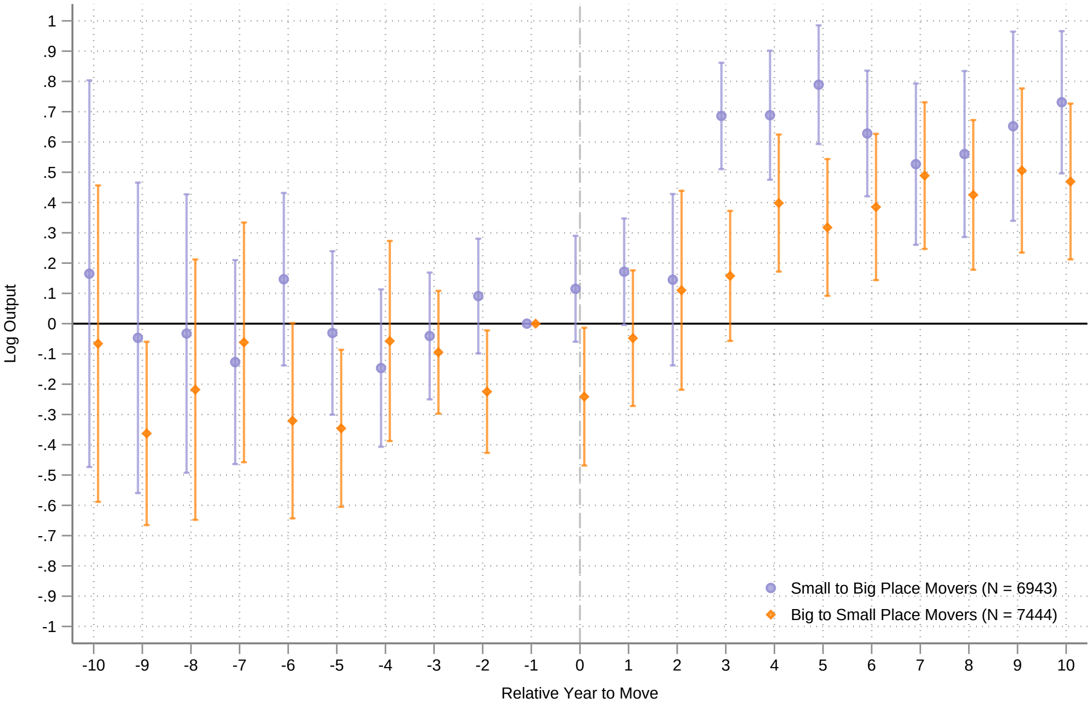
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers based on move size for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year 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: Output Change based on Age



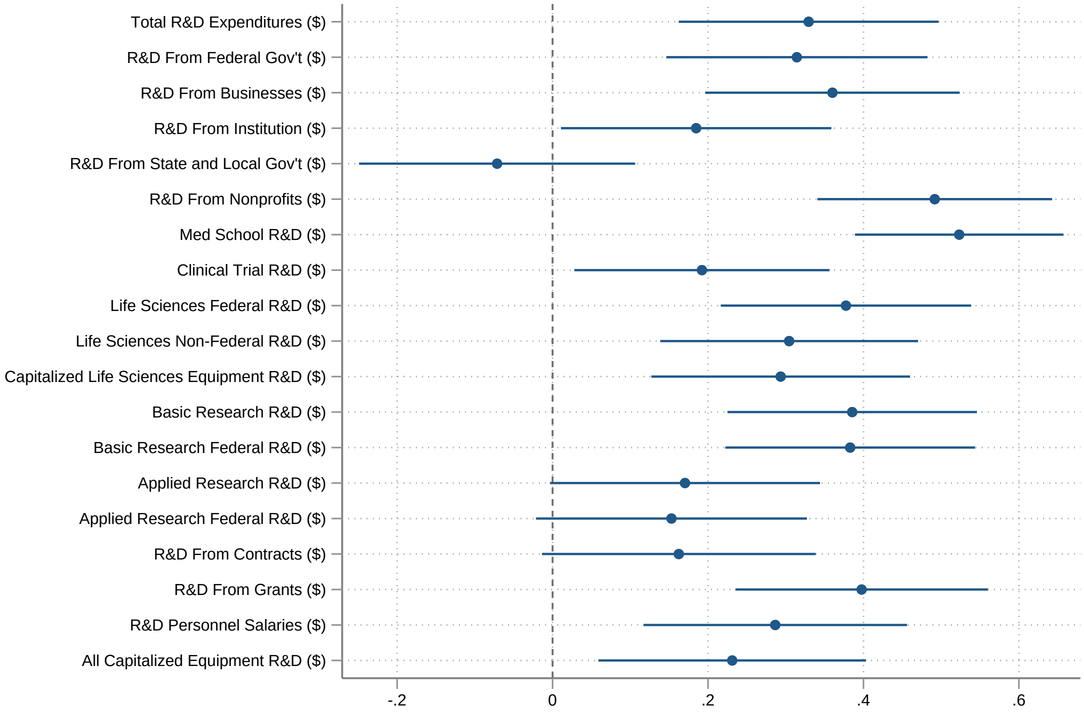
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers split by age for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, and relative-year 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: Output Change based on City Size



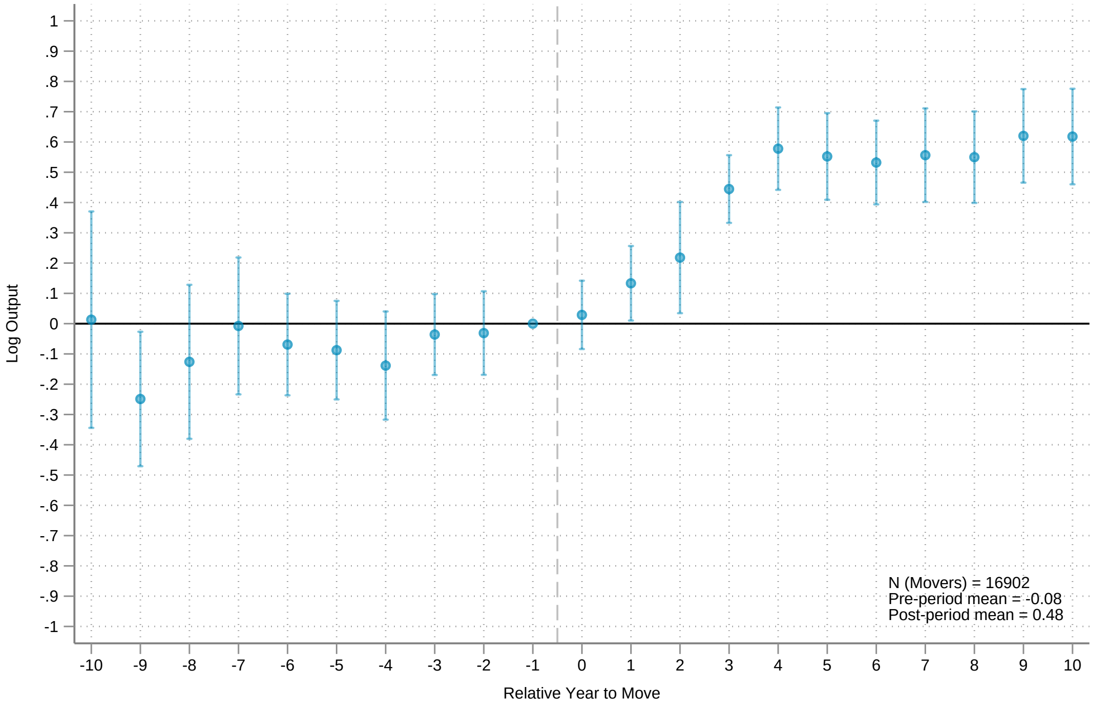
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers based on city size changes for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. In calculating city size, we ignore the researchers located at the mover's origin and destination location. 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



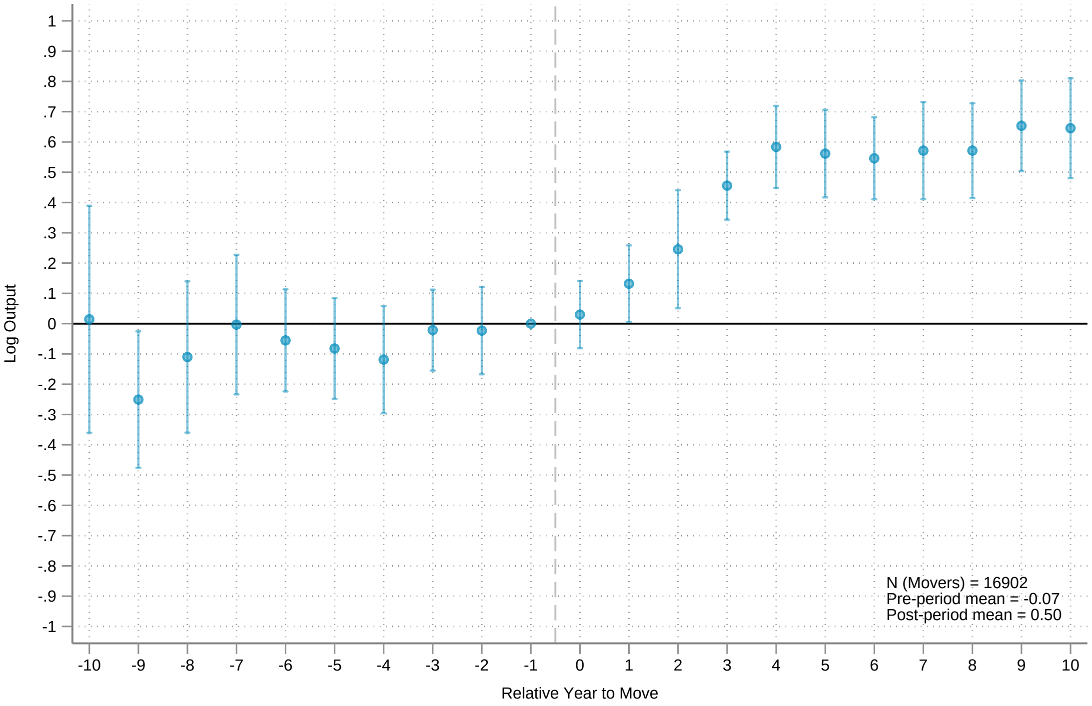
Notes: Figure shows results from regressing the estimated institution effects on various place-based factors from the NSF Higher Education Research and Development (HERD) Survey.

Appendix Figure A6: Specification including Field Fixed Effects



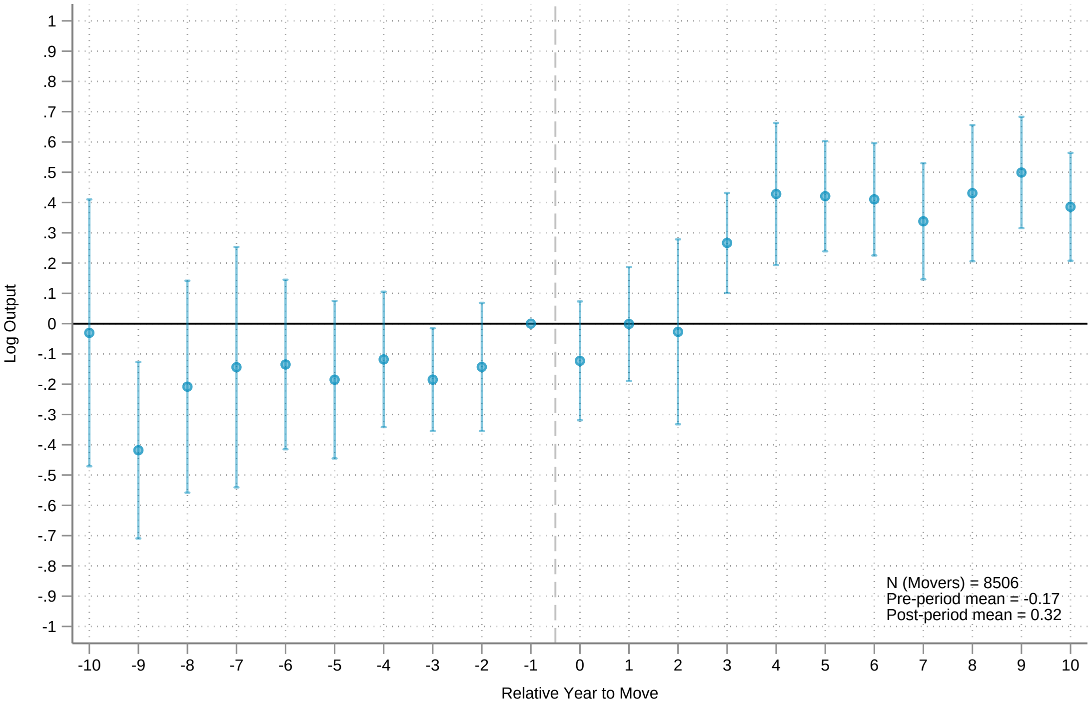
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 with the inclusion of subfield fixed effects for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, author, 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



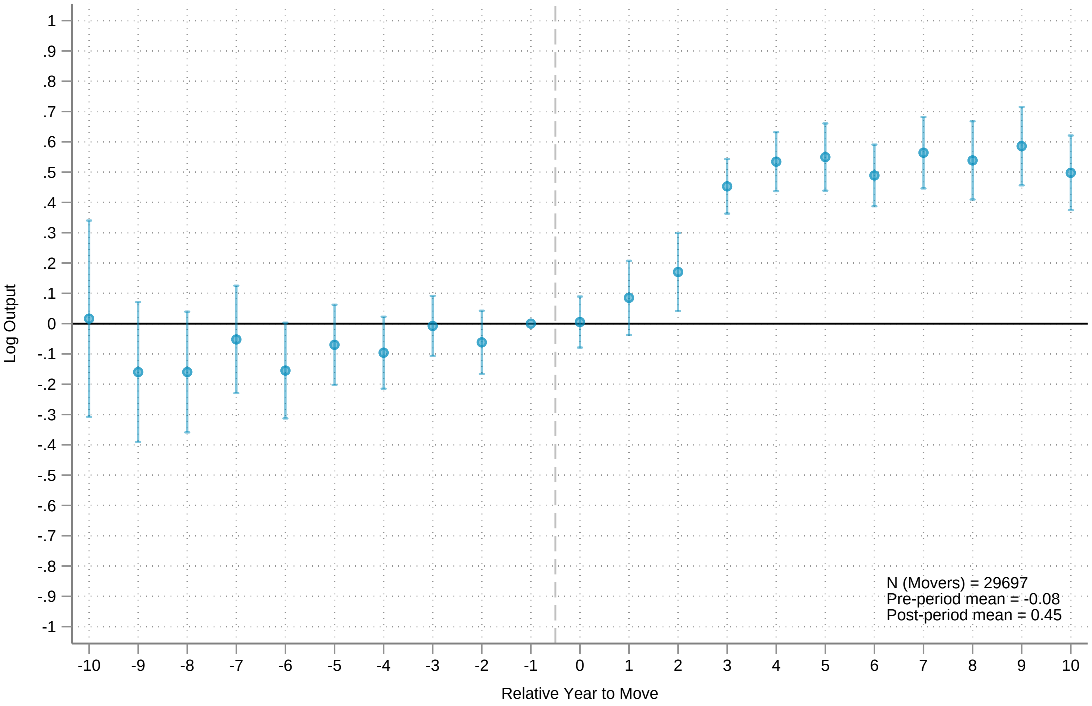
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 without the inclusion of relative year fixed effects for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year 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 2 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 2 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. Weighting Publications for Impact

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 output.

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 factors. 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.

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

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 output 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