Universities and the Production of Fundamental Knowledge in the Life-Sciences*

Amitabh Chandra† Connie Xu‡

October 3, 2024

Abstract

Fundamental knowledge in the life-sciences facilitates the discovery of medical innovations and subsequent knowledge, but is constrained by market-failures, limited public funding, and disproportionate production in a small set of institutions. To assess whether location causally increases the research productivity of scientists, we employ a scientist-mover design and find that $30-45\%$ of the geographic variation in fundamental research productivity is attributable to place-specific factors. We discuss the implications of these results for science policy.

*We thank George Daley and Amy Wagers for helpful comments.

†Harvard University and NBER, amitabh_chandra@harvard.edu

‡Harvard University, conniexu@g.harvard.edu

Fundamental discoveries in life-science have enabled new medicines and accelerated new knowledge that builds on these insights: the discovery of DNA, genes, mRNA, antibodies, and CRISPR has enabled new medicines and increased our cumulative understanding of physiology and disease (Zerhouni, 2005). This research is generally conducted in academic laboratories and naturally raises the question of whether some universities and locations are more productive than others. Quantifying the effect of institutions on the productivity of individual scientists is relevant for scientists' careers and the well-being of patients. Moreover, because fundamental research is likely to be underprovided by markets relative to the optimal social amount, understanding productivity differences across institutions becomes a key parameter for science policy, including how limited public and philanthropic subsidies should be allocated.

The location of where research is performed can increase the productivity of a scientist if it reduces the cost of inputs such as access to students, facilities, equipment, and mentors (Furman and Stern, 2011). Location also captures the proximity to other researchers and institutions that can facilitate learning, knowledge spillovers, and collaborations (Zucker et al., 1998). However, if more productive scientists prefer to live and work in places where they received their education, or where the weather is better (or worse), or places where solving dual-career challenges are easier, then it may appear that some locations are more productive in producing knowledge, but this be confounded by the locational preferences of scientists.

To understand whether the association between where a scientist works and their research productivity is causal, we implement a movers design by estimating event study specifications of an individual researcher's productivity on their "size of move", defined as the difference in productivity between the origin and destination location. If institutional factors explain 100% of a scientist's productivity, then moving to an institution that is 20% more productive ought to increase a scientist's productivity by the same amount; conversely, if productivity in producing fundamental knowledge is entirely a scientist-level phenomenon, then moves to institutions of higher or lower productivity should not affect an individual's research productivity. The movers design controls for individual and time fixed effects, which also reveal whether changes in an individual's productivity are predictive of where they move (which would not be desirable for identification of a place effect). This approach and the focus on the role of institution-level factors is closely connected to Lerner et al. (2024) who pioneer a scientist-mover design and estimate that university-level factors account for about a fifth of variation across universities in the extent to which patents cite knowledge produced in the life-sciences. As in Lerner et al. (2024) we are also interested in life-sciences research but focus on

the creation of knowledge regardless of whether it has commercial implications, for foundational discoveries drive discoveries in therapeutics over many decades. Moreover, as in Lerner et al. (2024) the precise source of institutional productivity remains shrouded– it may be better peers, students, resources, or governance. In the empirical work, we are able to separate the effect of institutional productivity from city-level productivity by examining the extent to which city-level productivity that ignores own-institutional productivity influences a scientist's productivity.

We define fundamental science research using a definition from the National Academy of Medicine, which encompasses traditional wet-lab research (such as cell division and protein degradation), and translational research, which bridges these insights with testing therapies in humans. The vast majority of this knowledge is published as scientific papers in academic journals. There are several reasons to believe that this research is underprovided relative to the social optimal: phenomena revealed by such research may include biological processes and objects, such as transcription and genes, that are not patentable because they are natural phenomena in which patents are impermissible (Kesselheim et al., 2013). Some discoveries may be removed from commercial applications, making them unable to satisfy the 'useful' criterion (also known as the Section 101 criteria) for obtaining a patent (Budish et al., 2015). Disclosing these discoveries in scientific journals also means that others may use the knowledge and combine it with other knowledge to make their own discovery; the combination of positive externalities and limited appropriability means that firms are unlikely to pursue fundamental science research. (Nelson, 1959; Akcigit et al., 2021; Azoulay and Li, 2022).

We measure research productivity as the frequency with which a scientist is listed as the first or last author in a paper, adjusting for citation counts, journal impact factors, and whether the research is cited in a patent. Our core sample consists of 563,754 papers from 15 journals from 1945 to 2023 that were cited 65,049,187 times, and show similar results in a sample of papers published in the leading three outlets— *Cell*, *Nature* and *Science* (henceforth, CNS) that were cited 31,450,159 times. We use information on the location of 556,821 first and last authors (213,690 US-based researchers) and their research productivity before and after moving from one institution to another.

We gather publication data from *OpenAlex*, a comprehensive open catalog of global research (Priem et al., 2022). These resources highlight another reason for studying knowledge creation in the life-sciences— 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 available. This would not be true for industries that rely on trade secrets over patents or where fundamental knowledge is not published in

journals.

These data reveal that the production of fundamental science research in the life-sciences is highly concentrated (Carlino and Kerr, 2015). At an aggregate level, the United States (US) dominates the field, producing more than half of the world's output. Notably, the gap in output between the US and its closest competitor in recent times, China, is more than five-fold. Three countries (US, China and the UK) account for 70 percent of the world's research output. Within the US, there is sharp geographic disparities in where productive researchers on average are located. California, Massachusetts, and New York disproportionately attract the most productive researchers, and within these states, the most productive researchers are located in the Boston, Bay Area, and New York metropolitan areas. We find that fundamental research from more productive cities also receive higher citations in patents, so this research is both of higher scientific quality (as measured by citations) and of higher commercial quality (as measured by paper-to-patent citations), which establishes the link between fundamental research and subsequent innovation in the life-sciences.

Our estimates suggest that at least 30-45% of the geographic variation in fundamental research productivity is accounted for by institutional-specific factors. This suggests that a scientist's location has a significant and independent influence on their research output, even after accounting for individual-level factors like experience and talent. Researchers who relocate to more productive environments, as measured by a positive move size, tend to experience a larger increase in their productivity, than the decline in productivity that is experienced when a researcher locates to a less productive institution. Disentangling the effect of institutional productivity further, we find that 20-25% of our estimated place effect is driven by productivity spillovers from institutional star researchers. In addition, researchers who moved before the age of 40 and those who moved to larger cities benefit more from relocation. This could be due to several factors, such as early career scientists being more adaptable in their research career and larger cities offering larger agglomeration externalities. Finally, we use alternative specifications that probe at the role of city-level productivity and find that research productivity spillovers are highly localized– city-level productivity that ignores own-institutional productivity has little influence on a scientist's productivity.

We correlate our place effects with various institution-level variables such as the amount of R&D spending. This correlational evidence reveals a positive association between place effects and factors like institutional R&D spending, the number of researchers receiving grants, and the presence of a supportive ecosystem for commercializing scientific research. In other words, institutions that invest heavily

in research and are located in areas with strong commercialization infrastructure exhibit larger place effects.

These estimates have implications for a number of allocations that pertain to science-policy. There is growing recognition of the importance of fostering talent for sustaining long-term innovation, and enhancing support for early-career scientists through targeted funding and mentorship programs (Azoulay et al., 2011, 2019, 2021). If institutions exert a causal effect on scientist's productivity, then such support may be most effective when used in the most productive clusters. Similarly, encouraging greater mobility of scientists towards more productive clusters, particularly younger ones through training grants at particular locations, would be a cost-effective way to increase their productivity and the return on public investment.

There is also an active debate on whether limited public funds should be distributed evenly across institutions, as would be the case if the key input for scientific discovery is money, or prioritized for institutions which are causally more productive, as would be the case if some institutions are able to do more than others with the same amount of funding. Our results suggest that the most productive institutions are causally able to generate more research, implying that efforts to distribute research funding without regard to the productivity of the location receiving it would be an inefficient allocation relative to the goal of producing the most research (but may achieve other social objectives). This insight would also justify colocating a number of public-private initiatives– such as Defense Advanced Research Projects Agency (DARPA), and Advanced Research Projects Agency- Health (ARPA-H)– near more productive clusters (Azoulay et al., 2019; D'Souza and Snyder, 2024).

Finally, our results suggest that disproportionate investments by venture capitalists in cities that produce a lot of fundamental science research (Chandra et al., 2022), may simply reflect better investment opportunities and more scientific human capital in these cities, as opposed to capital misallocation, where valuable investment opportunities are not capitalized. These examples underscore the potential of targeted investments and strategic science-policy to increase discovery in the life-sciences and ensuring public funding translates into substantial societal benefits.

Our results contribute to a larger research effort that measures various aspects of the production function that generates innovation. Our contribution focuses on one key determinant of fundamental science research– the location where it happens– and there is a rich literature noting other determinants (Iaria et al., 2018; Azoulay et al., 2019; Bloom et al., 2020; Myers, 2020; Akcigit et al., 2021; Arora et al., 2021; Hill and Stein, 2023, 2024; Acemoglu, 2023). This literature complements

other research that has focused on patenting, which is a downstream, but adjacent, consequence of such research. The research on patenting has uncovered the role of public-funding on subsequent patents, the role of university licensing offices on patenting, the role of parental income in encouraging children's patenting propensity, and the gender composition of scientific teams on patents (Azoulay et al., 2013, 2019; Myers and Lanahan, 2022; Gross and Sampat, 2023; Moretti et al., 2023; Jacob and Lefgren, 2011; Jensen and Thursby, 2001; Thursby et al., 2007; Lerner et al., 2024; Whittington, 2009; Koning et al., 2020; Bell et al., 2019; Williams, 2017; Bryan and Williams, 2021). If the same forces that influence patenting also influence fundamental science research– because they are often undertaken by the same people and are likely to be complementary activities– it would mean that they are even larger than previously estimated.

I. Data and Measurement

A. Defining Fundamental Science Research

The National Academy of Medicine (NAM), previously known as the Institute of Medicine, defines fundamental research in the life-sciences 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). The NAM definition includes two types of research: the discovery of biological and chemical phenomena that are not specifically focused on therapeutics or diseases. These include studies of anatomy, organisms, and other biological processes such as RNA, DNA, genes, cell signaling, and viral replication (this portion of the definition accords with the general notion of basic science research that is removed from therapeutics and diseases). The second component is translational science research that explores how these basic science insights can be used to produce new medicines.[1]

We operationalize the NAM definition of fundamental science research by querying Medical Subject Headings (MeSH) that are assigned to each publication in *PubMed*. MeSH are a hierarchically organized classification created by the NIH National Library of Medicine (NLM) for the purpose of indexing and cataloging biomedical articles in *PubMed*. We specifically query for the MeSH terms *Anatomy*,

[1]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 health-economics research (!). Some of these areas have public good attributes that are similar to the fundamental research definition used by NAM.

Chemicals and Drugs, Organisms, Phenomena and Processes, Diseases, Anesthesia and Analgesia, and *Therapeutics*. Remarkably, the main MeSH terms represented in our sample of fundamental science articles were receptors, DNA, RNA, neurons, and brain.

We include journal articles and clinical trials published between 1945[2] and 2023. This focus on journal articles excludes publications classified as comments, case reports, technical notes, letters, and reviews. We further adjust for research quality in a simple and transparent manner by limiting our query to publications from fifteen top life-sciences journals: *Cell*, *Nature*, *Science*, *Nature Cell Biology*, *Nature Genetics*, *Nature Medicine*, *Nature Biotechnology*, *Nature Chemical Biology*, *Nature Neuroscience*, *Neuron*, *Cell Stem Cell*, *PLOS One*, *Oncogene*, *Journal of Biological Chemistry* and *the FASEB Journal*. As a sensitivity analysis (Appendix Section D), we find qualitatively similar results when we subset to a limited set of high-quality papers published in the leading three outlets— *Cell*, *Nature*, and *Science* (henceforth, CNS). Our core sample consists of 563,754 publications, authored by 1,434,267 different scientists; the CNS sample includes 119,730 publications by 333,020 authors. As shown in Appendix Table A1, the share of fundamental science articles published in these journals ranges from 38% (for *Science*) to 99% (for *Oncogene*), with the remaining articles pertaining to the physical and social sciences.

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). *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*[3]. As of the latest data release, *OpenAlex* indexes over 250 million works by 95 million authors affiliated with 109,000 different institutions. These data are updated frequently, often every few hours through its API.

OpenAlex offers several advantages over comparable databases like *PubMed* or *Web of Science* for identifying the geography of research because of intentional

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

[3]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.

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

Finally, we use *OpenAlex* to impute the age of fundamental science researchers to understand how institutions affect life-cycle productivity. This is done by identifying the first time that a researcher appears in a paper indexed by *OpenAlex*. We assign a researcher's age to be 25 in the year that they first publish a paper[4].

B. Measuring Research Productivity

We move from publications to the research productivity of individual scientists by characterizing the "effective" research contribution of each author associated with a publication. We first divide a paper by the number of authors and then adjust for research impact by assigning paper-specific weights based on the latest 5-year impact factor of the journal from Clarivate (2023) and the average citation count per year since publication (the two weights mean that higher weights are assigned to papers in journals with higher 5-year impact factors, and also allows a given paper to be weighted more if it generated more cites). These weights are constructed in a manner that ensures our final measure of research productivity still sums up to the total number of papers in our sample. Appendix Section B explains in detail the math behind our productivity measure. In addition, the top panel of Appendix Table A2 presents various methods of weighting our productivity measure and shows that these are all tightly correlated at the country, city, and institution-level.

For our main analysis, we restrict our core sample to researchers listed either

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

as the first or last author on a publication; the first author typically represents the person who contributed the most to the work and writing, while the last author usually represents the principal investigator. This approach accounts for the varying tendencies to include additional coauthors across locations. Consequently, the total number of authors on a paper is at most two (Section E replicates our analysis for the full set of authors). Overall, we find that place-specific factors matter less for the sample of first and last authors than for the sample of all authors. This decrease in magnitude of around 10% is likely driven by the fact that the first and last authors are more likely to represent influential researchers in the field relative to the middle authors. Thus, the relative share of person-specific factors should be higher for the first and last authors.

C. Defining the Geography of Research

While authors can be affiliated with multiple institutions in a year, we are interested in the home institution where they physically carry out their research. We infer each author's home institution by gathering the full publication history of each author and leveraging publication metadata to identify the most consistent affiliation for an author in any given year. For authors with multiple affiliations in one year, our method starts by selecting the unique modal institution in a year and then applies rules that prioritize continuous affiliations and identify "sandwich" patterns[5].

Having assigned each author to a unique institution in a year, we are able to examine research productivity not only at the institutional level but also at broader levels including cities and countries. Our definition of cities aims to capture economic areas around the world. Thus, for areas within the US we use Metropolitan Statistical Areas (MSAs) as our definition of cities, and combine major MSAs to integrated labor-markets. In the rest of the paper, we refer to these classifications as "cities". For example, we define the Bay Area, CA city as the combination of San Francisco-Oakland-Hayward and San Jose-Sunnyvale-Santa

[5]The goal of our algorithm is to reduce multiple affiliations to a single one. If a unique modal affiliation exists for an author in a given year, we choose that one. We break ties by first identifying any "sandwich" patterns across consecutive years. For example, if institution A appears in 2008 and 2010 but not in 2009, we assign institution A as the home affiliation for all those years. If multiple affiliations remain, we prioritize continuous affiliation by assigning the current year's primary institution based on affiliations from the prior year and then from the subsequent year. Specifically, if in any given year an author is affiliated with both institutions A and B, we first assign the home institution as A if A appeared in the prior year. If there is no affiliation from the previous year, we assign the home institution as A if A appears in the subsequent year. Finally, if there are any remaining multiple affiliations after all these steps, we randomly choose one and assign it as the home institution. This last step affects less than 1% of author-years.

Clara (similarly, we define the Research Triangle Park, NC as the combination of Durham-Chapel Hill and Raleigh-Cary). Table 1 lists the top 20 US cities by number of fundamental life scientists since 2015.

Finally, to account for differential changes in productivity across life-sciences subfields stemming from breakthrough discoveries (for example, relative to the fields of structural or cancer biology, there could be larger advances in the field of genetics because of the mapping of the human genome), we categorize each author in each year into one of five life-sciences research field. We use paper titles, MeSH terms, abstracts, and OpenAlex concept terms of each author's papers as inputs for a k-means textual clustering algorithm. By running this algorithm to each author-year rather than to each author, we allow for changes in scientists' research focus over time. The five life-sciences fields represent (1) molecular and cellular biology, (2) cancer biology, (3) genetics and genomics, (4) neuroscience and physiology, and (5) structural biology. Appendix Section C describes the tokenized features that define each research field.

To illustrate our approach, consider the following two papers that contributed to the development of the CRISPR-Cas9 genome editing technique, a groundbreaking discovery that earned Emmanuelle Charpentier and Jennifer A. Doudna the 2020 Nobel Prize in Chemistry. The first paper, "*A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity*", was published in *Science* in 2012 and has been cited 13,832 times. To construct our measure of research productivity, we first allocate an effective contribution of 0.5 to both Martin Jinek (the first author) who was at UC Berkeley (in the Bay Area) and Emmanuelle Charpentier (the last author) who was at Umeå University. In 2012, both Martin Jinek and Emmanuelle Charpentier were classified as working in the subfield of molecular and structural biology[6]. Using the latest 5-year impact factor of *Science* (54.5) and the average citation count per year since publication ($13832/12 \approx 1152$), we reweight our measure of productivity to represent the relative influence of this paper in comparison to other works in our sample. Consequently, both Martin Jinek and Emmanuelle Charpentier (and their affiliated institutions) receive a recalibrated contribution of 57.25 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 (the first author) and Jennifer A. Doudna (the last author), who were both at UC Berkeley. In 2014, both Samuel Sternberg and Jennifer A. Doudna were classified

[6]Based on our subfield classification, Martin Jinek starts his research in neuroscience and physiology before working a bit in genetics and genomics and focusing on molecular and cellular biology. Emmanuelle Charpentier starts her career working in neuroscience and physiology before expanding her research to molecular and cellular biology and genetics and genomics

as working in the subfield of molecular and structural biology[7]. 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 6.6 articles for this publication. To put these research productivity magnitudes in context, Table 2 shows that the average author in an elite set of scientific papers has an annual productivity of 0.72 articles and a lifetime productivity of 1.64 articles (measured up until the end of our time frame in 2023).

II. Geographic Variation in Fundamental Science

A. Researcher Productivity

We first motivate our choice to focus our main analysis on US scientists by documenting the large geographic disparity in the global production of fundamental science research. Figure 1 shows the aggregate share of output produced by each country since 1945[8]. 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.[9] Thus, given the US' large role in fundamental science research, we do not view our restriction to US-based scientists as a major limitation (in our original sample of 563,754 papers, 272,604 were authored by 213,690 US-based researchers).

To explore how researcher productivity varies across different geographies in our

[7]Throughout his career, Samuel Sternberg shifts from molecular and cellular biology to genetics and genomics. Jennifer Doudna begins her career in neuroscience and physiology, before moving to molecular and cellular biology, and finally working mainly in genetics and genomics.

[8]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.

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

sample, we analyze the cross-sectional variation in our productivity measure. Figure 2 presents a city-level map of the average productivity of researchers, revealing substantial differences in the locations where highly productive scientists are concentrated. We see that the most productive researchers are predominantly located along the Northeast and West Coast.[10]

Table 3 presents the top 20 states and cities by average researcher productivity. On average, researchers in California, Massachusetts, and New York are more than twice as productive as those in other states. Notably, the metropolitan areas of Boston, the Bay Area, and New York City emerge as key centers of scientific productivity within these states. Table 4 further highlights the top five institutions in the ten most productive cities. While each city has a few dominant institutions, there is also a balanced presence of highly productive researchers in both universities and private research organizations. Appendix Table A4 provides the list of the top 50 institutions by average researcher productivity.

B. Fundamental Science and Papers-in-Patents

So far, we have have not established whether fundamental science research has direct implications for inventive activity, as opposed to representing knowledge that is fundamental but that is quite distant from inventions. Building on work by (Bryan et al., 2020), (Marx and Fuegi, 2020), we measure one form of commercial spillovers from fundamental research by looking at the propensity of published research to be cited in patents. This measure of "papers in patents" productivity is similar to our measure of research productivity except the paper-specific weights adjust for the number of times the paper has been cited within the body text of patents. 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. Appendix Table A5 shows the top 20 states, cities, and institutions based on average researcher papers in patents citation productivity.

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

[10]When calculating average productivity across author-years within a given geography, we weight each author-year by the number of authors in that region. This adjustment accounts for the fact that larger cities often have more outliers, with some authors contributing fewer publications, which skews the distribution.

by patent lawyers (Bryan et al., 2020)).

We plot the city-level relationship between total research productivity and total "papers in patents" productivity in Figure 3. The slope is effectively 1, which allows us to infer that clusters 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. Appendix Table A2 and A3 show the sensitivity of our results to using front page vs. patent body citations – regardless of the chosen measure, there is a strong correlation between producing fundamental science research and having it cited in patents at the country, city, and institution-level.

III. Empirical Strategy

A. Econometric Model

We present an empirical framework using a "movers" analysis that closely follows that of (Finkelstein et al., 2016) and decomposes the observed variation in fundamental research productivity into place- and person-based components. Place-based components capture local institutional factors such as firm-level R&D expenditures and person-based components capture individual characteristics such as educational background and motivation. Our model of research production is similar in spirit to the model in (Lerner et al., 2024)[11].

In a simplified manner, we can define a scientist's log productivity, y_{ijt}, for scientist i at institution j in year t as[12]

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

where α_i is an individual fixed effect, γ_j is an institution fixed effect, and τ_t is a calendar year fixed effect.

To decompose the variation in average log productivity across institutions into place- and person-based components, we define $\overline{y_{jt}}$ as the average of y_{ijft} across

[11]The paper by (Lerner et al., 2024) is most closely related to ours. In their setting, they are interested in unpacking the heterogeneity in university commercialization. We focus on unpacking the upstream heterogeneity in institutional fundamental research production

[12]This model can be expanded to capture observable researcher characteristics by including $x_{it}\beta$. In this more extensive model, the researcher-level averages would be for $\alpha_i + x_{it}\beta$ rather than just α_i. For notational convenience, we describe the motivating theory with the simplified model instead. Indeed, when we take the model to the data, we incorporate subfield fixed effects which would fall under $x_{it}\beta$

researchers working at institution j in year t, and $\overline{y_j}$ as the average of $\overline{y_{jt}}$ across time. Similarly, $\overline{\alpha_{jt}}$ and $\overline{\alpha_j}$ represent researcher-level averages. The difference in average productivity between any two institutions j and j' is given by:

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

We define the share of the difference attributable to the institution as:

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

and the share attributable to researchers as:

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

The key parameters of interest in this model are the institution fixed effects, γ_j which are only identified if our sample includes scientist-movers who relocate between institutions; without movers, we cannot separately identify the impact of place from the unobserved heterogeneity across individuals.

In order to interpret these place effects causally, we require that changes in unobserved person-based determinants of research productivity are not correlated with the difference in average productivity between destination and origin institutions chosen by the researcher. This assumption would be violated if "rising stars" systematically relocate from low- to high-productivity institutions. Violations of this kind will appear in the pre-trends when we estimate our model empirically with an event study specification. The second assumption in our model is that α_i and γ_j are additvely separable in log productivity, which implies that the level of place and person characteristics affect productivity multiplicatively. In other words, moving to a place that is more conducive to productivity matters more for those researchers starting with a lower level of productivity. Finally, our approach assumes that the γ_j estimated for movers is the same as that for non-movers. If this assumption is violated, then our estimates would only apply to the set of researchers who relocate rather than the entire population of researchers.

B. US-based Scientist Movers

Using our unique panel of author-year affiliations, we define movers as researchers who are affiliated with different institutions in two 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 time of move. We find that movers begin accumulating coauthors from their future destination institution approximately three years before the first publication affiliated with that institution. Therefore, we define the move year as three years prior to the first observed publication with the destination institution. Figure 4 illustrates this approach, showing that the share of coauthors from the destination institution only increases after our newly defined move year.

Among the set of US-based researchers, we are able to identify 32,457 scientist-movers. On average, these movers relocate 1.4 times in their lifetime. Due to the role that the accuracy of OpenAlex's author disambiguation algorithm and our affiliation imputation method play in identifying movers, our empirical analysis focuses on a set of 19,325 researchers who only move once. Table 2 provides summary statistics on movers and non-movers in our sample. On average, movers tend to come from slightly smaller clusters but produce more output on average than non-movers. Additionally, one-time movers represent about 60% of the mover sample and are generally representative of the broader mover population.

C. Event Study Estimation

We implement an event study approach to visually observe the dynamic evolution of place effects on scientist's productivity, and assess the robustness of our assumptions through the pre-trends. To implement the event study approach, we need to scale log productivity, y, so that the direction and magnitude of the jump on the move are informative regardless of the origin and destination. 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. Thus, we define the move size, δ_i, for a mover i who goes from institution $o(i)$ to institution $d(i)$ as the difference in the average log productivity between the other researchers in the mover's destination and origin:

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

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

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

This expression implies that y_{it}^{scaled} will equal 0 if the researcher's productivity exactly equals the average productivity at their origin institution and exactly equals 1 if the researcher's productivity exactly equals the average productivity at their destination institution. Thus, plotting the averages of y_{it}^{scaled} by relative year, $\theta_{r(i,t)}$ would produce an event study figure where the size of the jump at the time of the move corresponds to the average value of S_{place} across all movers. Here, if mover i relocates in year t_i^*, the relative year is defined as $r(i,t) = t - t_i^*$. Thus, $r(i,t) = -1$ represents the last year in the origin institution and $r(i,t) = 0$ represents the first year in the destination location.

In our data, we see from Figure A1 that δ_i is often close to 0, which will make a regression with y_{it}^{scaled} on $\theta_{r(i,t)}$ behave poorly. Thus, we rearrange our equation to avoid dividing by δ_i as follows:

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

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

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

where τ_{ft} and $\rho_{r(i,t)}$ are field by year and relative year fixed-effects, respectively. We combine $\overline{y_{o(i)}}$ and α_i into a single individual-level fixed effect, $\widetilde{\alpha}_i$. By including relative year fixed effects in our specification, we also allow for a common trend in productivity for movers either pre- or post-move as long as this trend is common across all movers regardless of the size and direction of their move. This is important if we think that there are specific reasons correlated with one's productivity that lead a researcher to relocate when they do. Finally, we include field-by-year fixed effects to estimate our results within a researcher's field.

The parameters of interest from Equation 2 are the relative year coefficients, $\theta_{r(i,t)}$, that are normalized to zero in the year before the move $r(i,t) = -1$. These coefficients measure changes in productivity (expressed as a percent) in the years around the move scaled by the size and direction of the move, δ_i. Moreover, the relative quality of one's destination should not be predictive of productivity before a move but should be predictive of productivity after a move, and $\theta_{r(i,t)}$ should be statistically indistinguishable from zero preceding a move and non-zero following a move. The magnitude of the jump in $\theta_{r(i,t)}$ after a move captures on average how much place-specific factors influence productivity. $\theta_{r(i,t)} = 1$ would imply that the mover's productivity has fully converged to that in the destination location, and an estimate of $\theta_{r(i,t)} = 0$ would imply no convergence. We report robust standard errors clustered at the institution level.

IV. Estimates of Institution and Researcher Effects

A. Baseline results

We motivate the event study estimates from the movers analysis by first showing the identifying variation behind our movers specification. Appendix Figure A1 presents the distribution of $\hat{\delta}_i$, the difference in average log productivity in a mover's destination and origin institution, which is the key dependent variable. A value of 0 means that the destination institution has the same average research productivity as the origin institution. This distribution is centered around -0.11 and is relatively symmetric, indicating an approximately equal number of moves from low-to-high productivity institutions ($\hat{\delta}_i > 0$) as there are moves from high-to-low productivity institutions ($\hat{\delta}_i < 0$). The standard deviation is 1.33, with a significant number of moves with a magnitude greater than 5[13].

As an initial exploration of the impact of moving to a new institution on a scientist's productivity, we plot the post-move change in scientist productivity (on the y-axis) against the size of the move, which is the same $\hat{\delta}_i$ (on the x-axis). We define a scientist's change in productivity as their average post-move productivity minus their average pre-move productivity. The slope of this graph indicates the extent to which moving to an institution with higher average productivity influences a scientist's productivity. If geographic variation in research productivity was entirely due to the impact of place, then the slope would 1. Conversely, if individual researcher factors fully explained the variation, the slope would be 0. Figure 5 plots this relationship and depicts a slope of around 0.45. This simple analysis suggests that, on average, nearly half (46%) of the variation in scientist productivity can be attributed to institutional factors, while the remaining slightly more than half (54%) 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 the additive separability of α_i and γ_j in equation 1.

We report event study results from equation 2 in Figure 6a for the 1945-2023 time period. The figure plots the estimates of $\theta_{r(i,t)}$ and their 95% confidence intervals, after normalizing $r(i,t) = -1$ to be zero. The standard errors are clustered at the institution level to account for intra-institutional correlation in research productivity among researchers.

[13]Since our research productivity captures "effective" paper contributions, a measure of 5 means that the researchers in the destination institution produce 5 more papers on average than in the mover's origin institution

The figure shows that there is no significant trend in $\theta_{r(i,t)}$ in the decade before relocation. This supports our identifying assumption that unobserved person-based determinants of research productivity are not correlated with the productivity difference between the destination and origin institutions chosen by the researcher. There is no significant effect of place until three years after the move, for which there is then a discontinuous jump to about 0.3. This estimate remains relatively stable in the following year, with an increase to around 0.45– together, these findings suggest that the average institutional share of the variation in research productivity overtime is approximately 30-45%. This magnitude is slightly less than the static measure we see from Figure 5 because we're observing the dynamic change in place-specific factors and we're controlling for observable researcher and field characteristics.

To examine whether the results are different for more recent years, particularly in the internet era, which may have made it easier to collaborate with scientists at different institutions, Figure 6b focuses on scientists who moved in more recent years, from 1995 to 2023. Despite the digital age making collaboration easier, the role of institutions in explaining scientists' research productivity remains similar in magnitude. This consistency suggests that even in the era of advanced digital communication, the institutional environment continues to play a crucial role in influencing research productivity (estimates for other time periods are shown in Figure A2).

B. Extensions

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

The first mechanism we explore is based on the direction of the move. In our setting, a positive move ($\hat{\delta}_i > 0$) is one in which a researcher moves to an institution with a higher research productivity, whereas a negative move ($\hat{\delta}_i < 0$) is one in which a researcher moves to a lower productivity institution. Higher productivity institutions likely have more resources and better help facility scientists to work at the frontier of the field with more advanced research techniques. Assuming that researchers incorporate these benefits into their own research when moving to higher productivity places but don't unlearn these skills when moving to lower productivity places, then we would expect to see larger place effects for positive moves. We illustrate the estimated place effects by move size in Figure A3. These plots confirm our theory as we see a stark difference in place effects between

scientists moving to higher- versus lower-productivity institutions. The estimated place effect for positive moves is significantly larger and goes from 0.4 to 0.6, whereas the place effect for negative moves is closer to 0.3.

Given the complementary strand of research on the role of "star" researchers in driving innovation, we also examine the influence of stars in explaining our place effect estimates (Azoulay et al., 2019, 2011). To do this, we define "stars" as researchers in the top 95th percentile of productivity and modify our definition of δ_i to represent the difference between the average productivity of stars in the destination and origin institutions. Figure A4 illustrates the distribution of this revised move size, while Figure A5 presents the estimated place effect using this new definition. The results indicate that stars significantly contribute to the productivity of movers, with the estimated place effect being around 0.2. In other words, stars are driving around 40-50% of the place effect we see in our baseline results.

The next dimension we explore is the researcher's age. This is motivated by previous research examining the changing life-cycle productivity of a researcher (Dietz and Bozeman, 2005; Jones, 2010, 2009; Levin and Stephan, 1991; Myers, 2020). Consistent with this literature, we explore whether movers who move before the age of 40 are more likely to be influenced by their environment compared to researchers who relocate later in their career (which implies larger place effects for researchers moving at an earlier age). Figure A6 shows the estimates of our event study for movers according to whether they are above or below the age of 40 at the time of move. The results show economically large place-based effects for researchers no matter when they move, with place effects influencing those who move before the age of 40 slightly more.

Finally, we explore the role of agglomeration effects by characterizing moves based on changes in the cluster size of the destination and origin location. A large literature has documented a positive correlation between cluster size and innovation (Carlino and Kerr, 2015; Kerr and Robert-Nicoud, 2020; Moretti, 2021). In our context, we expect the place effects to be larger for researchers moving from smaller to large clusters. To capture meaningful changes in cluster size, we define cluster size at the city-level rather than at the institution-level[14]. For each author, their cluster size in a given year is the number of other researchers working in the same city (excluding their coauthors). With this definition, we define movers moving from small to larger clusters if their destination location has more researchers than their origin location. Figure A7 plots the estimates of our event study for researchers moving from small to large clusters and from large to small

[14]While we can define cluster size at the institutional level, we encounter the issue of small sample sizes at some institutions, leading to noisier measures.

clusters. These results show that the place effects are around 0.5 for those moving to larger clusters and around 0.4 for those moving to smaller clusters. This implies that increases in cluster size are more influential to a researcher's productivity and is consistent with the agglomerations literature.

C. Robustness

We assess the sensitivity of our results by constructing the move size at different levels of aggregation. Rather than defining δ_i at the institution-level, we define an alternative move size at the city-level. This approach accounts for the broader influence of the city, including researchers at other institutions, on scientist productivity. Figure A8 illustrates the distribution of this city-level move size, while Figure A9 shows the estimated place effects using this specification. The city-level estimates are qualitatively similar to, but slightly larger than, the institution-level estimates, indicating that most of the productivity diffusion in a place is concentrated within the author's own institution.

To further test this hypothesis, we modify our city-level move size definition to exclude a mover's own institutional productivity in both the origin and destination city. This alternative measure captures the impact of productivity changes driven by researchers outside of the mover's institutions. Figure A10 presents the distribution of this move size and Figure A11 displays the estimated place effects. The results show that the estimated place effects remain consistently around zero, supporting our theory that the impact of place on a researcher's productivity is primarily localized to their institution, with minimal influence from surrounding researchers in the same city.

D. Correlates of Place Effects

Having documented the important role of place-based factors in shaping researcher productivity, we take a first step to unpack what makes certain places better in facilitating productive research by correlating our estimated institution place effects, $\hat{\gamma}_j$, with various institution characteristics. We recover institution-level place effects by estimating a two-way fixed effects regression with researcher and institution fixed effects following equation 1.

We gather institution-level characteristics from a couple of data sources. The National Science Foundation has conducted an annual Higher Education Research and Development (HERD) survey since 1972. This survey collects data on research

and development expenditures at US colleges and universities, providing insights into funding sources, research activities, and spending trends across various fields of study. Due to a survey change that expanded the scope of questions asked in 2010, we only use data collected after 2010. We focus on the different sources of where R&D spending comes from as well as how much the institution spends on basic science (for all fields) and on life-science R&D. Using principal component analysis, we also construct an R&D index based on the set of chosen HERD variables. The place effect correlates shown for HERD institutional characteristics are restricted to the sample of universities we can match in our data (N=102). We also aggregate our "papers in patents" measure to the institution level and use this as a measure of the commercialization environment.

We transform each feature into a z-score and estimate a bivariate OLS regression, where the dependent variable is our estimated institution fixed effect, and the independent variable is a single institution feature. Figure A12 shows the results of these bivariate regressions. Although these results are only correlational, they suggest that institutions that spend more on R&D, have more researchers receiving grant money, and have a more supportive ecosystem for commercialization are associated with larger institution place effects.

V. Conclusion

Our study sheds light on the critical role of place in influencing the productivity of fundamental science research. We estimate that place-specific factors account for 30-45% of the observed variation in productivity.

Our analysis raises several questions for future research. One area of interest is the factors that influence the mobility of scientists. Previous studies have pointed to the importance of funding availability, personal circumstances, and institutional support systems. Another important topic is the role of institutional characteristics in fostering high research productivity. Although our results emphasize the importance of geographic location, they also suggest that institutional policies and culture, such as support for early-career researchers and high-risk, high-reward projects, play a crucial role. Exploring how different institutions structure their research environments and the impact of these structures on productivity can inform best practices for maximizing the effectiveness of research funding.

In conclusion, our findings underscore the significant role of institutional factors in shaping the landscape of fundamental life-sciences research. By strategically investing in research clusters and fostering supportive institutional environments,

policymakers can enhance the overall impact of scientific research.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chandra, A., C. Foroughi, and L. Mostrom (2022). Venture Capital– Led Entrepreneurship in Health Care. In *The Role of Innovation and Entrepreneurship in Economic Growth*. The University of Chicago Press.

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.

D'Souza, A. and C. M. Snyder (2024, May). Can Operation Warp Speed Serve as a Model for Accelerating Innovations Beyond Covid Vaccines? In *Entrepreneurship and Innovation Policy and the Economy, volume 4*. University of Chicago Press.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 1: Largest Fundamental Science Research Clusters by Number of Scientists (2015-2023)

	MSA	Avg. Cluster Size
1	New York-Newark-Jersey City, NY-NJ-PA	24574.62
2	Boston-Cambridge-Newton, MA-NH	23470.67
3	Washington-Arlington-Alexandria, DC-VA-MD-WV	20655.14
4	Bay Area, CA	17407.32
5	Los Angeles-Long Beach-Anaheim, CA	10457.83
6	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	9180.19
7	Chicago-Naperville-Elgin, IL-IN-WI	8683.64
8	Research Triangle Park, NC	7780.33
9	San Diego-Carlsbad, CA	7765.88
10	Houston-The Woodlands-Sugar Land, TX	7753.82
11	Baltimore-Columbia-Towson, MD	7425.37
12	Seattle-Tacoma-Bellevue, WA	6832.95
13	Atlanta-Sandy Springs-Roswell, GA	6077.65
14	Ann Arbor, MI	4288.82
15	Pittsburgh, PA	3981.57
16	St. Louis, MO-IL	3913.98
17	New Haven-Milford, CT	3777.11
18	Madison, WI	3304.92
19	Minneapolis-St. Paul-Bloomington, MN-WI	3244.96
20	Ithaca, NY	3225.31

Notes: Table shows the top 20 MSAs by average cluster size across 2015-2023. Cluster size is calculated as the number of (first & last) authors contributing to fundamental science research in a MSA-year. Even if authors do not publish fundamental science research every year in our set of top 15 life-sciences journals, we are able to impute where they are living in a given year based on each author's full set of publication history across OpenAlex.

Table 2: Summary Statistics

	Mean	Std. Dev.
A. Individual-Level		
Everyone (N = 183,387)		
Years Actively Publishing	2.04	2.67
Average Team Size	1.98	0.13
Annual Cluster Size	7180.17	7177.11
Annual Productivity	0.72	3.79
Lifetime Productivity	1.64	9.78
Number of Moves	0.25	0.64
Non-Movers (N = 150,930)		
Years Actively Publishing	1.42	1.45
Average Team Size	1.98	0.13
Annual Cluster Size	7201.11	7427.47
Annual Productivity	0.68	3.94
Lifetime Productivity	0.99	5.78
Movers (N = 32,457)		
Years Actively Publishing	4.90	4.52
Average Team Size	1.96	0.12
Annual Cluster Size	7082.75	5873.48
Annual Productivity	0.87	2.96
Lifetime Productivity	4.67	19.33
One-time Movers (N = 19,325)		
Years Actively Publishing	3.85	3.26
Average Team Size	1.97	0.12
Annual Cluster Size	7108.23	5997.96
Annual Productivity	0.81	2.63
Lifetime Productivity	3.33	14.68
B. City-Level		
Number of Institutions	84.91	85.67
Annual Cluster Size	7199.39	7656.73
Annual Productivity	0.95	4.24
Lifetime Productivity	123.93	910.54

Notes: Table shows summary statistics for our sample. Panel A presents the individual-level statistics and Panel B presents the city-level statistics.

Table 3: Places with Most Productive Researchers on Average

Rank	Place	Average Author Productivity
A. State		
1	California	2.01
2	Massachusetts	1.84
3	New York	1.20
4	Maryland	0.67
5	Texas	0.46
6	Pennsylvania	0.35
7	Washington	0.26
8	Connecticut	0.23
9	Illinois	0.23
10	North Carolina	0.21
11	Missouri	0.19
12	Michigan	0.17
13	New Jersey	0.16
14	Ohio	0.11
15	Tennessee	0.11
16	Georgia	0.11
17	Virginia	0.09
18	Florida	0.08
19	Minnesota	0.08
20	Wisconsin	0.08
B. City		
1	Boston-Cambridge-Newton, MA-NH	3.01
2	Bay Area, CA	1.94
3	New York-Newark-Jersey City, NY-NJ-PA	1.82
4	Washington-Arlington-Alexandria, DC-VA-MD-WV	0.86
5	San Diego-Carlsbad, CA	0.78
6	Los Angeles-Long Beach-Anaheim, CA	0.59
7	Baltimore-Columbia-Towson, MD	0.48
8	Seattle-Tacoma-Bellevue, WA	0.43
9	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.41
10	Houston-The Woodlands-Sugar Land, TX	0.40
11	New Haven-Milford, CT	0.36
12	Research Triangle Park, NC	0.35
13	Chicago-Naperville-Elgin, IL-IN-WI	0.33
14	St. Louis, MO-IL	0.28
15	Dallas-Fort Worth-Arlington, TX	0.26
16	Ithaca, NY	0.23
17	Ann Arbor, MI	0.21
18	Pittsburgh, PA	0.15
19	Trenton, NJ	0.15
20	Atlanta-Sandy Springs-Roswell, GA	0.14

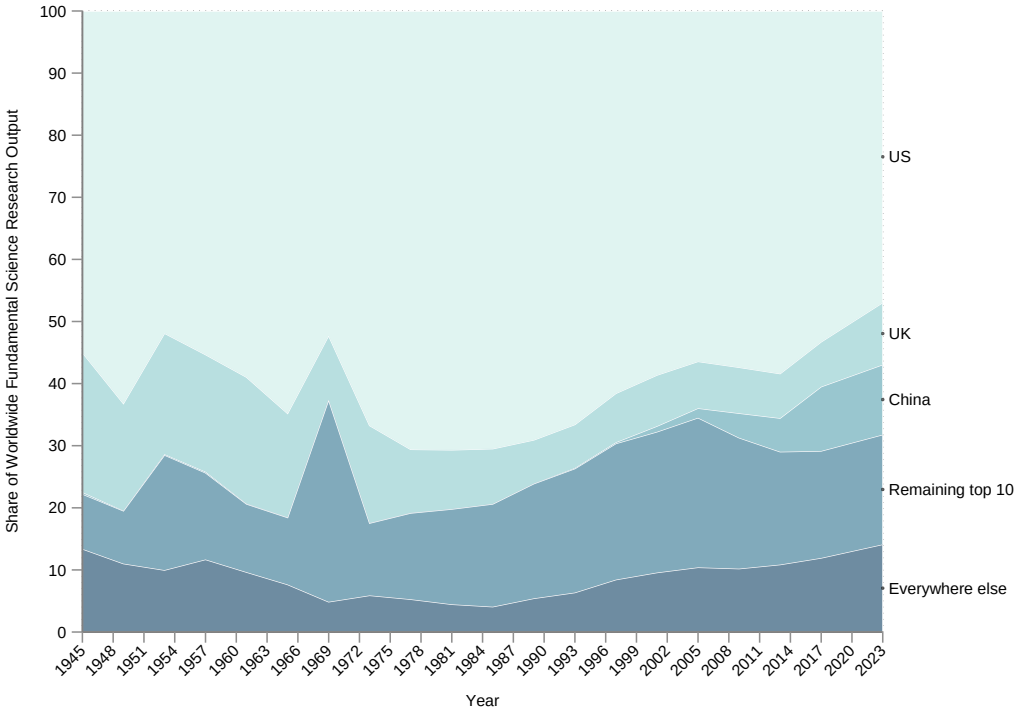
Notes: Table shows the top 20 states and cities by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the total number of researchers in each location. Journals included are: *Cell, Nature, Science, Nature Cell Biology,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

Table 4: Leading Institutions in the Ten Most Productive Cities by Average Author Productivity

Rank	Institution	Average Author Productivity
I. Boston-Cambridge-Newton, MA-NH		
1.	Harvard University	4.5604
2.	Massachusetts Institute of Technology	2.8697
3.	Mass General Brigham	1.9362
4.	Dana-Farber Cancer Institute	0.8869
5.	Boston Children's Hospital	0.4071
II. Bay Area, CA		
1.	Stanford University	3.0957
2.	University of California, San Francisco	1.9653
3.	University of California, Berkeley	1.1521
4.	Google (United States)	0.1753
5.	Lawrence Berkeley National Laboratory	0.1722
III. New York-Newark-Jersey City, NY-NJ-PA		
1.	Columbia University	1.3856
2.	Memorial Sloan Kettering Cancer Center	0.9675
3.	Rockefeller University	0.8372
4.	New York University	0.8236
5.	Icahn School of Medicine at Mount Sinai	0.6189
IV. Washington-Arlington-Alexandria, DC-VA-MD-WV		
1.	National Institutes of Health	2.0850
2.	University of Maryland, College Park	0.2093
3.	Carnegie Institution for Science	0.0830
4.	Georgetown University	0.0828
5.	Molecular Oncology (United States)	0.0765
V. San Diego-Carlsbad, CA		
1.	University of California, San Diego	1.3885
2.	Scripps Research Institute	0.5718
3.	Salk Institute for Biological Studies	0.5527
4.	La Jolla Institute For Allergy & Immunology	0.1332
5.	Sanford Burnham Prebys Medical Discovery Institute	0.1136
VI. Los Angeles-Long Beach-Anaheim, CA		
1.	University of California, Los Angeles	0.8181
2.	California Institute of Technology	0.6042
3.	University of Southern California	0.3373
4.	University of California, Irvine	0.2102
5.	Cedars-Sinai Medical Center	0.1292
VII. Baltimore-Columbia-Towson, MD		
1.	Johns Hopkins University	1.6879
2.	University of Maryland, Baltimore	0.0985
3.	Department of Embryology	0.0162
4.	Lieber Institute for Brain Development	0.0151
5.	Osiris Therapeutics (United States)	0.0117
VIII. Seattle-Tacoma-Bellevue, WA		
1.	University of Washington	1.0220
2.	Fred Hutch Cancer Center	0.2523
3.	Institute for Systems Biology	0.0867
4.	Allen Institute for Brain Science	0.0732
5.	Allen Institute	0.0563
IX. Philadelphia-Camden-Wilmington, PA-NJ-DE-MD		
1.	University of Pennsylvania	1.0443
2.	Children's Hospital of Philadelphia	0.1102
3.	The Wistar Institute	0.0749
4.	Thomas Jefferson University	0.0617
5.	Fox Chase Cancer Center	0.0440
X. Houston-The Woodlands-Sugar Land, TX		
1.	The University of Texas MD Anderson Cancer Center	0.6632
2.	Baylor College of Medicine	0.6169
3.	The University of Texas Medical Branch at Galveston	0.0799
4.	The University of Texas Health Science Center at Houston	0.0664
5.	Rice University	0.0578

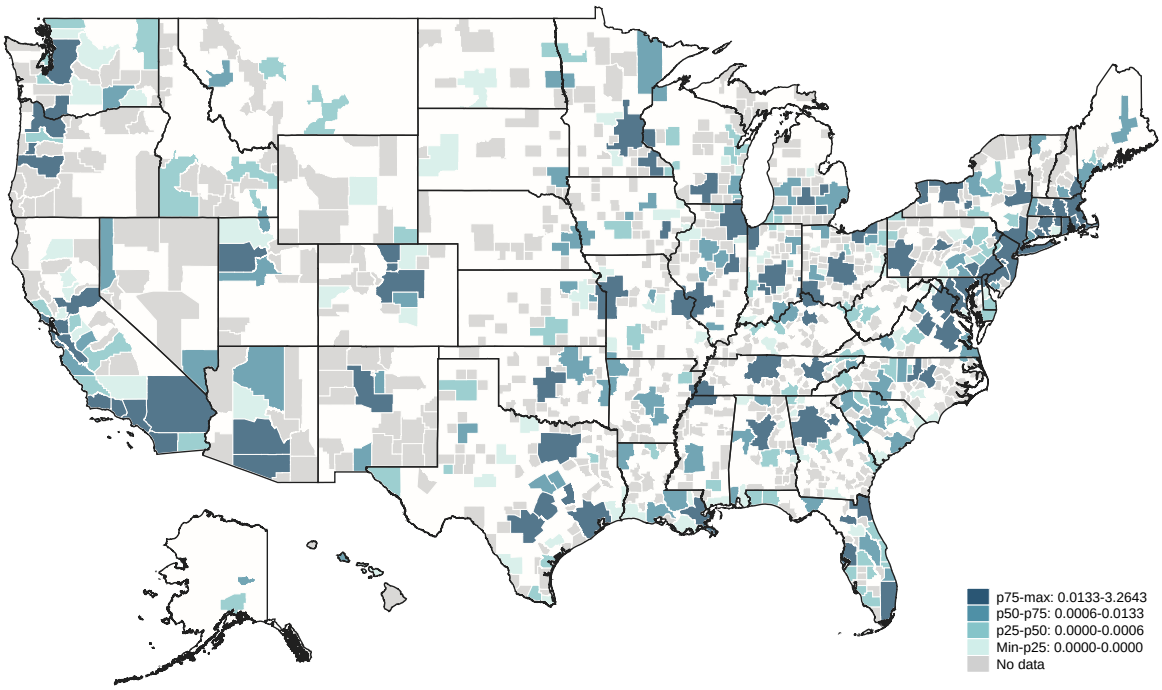
Notes: Table shows the leading institutions in the ten most productive cities by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers in each geography. Journals included are: *Cell, Nature, Science,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

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



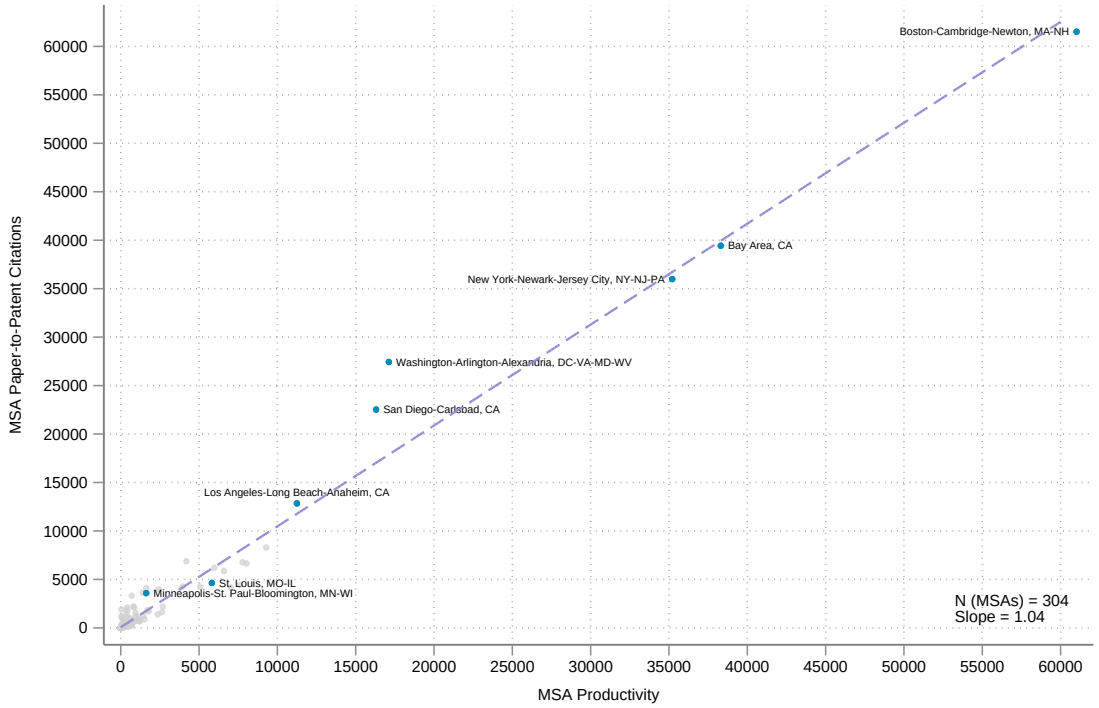
Notes: Figure shows the trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical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



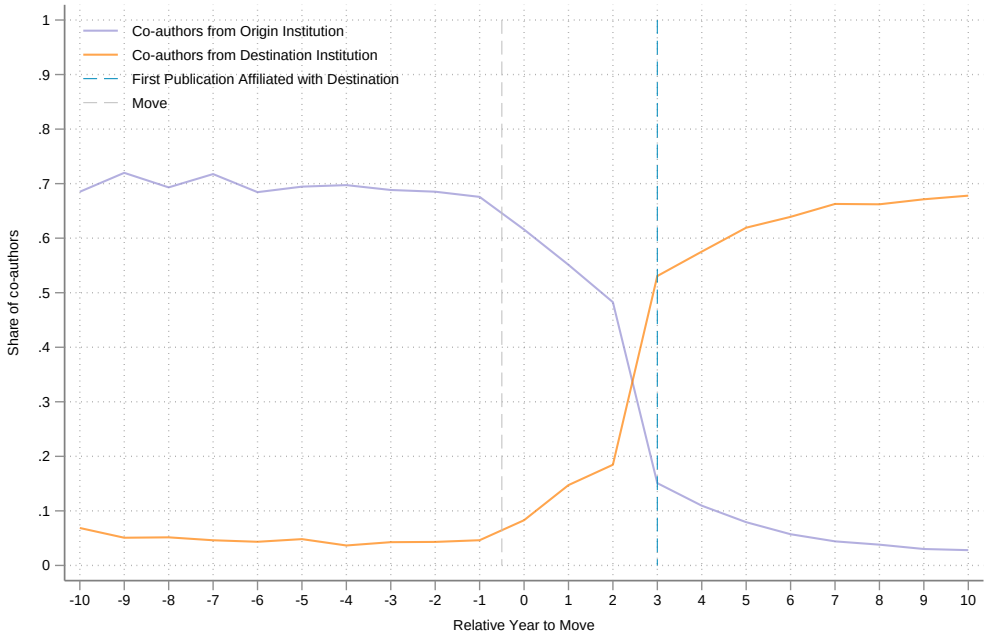
Notes: Figure shows a heat map of average fundamental science researcher productivity at the MSA-level between 1945 and 2023 in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature 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.

Figure 3: Fundamental Science Research Cited in Patents (1945-2023)



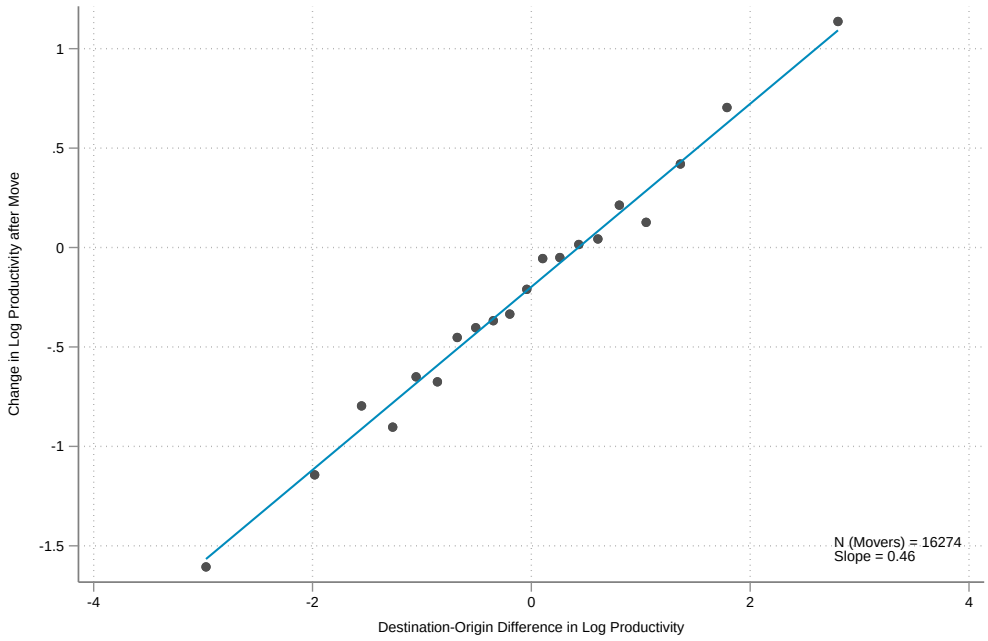
Notes: Figure shows a scatter plot of the number of papers in patents citations against the citation-adjusted measure of fundamental science research produced at the MSA-level. Around 15.5% of papers in our sample have ever been cited in a patent. There are N = 304 MSAs in our sample.

Figure 4: 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 defined as three year prior to the first publication at the mover's destination.

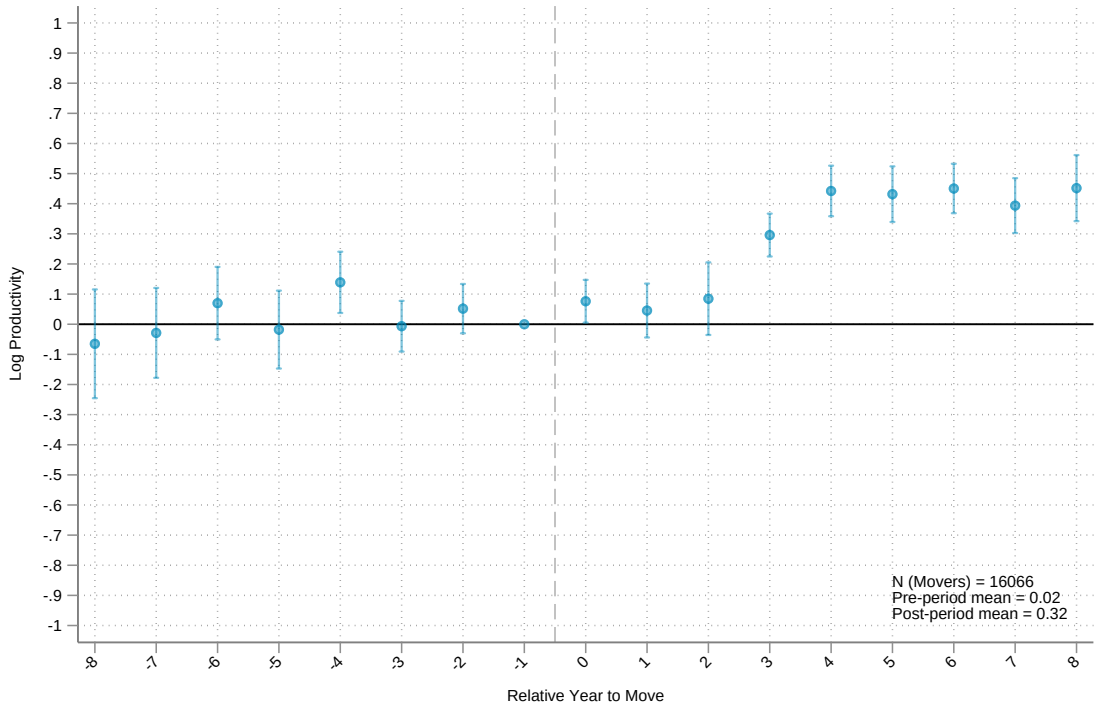
Figure 5: Change in Log Productivity by Move Size



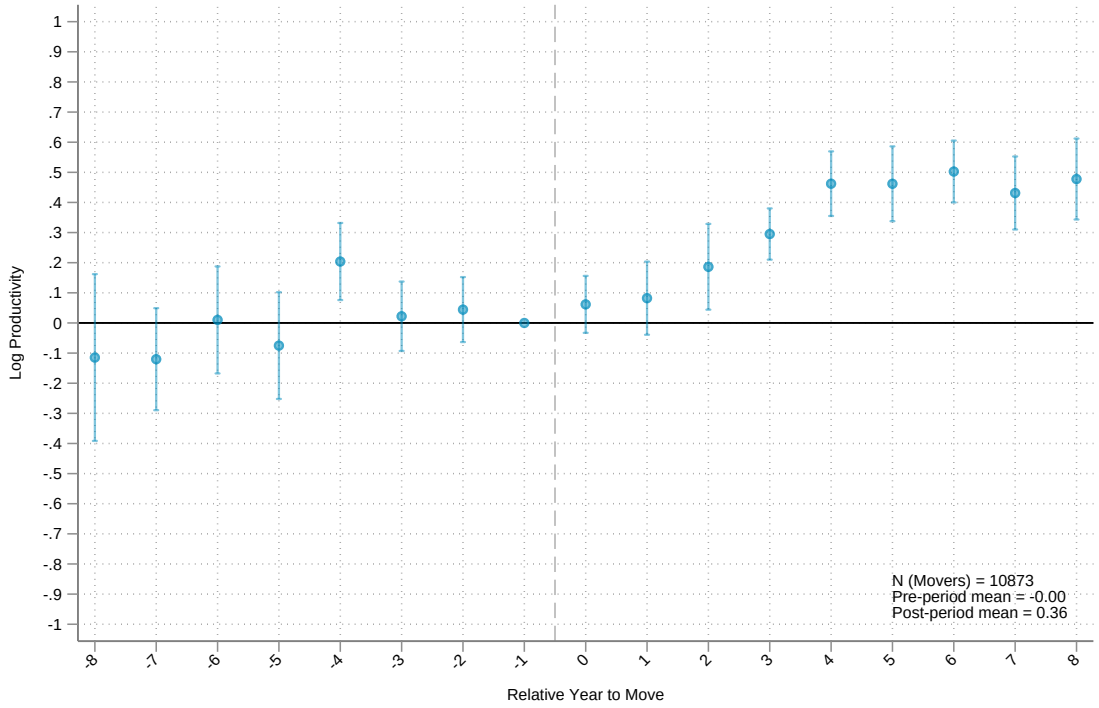
Notes: Figure shows the change in log productivity relative to a mover's productivity move size. Move size is defined as the difference in average produtivity between a mover's origin and destination institution.

Figure 6: Productivity 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 productivity and we control for a set of year, field, field x 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 (N = 16,066). Panel (a) shows the results for movers publishing anytime between 1945-2023 and panel (b) shows the results for movers publishing between 1995-2023 (N = 10,873).

Appendix

A. Additional Exhibits

Appendix Table A1: Journal Sample Sizes

Sample	Journal	% Fundamental	# Fundamental
CNS	Science	37.84	37,278
	Nature	80.92	60,615
	Cell	95.22	14,725
Leading Life-Sciences Journals	Neuron	94.48	7,200
	Nature Genetics	96.54	4,968
	Nature Medicine	85.66	4,111
	Nature Biotechnology	57.24	2,701
	Nature Neuroscience	91.11	3,670
	Nature Cell Biology	95.91	2,932
	Nature Chemical Biology	87.66	2,038
	Cell Stem Cell	92.43	1,514
Other Life-Sciences Journals	PLOS One	84.37	192,504
	Journal of Bio Chem	98.60	165,346
	Oncogene	98.89	17,660
	The FASEB Journal	96.30	11,696

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: Productivity Measure Robustness

	Country-Level	City-Level	Institution-Level
Research Productivity			
Unweighted vs. Impact Factor Weighted	0.9952	0.8997	0.8588
Unweighted vs. Impact Factor & Citation Weighted	0.9996	0.9961	0.9897
Papers-in-Patents Citations			
All Cites vs. Front Page Cites	0.9992	0.9921	0.9786
All Cites vs. Body Cites	1.0000	0.9961	0.9897

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: Productivity and Papers-in-Patents Citation Correlation Robustness

	State-Level	City-Level	Institution-Level
Research Productivity vs. Papers-in-Patents Citations			
Impact Factor & Citation Weighted vs. All Cites	0.9949	0.9920	0.9606
Impact Factor & Citation Weighted vs. Front Page Cites	0.9940	0.9927	0.9681
Impact Factor & Citation Weighted vs. Body Cites	0.9933	0.9854	0.9319

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

Appendix Table A4: Top 50 Institutions by Average Author Productivity

Rank	Institution	Author Productivity
1	Harvard University	4.5604
2	Stanford University	3.0957
3	Massachusetts Institute of Technology	2.8697
4	National Institutes of Health	2.0850
5	University of California, San Francisco	1.9653
6	Mass General Brigham	1.9362
7	Johns Hopkins University	1.6879
8	University of California, San Diego	1.3885
9	Columbia University	1.3856
10	Yale University	1.3046
11	University of California, Berkeley	1.1521
12	University of Pennsylvania	1.0443
13	University of Washington	1.0220
14	Memorial Sloan Kettering Cancer Center	0.9675
15	The University of Texas Southwestern Medical Center	0.8877
16	Dana-Farber Cancer Institute	0.8869
17	Cornell University	0.8792
18	Washington University in St. Louis	0.8726
19	Rockefeller University	0.8372
20	New York University	0.8236
21	University of California, Los Angeles	0.8181
22	University of Michigan–Ann Arbor	0.7631
23	The University of Texas MD Anderson Cancer Center	0.6632
24	Icahn School of Medicine at Mount Sinai	0.6189
25	Baylor College of Medicine	0.6169
26	California Institute of Technology	0.6042
27	Duke University	0.5928
28	Scripps Research Institute	0.5718
29	University of North Carolina at Chapel Hill	0.5619
30	Salk Institute for Biological Studies	0.5527
31	University of Chicago	0.5381
32	Princeton University	0.5333
33	New York Genome Center	0.5163
34	University of Wisconsin–Madison	0.4404
35	Northwestern University	0.4217
36	Cold Spring Harbor Laboratory	0.4200
37	Boston Children's Hospital	0.4071
38	University of Pittsburgh	0.3732
39	Beth Israel Deaconess Medical Center	0.3476
40	University of Southern California	0.3373
41	University of Minnesota	0.3288
42	Emory University	0.3189
43	University of Massachusetts Chan Medical School	0.3069
44	Vanderbilt University	0.3010
45	Boston University	0.2747
46	St. Jude Children's Research Hospital	0.2690
47	Albert Einstein College of Medicine	0.2664
48	University of Utah	0.2646
49	Fred Hutch Cancer Center	0.2523
50	The University of Texas at Austin	0.2461

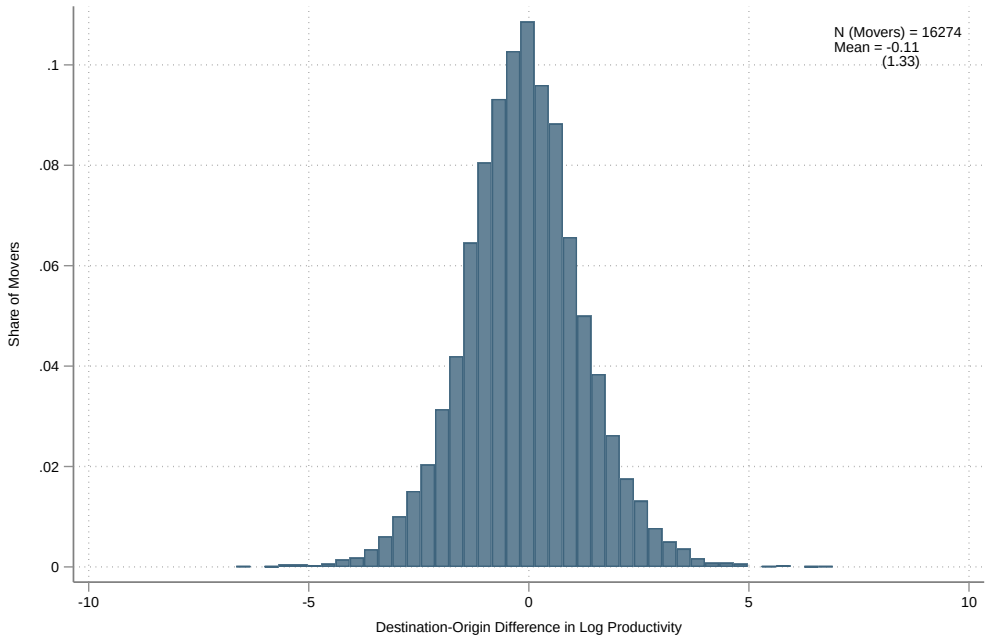
Notes: Table shows the top 50 institutions by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers at each institution. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics,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

Appendix Table A5: Top Places by Average Researcher Papers-in-Patents Citation

Rank	Place	Researcher Papers-in-Patents Citation
A. State		
1.	California	0.8164
2.	Massachusetts	0.6867
3.	New York	0.4739
4.	Maryland	0.3173
5.	Texas	0.1272
6.	Pennsylvania	0.1201
7.	New Jersey	0.0917
8.	Illinois	0.0852
9.	Washington	0.0803
10.	Connecticut	0.0777
11.	Ohio	0.0771
12.	Michigan	0.0763
13.	North Carolina	0.0742
14.	Virginia	0.0656
15.	Missouri	0.0645
16.	District of Columbia	0.0635
17.	Wisconsin	0.0554
18.	Colorado	0.0454
19.	Utah	0.0449
20.	Minnesota	0.0398
B. City		
1.	Boston-Cambridge-Newton, MA-NH	1.0880
2.	Bay Area, CA	0.6862
3.	New York-Newark-Jersey City, NY-NJ-PA	0.6860
4.	Washington-Arlington-Alexandria, DC-VA-MD-WV	0.5205
5.	San Diego-Carlsbad, CA	0.3684
6.	Los Angeles-Long Beach-Anaheim, CA	0.2785
7.	Baltimore-Columbia-Towson, MD	0.1484
8.	Ithaca, NY	0.1412
9.	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.1410
10.	Seattle-Tacoma-Bellevue, WA	0.1322
11.	Chicago-Naperville-Elgin, IL-IN-WI	0.1240
12.	Research Triangle Park, NC	0.1223
13.	Dallas-Fort Worth-Arlington, TX	0.0878
14.	St. Louis, MO-IL	0.0871
15.	Houston-The Woodlands-Sugar Land, TX	0.0865
16.	New Haven-Milford, CT	0.0858
17.	Madison, WI	0.0843
18.	Ann Arbor, MI	0.0838
19.	Columbus, OH	0.0777
20.	Salt Lake City, UT	0.0718
C. Institution		
1.	Harvard University	1.8057
2.	Massachusetts Institute of Technology	1.4815
3.	National Institutes of Health	1.3343
4.	Stanford University	0.8981
5.	University of California, San Francisco	0.6650
6.	Mass General Brigham	0.6559
7.	Cornell University	0.5951
8.	University of California, Berkeley	0.5610
9.	Johns Hopkins University	0.5344
10.	California Institute of Technology	0.4920
11.	Columbia University	0.4849
12.	Rockefeller University	0.4603
13.	University of California, San Diego	0.3917
14.	Yale University	0.3515
15.	University of California, Los Angeles	0.3492
16.	University of Washington	0.3463
17.	University of Michigan–Ann Arbor	0.3439
18.	The Ohio State University	0.3253
19.	Washington University in St. Louis	0.3244
20.	University of Wisconsin–Madison	0.3139

Notes: Table shows the top 20 states, cities, and institutions by average researcher papers-in-patents citation in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers in each geography or 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.

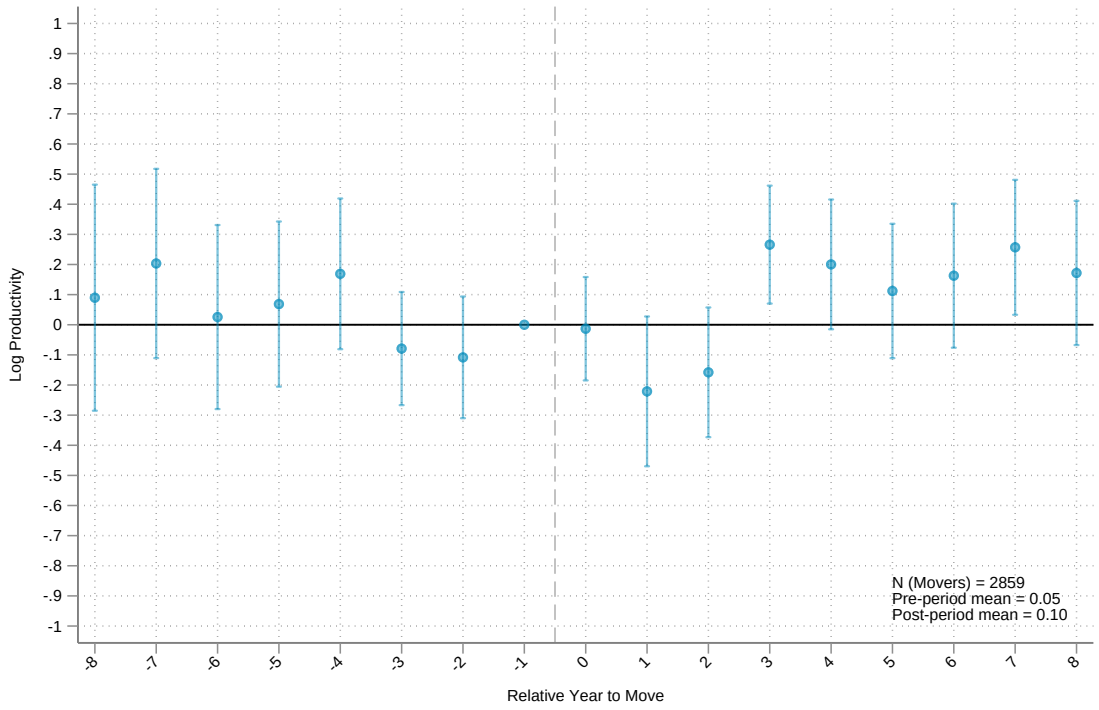
Appendix Figure A1: Distribution of δ_i



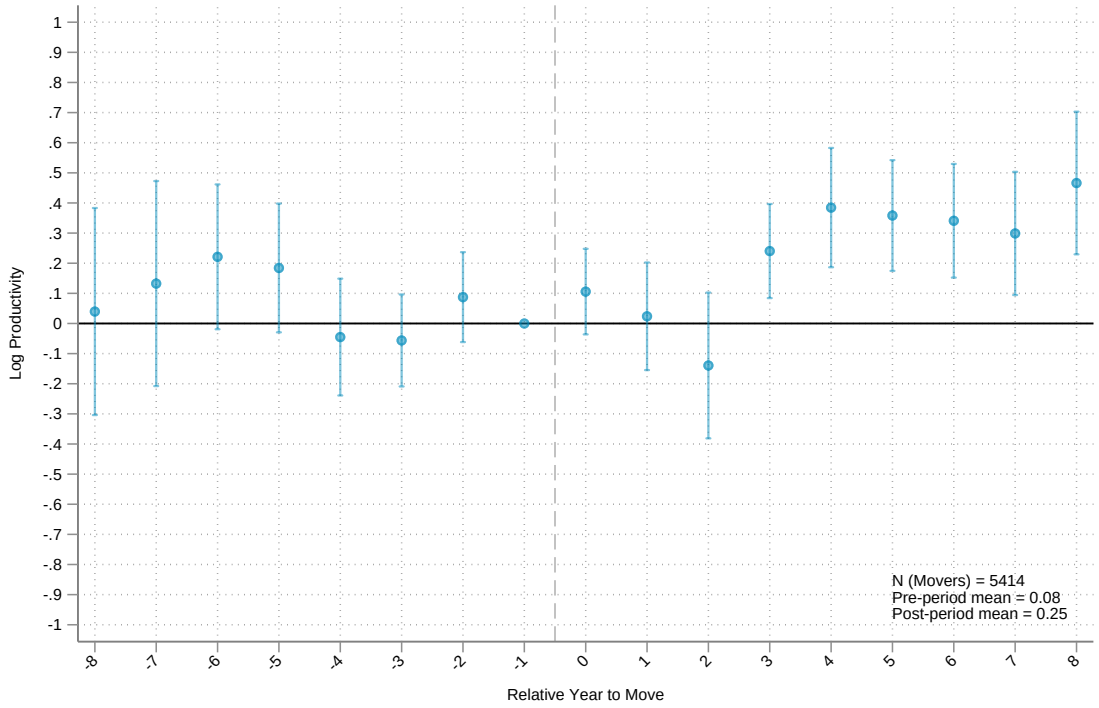
Notes: Figure shows the distribution of δ_i computed at the institution-level for scientist-movers. Here, δ_i measures the size of the move and is constructed as the average difference in productivity between the destination and origin institution.

Appendix Figure A2: Productivity Change in Response to Move (Other Time Periods)

(a) 1945-1980

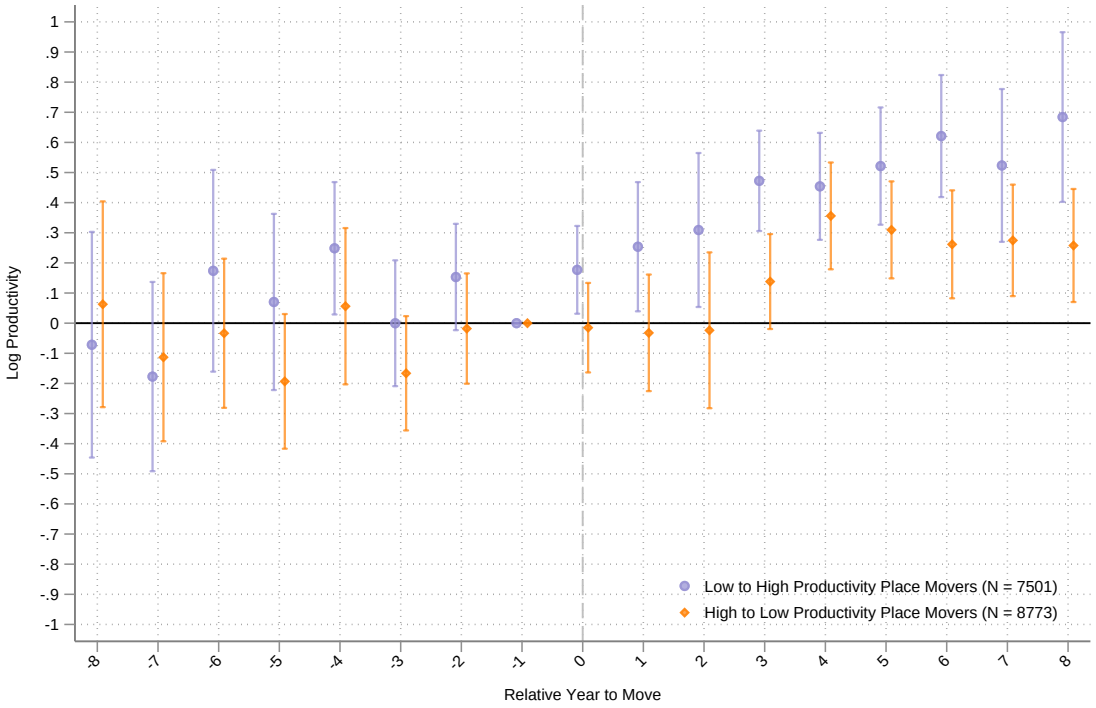


(b) 1980-1995



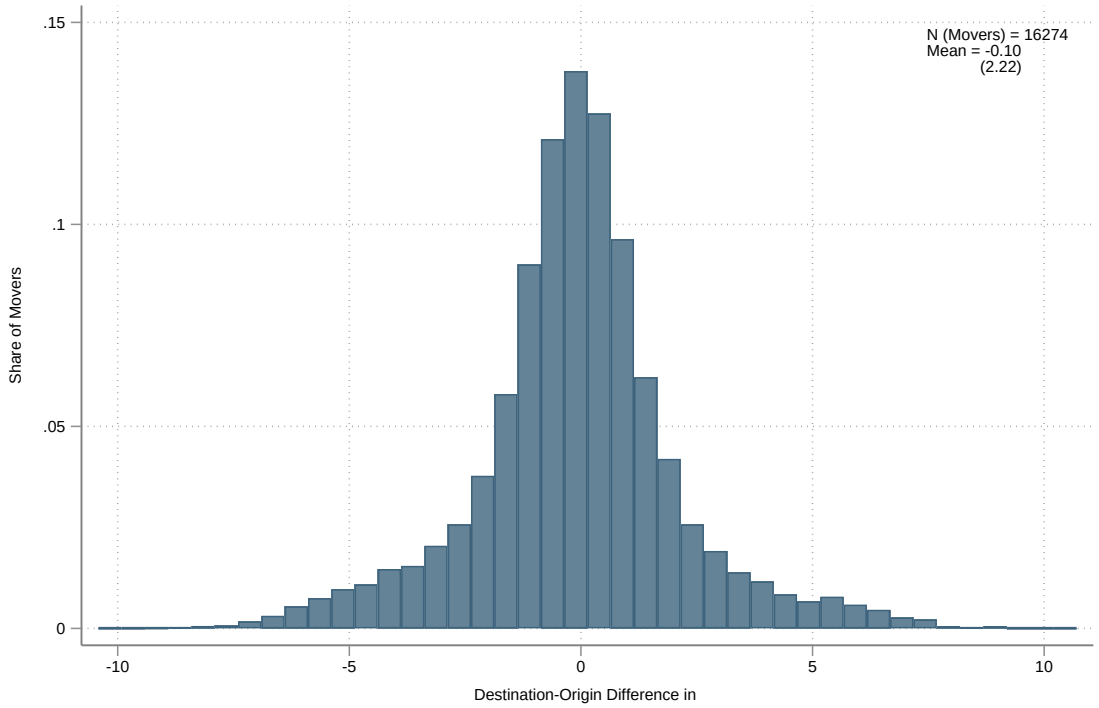
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, field, field x year, and author fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023. Panel (a) shows the results for movers publishing anytime between 1945-1980 (N = 2,859) and panel (b) shows the results for movers publishing between 1980-1995 (N = 5,414).

Appendix Figure A3: Productivity Change based on Move Size



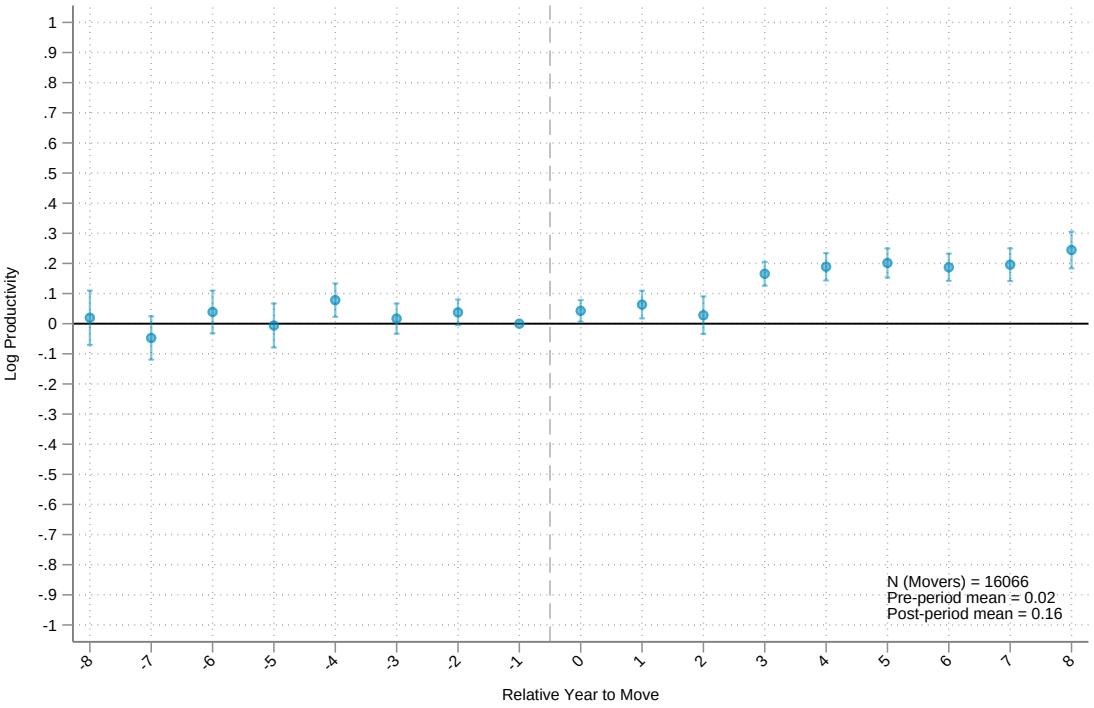
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers moving from high to low productivity places and from low to high productivity places for the time period between 1945-2023. The dependent variable is log productivity and we control for a set of year, field, field x year, and author fixed effects in our specification. Standard errors are clustered at the institution-level. Panel (a) shows the results for high-to-low movers publishing (N = 7,501) and panel (b) shows the results for low-to-high movers publishing (N = 8,773).

Appendix Figure A4: Distribution of Move Size Defined by Top 95th Percentile of Researchers at Institution



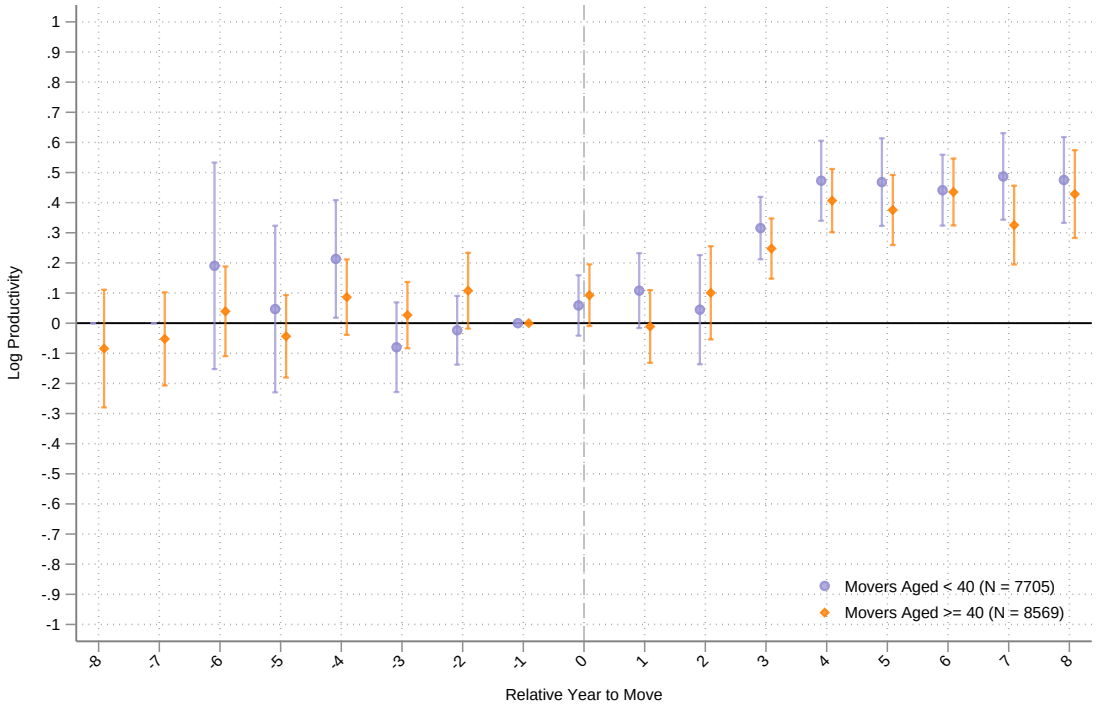
Notes: Figure shows the distribution of δ_i computed at the institution-level for scientist-movers based on the difference in productivity for the top 95th percentile researchers between the destination and origin institution.

Appendix Figure A5: Productivity Change in Response to Institution-level Star Productivity



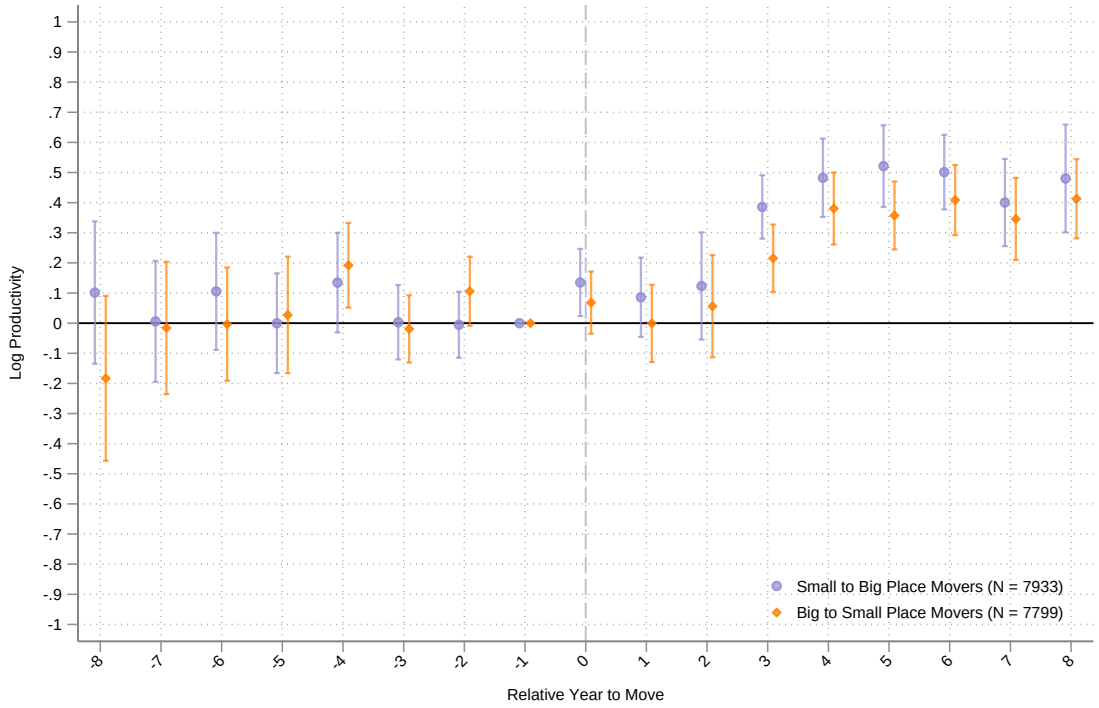
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers between 1945-2023 using δ_i defined off the top 95th percentile of researchers in an institution. The dependent variable is log productivity and we control for a set of year, field, field x year, and author fixed effects in our specification. Standard errors are clustered at the institution-level.

Appendix Figure A6: Place Effect Heterogeneity by Age



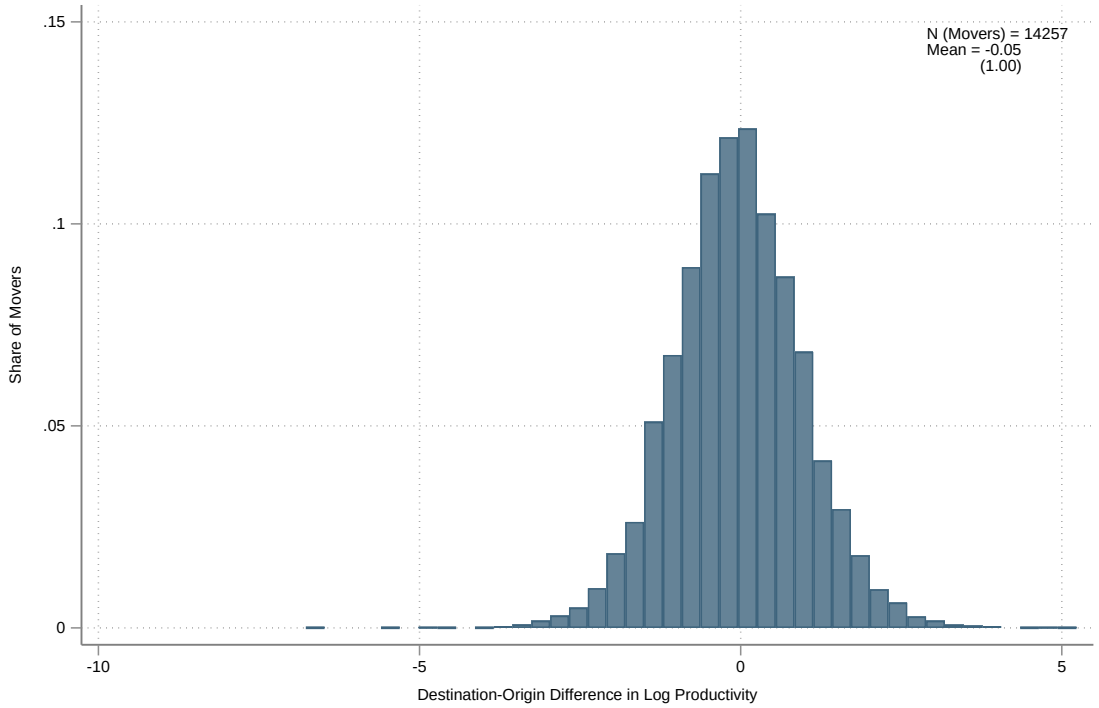
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers separated by their age at the time of move for the time period between 1945-2023. The dependent variable is log productivity and we control for a set of year, field, field x year, and author fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023. Panel (a) shows the results for researchers who move when they are less than 40 (N = 7,705) and panel (b) shows the results for movers older than 40 (N = 8,569).

Appendix Figure A7: Productivity Change based on Cluster Size Change



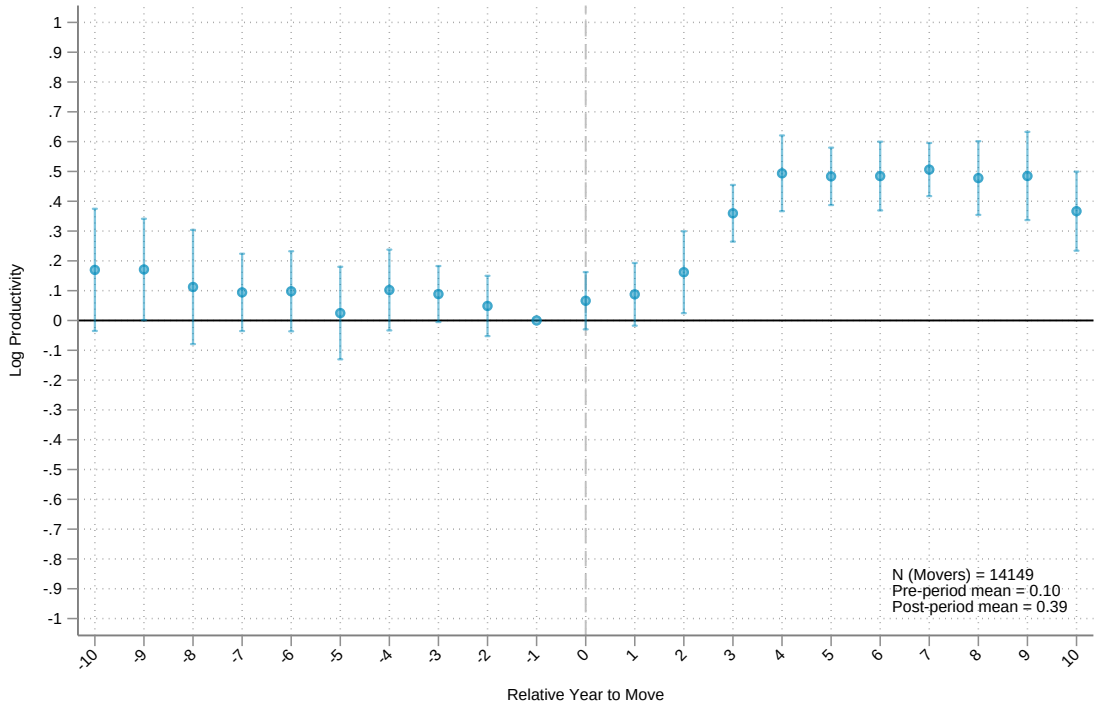
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers by the change in their origin and destination cluster size between the time period 1945-2023. The dependent variable is log productivity and we control for a set of year, field, field x year, and author fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023. Panel (a) shows the results for small-to-large cluster movers (N = 7,933) and panel (b) shows the results for large-to-small cluster movers (N = 7,799).

Appendix Figure A8: Distribution of City-Level Move Size



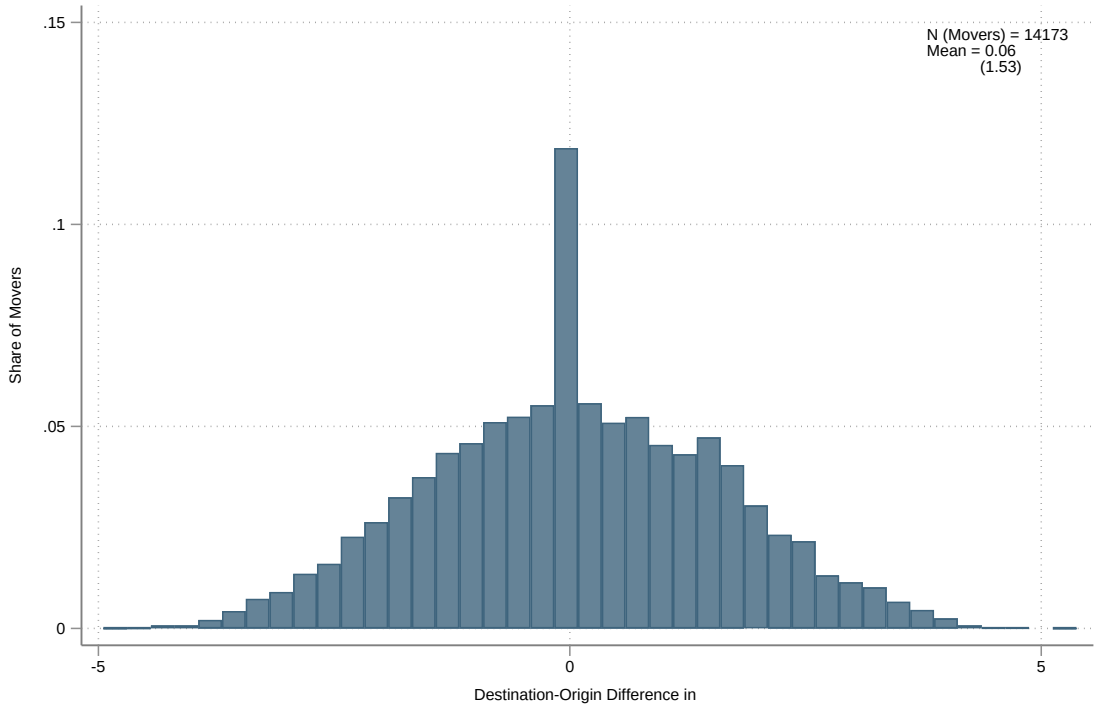
Notes: Figure shows the distribution of δ_i computed at the city-level for scientist-movers. Here, δ_i measures the size of the move and is constructed as the average difference in productivity between the destination and origin city.

Appendix Figure A9: Productivity Change in Response to City-level Productivity Move Size



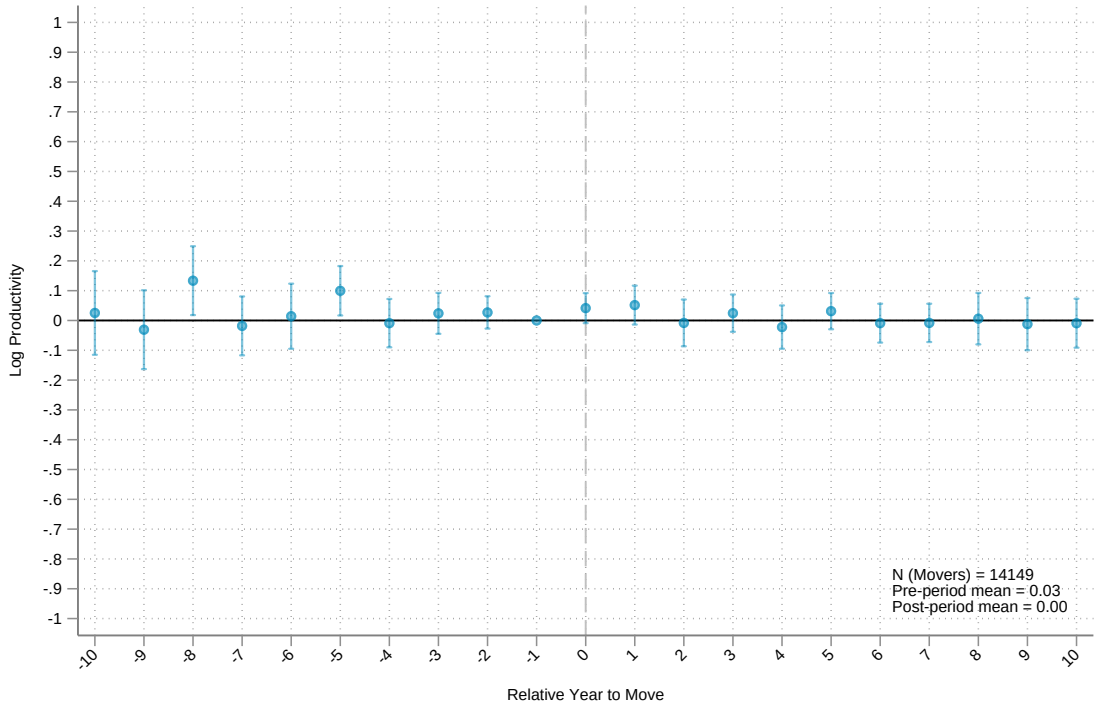
Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers by the change in their origin and destination cluster size between the time period 1945-2023 using the city-level δ_i. The dependent variable is log productivity and we control for a set of year, field, field x year, 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 A10: Distribution of City-Level Move Size Excluding Author's Own Institution



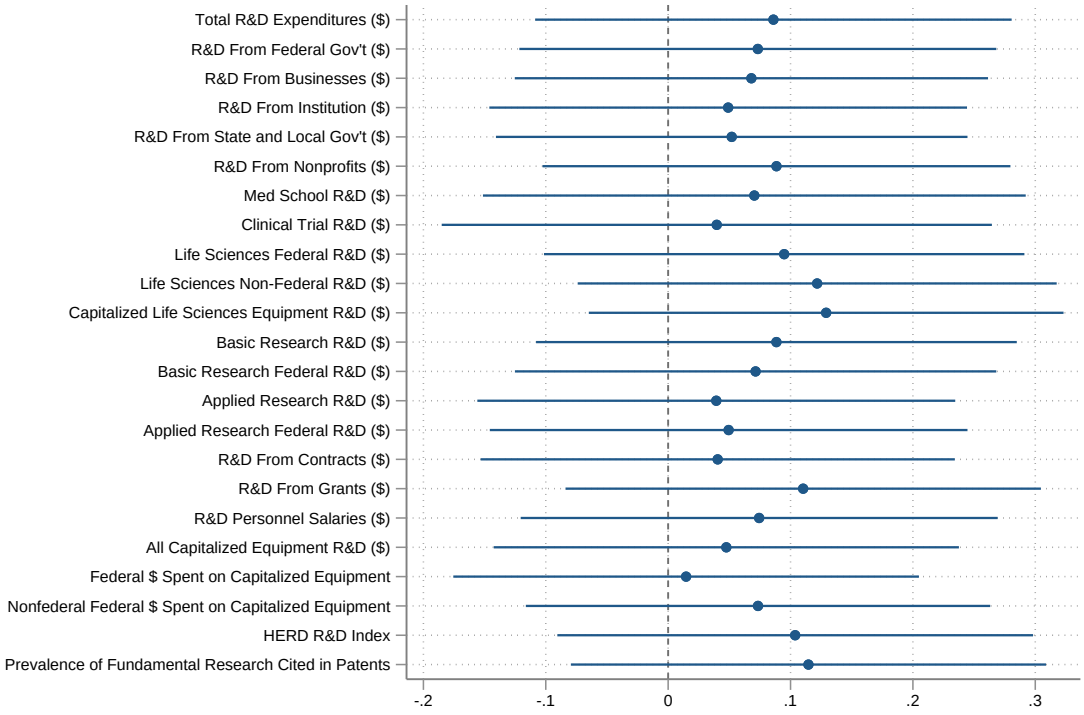
Notes: Figure shows the distribution of δ_i computed at the city-level for scientist-movers excluding the mover's own institution.

Appendix Figure A11: Productivity Change in Response to City-level Productivity Move Excluding Author's Institution Size



Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers by the change in their origin and destination cluster size between the time period 1945-2023 using the city-level δ_i excluding the author's own institution productivity. The dependent variable is log productivity and we control for a set of year, field, field x 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 A12: Correlates of Institution Effects



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

B. Productivity Weighting

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

Letting j denote journals and i denote papers, define

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

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

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

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

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

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

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

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

relative to other papers in the same journal. We calculate this as

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

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

Finally, we divide these paper-level weights by the number of authors A_i on a paper. Our final impact-factor and citation weighted productivity measure is equal to

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

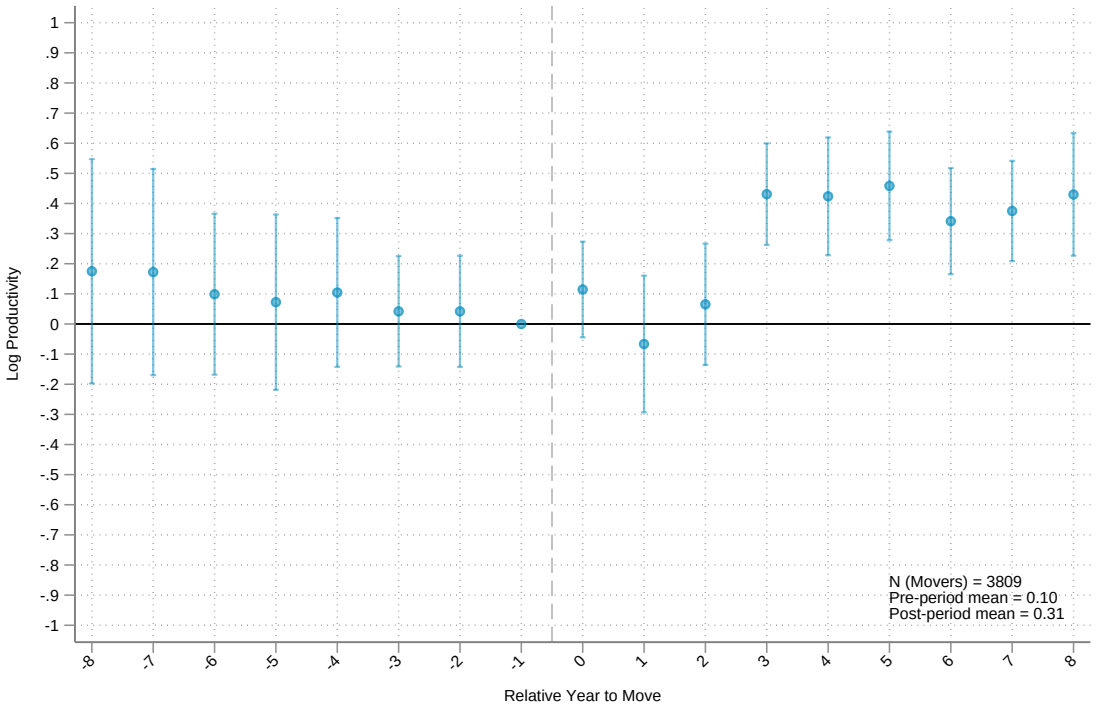
C. Constructing Life-Sciences Subfields

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

1. Molecular and Cellular Biology: cells, cell, expression, activation, protein, human, signaling, receptor, apoptosis, stem, kinase, growth, metabolism, gene, increased
2. Cancer Biology: cancer, breast, tumor, cells, cell, expression, tumors, lung, cancers, survival, growth, human, carcinoma, therapy, genetics
3. Genetics and genomics: genes, gene, genetics, dna, genetic, genome, expression, mutations, human, variants, identified, sequencing, sequence, loci, disease
4. Physiological and Neuroscience: physiology, disease, metabolism, species, neurons, increased, brain, infection, higher, drug, high, among, blood, also, human
5. Structural biology: protein, binding, proteins, domain, structure, complex, residues, enzyme, acid, site, receptor, cells, membrane, amino, sequence

D. Results for Cell, Science, and Nature

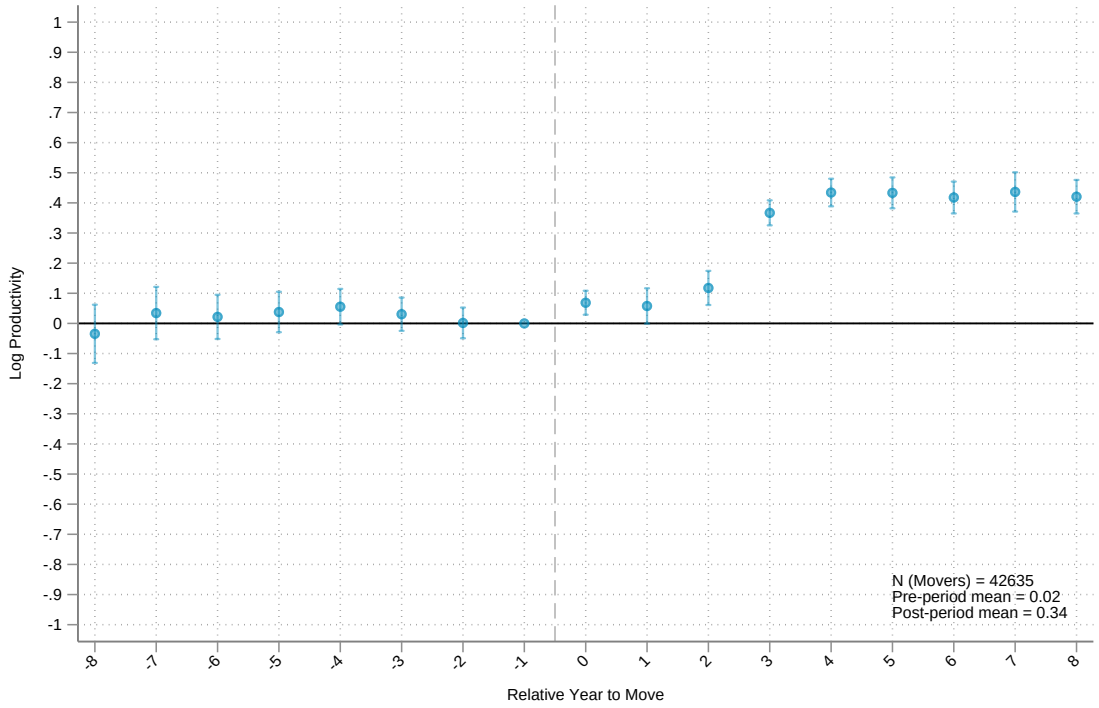
Appendix Figure A13: Productivity Change in Response to Move



Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for the sample of authors publishing in CNS. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, field, field x 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 (N = 3,809).

E. Results Including All Authors

Appendix Figure A14: Productivity Change in Response to Move



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, field, field x 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 (N = 42,635).