# Productivity Differences in Fundamental Life-Sciences Discovery*

Amitabh Chandra† Connie Xu‡

November 26, 2024

**Abstract**

Fundamental knowledge in the life sciences is generally discovered in universities and hospitals, and has profound implications for welfare by its effect on clinical medicine, new medical innovations, and subsequent scientific knowledge. To assess whether the institution where research is produced causally increases the productivity of scientists, we employ a scientist-mover design and find that 40−50% of the variation in research productivity between institutions is causally attributable to institutions. A significant share, 20−25% of the institutional effect reflects star researchers, with a small role for research expenditures. Institution effects are highly localized– a scientist's productivity is unaffected by the productivity of scientists outside the institution but in the same city. The magnitude of these effects has not decreased in the wake of technologies that make cross-institutional collaborations easier.

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

†Harvard University and NBER, amitabh_chandra@harvard.edu

‡Harvard University, conniexu@g.harvard.edu

Fundamental discoveries in life sciences such as the discovery of DNA, genes, mRNA, antibodies, and CRISPR have enabled new drugs and increased our cumulative understanding of physiology and disease (Zerhouni, 2005). This research is usually conducted in universities and hospitals and naturally raises the question of whether some institutions are causally more productive than others. Moreover, because fundamental science discoveries are likely to be underprovided by the market – understanding the forces that produce these limited discoveries has large implications for welfare, and how limited public and philanthropic resources 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, mentors, and collaborators (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 in places where they received their education, where the weather is better, or places where solving dual-career challenges is easier, then it may appear that some locations or institutions are more productive at producing knowledge, but this is confounded by the locational preferences of their scientists.

To understand whether the association between where a scientist works and their research productivity is causal, we implement a movers design by estimating an event study specification of an individual researcher's productivity on their "size of move", defined as the difference in institutional 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 should increase a scientist's productivity by 20%; conversely, if productivity in producing fundamental knowledge is entirely a scientist-level phenomenon, then moving to institutions of higher or lower productivity will not affect an individual's research productivity. Because the movers design links changes in an individual scientist's productivity to changes in their location, it controls for confounders like individual and time effects. This design can also reveal whether scientists who are becoming more productive move to more productive institutions, which would not be desirable for the identification of institution effects.

Our contribution is closely related to Lerner et al. (2024) who pioneered using a scientist-mover design of "wandering scholars" to estimate that university-level factors account for a fifth of the variation across universities in the production of commercially relevant knowledge, as measured by the degree to which patents cite academic research. Like them, we are interested in life sciences research

but focus on the creation of knowledge regardless of whether it has commercial implications; fundamental research is of interest in its own right even if it has no commercial implications. Moreover, even if it does affect patents, clinical practice, or therapeutics, may do so over many decades and in complex and non-linear ways that may render their estimates, which focus on relatively direct and immediate connections, quite conservative (Myers and Lanahan, 2022).[1]

To perform this analysis, we need to define fundamental science research, measure it in a systematic manner, and link it to the institution to which it was attributed. We define fundamental science research using the definition of the National Academy of Medicine[2] This definition encompasses traditional wet laboratory research (such as cell division and protein degradation), and translational research, which encompasses early testing of therapies in humans. The definition highlights several reasons for why fundamental research may be underprovided relative to the social optimal: fundamental science research understands biological processes and objects, such as transcription and genes, that are not patentable because they are natural phenomena for which patents are impermissible(Kesselheim et al., 2013). Some discoveries may be removed from commercial applications, making them unsuitable for the 'useful' criterion (also known as Section 101 criteria) to obtain 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, which makes understanding the institutions where such research is produced even more consequential(Nelson, 1959; Akcigit et al., 2021; Azoulay and Li, 2022; Myers and Lanahan, 2022).

We measure research productivity as the frequency with which a scientist is listed as the first or last author in a paper, adjusting for citation counts to the paper and journal impact factors (we show that these weighting choices are not consequential for our results). We gather publication data from *OpenAlex*, a comprehensive open catalog of global research (Priem et al., 2022). Our sample consists of 563,702 papers from 15 journals from 1945 to 2023 that were cited 65,046,333 times, and we show similar results in a smaller sample of papers published in

[1]Another reason for studying knowledge creation in the life-sciences is that it is an industry where the complete life-cycle of R&D– from academic research to patents to clinical trials, to new drug approvals– is observed. This would not be true for industries that rely on trade secrets over patents or where fundamental knowledge is not published in journals.

[2]The National Academy of Medicine (NAM) 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 leading three outlets— *Cell*, *Nature* and *Science* (henceforth, CNS) that were cited 31,449,839 times. We used information on the location of 556,769 first and last authors (212,904 US-based researchers) and their research productivity before and after moving from one institution to another (we validate this approach by examining the coauthorship patterns of scientists and whether a move predicts a change in coauthorship).

These data reveal that the production of fundamental science research in 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 metropolitan areas of Boston, the Bay Area, and New York. We find that fundamental research from more productive cities also receives 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 $40-50\%$ of the geographic variation in fundamental research productivity is accounted for by institutional-specific factors, which 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. The effect is asymmetric: researchers who relocate to more productive environments, as measured by a positive move size, experience a larger increase in their productivity than the decline in productivity when a researcher locates to a less productive institution. Disentangling the effect of institutional productivity further, we find that $20-25\%$ of the estimated institutional effect is driven by the presence of star researchers, who are at the 95th percentile of research productivity. Finally, our results suggest that productivity spillovers from researchers outside the institution, at other institutions, have little influence on a scientist's productivity; institutions affect scientist's productivity in a highly localized manner, whereas highly productive researchers at other institutions in the same city, do not.

We correlate our place effects with various institution-level variables such as the amount of R&D spending. This correlational evidence reveals a weakly 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, institutional research expenditure is not a first-order factor in influencing place effects.

These results have implications for a number of allocations that pertain to science-policy. There is 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 imply 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 that are unrelated to producing knowledge). At the same time, we find that institutional effects on scientist's productivity are localized, and do not spillover onto other scientists in the same city. This suggests that individual institutions in smaller locations can be consequential producers of fundamental science knowledge, for there is little evidence that clusters matter for research productivity. Finally, there is growing recognition of the importance of public support for the scientific workforce and early-career scientists through targeted funding and mentorship programs (Azoulay et al., 2011, 2019, 2021). If institutions exert a causal effect on scientists' productivity, then such may be more effective when used in the most productive institutions.

# I. Data and Measurement

## A. Measuring Fundamental Science Research

The NAM definition of fundamental science research, which we noted earlier, 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.[3]

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

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, Chemicals and Drugs, Organisms, Phenomena and Processes, Diseases, Anesthesia and Analgesia,* and *Therapeutics*. Reassuringly, the principal 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[4] 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*.[5] 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 37% (for *Science*) to 98% (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

definition used by NAM.

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

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

and standardizing information from more than 249,000 sources, such as publication databases like *PubMed* and Microsoft Academic Graph (MAG) as well as metadata database organizers like *Crossref* and *ORCID*.[6] As of the latest data release, *OpenAlex* indexes over 250 million works by 95 million authors affiliated with 109,000 different institutions. These data are updated frequently, often every few hours through its API.

*OpenAlex* offers several advantages over comparable databases like *PubMed* or *Web of Science* for identifying the geography of research because of intentional 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[7].

## B. Measuring Scientist's Research Productivity

We move from publications to the research productivity of individual scientists in these publications by characterizing the "effective" research contribution of each author associated with a publication. We first divide a paper by the number of

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

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

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 imply that larger 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 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 as the first or as the 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). Our focus on first and last authors results in conservative estimation of place effects, for we find that place-specific factors are around 10% smaller for them relative to the sample of middle authors. This is likely driven by the fact that the first and last authors are more likely to represent influential researchers, who are less affected by place effects, relative to the middle authors.

## C. Measuring Institutional Productivity

We are interested in identifying the home institution where scientists physically carry out their research which is challenging because some authors are affiliated with multiple institutions at the same time. We infer each author's principal institution by gathering the full publication history of each author and using publication metadata to identify the most consistent affiliation for an author in any given year. If there is a unique modal affiliation for an author in a given year, we choose that one. We break ties by first identifying any "sandwich" patterns in consecutive years. For example, if institution A appears in 2008 and 2010 but not in 2009, we designate institution A as the principal affiliation for all years. If an author is affiliated with both institutions A and B in a given year, we first assign the principal institution as A if A appeared in the prior year. If there is no affiliation from the previous year, we assign the principal 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 principal institution. This last step affects less than 1% of author-years.

Having assigned each author to a unique institution in a year, we are able to examine research productivity 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 Clara (similarly, we define the Research Triangle Park, NC as the combination of Durham-Chapel Hill and Raleigh-Cary).

Finally, we account for differential changes in productivity across life-sciences subfields stemming from breakthrough discoveries we categorize each author-year into one of five life-sciences research fields (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). 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. 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 at the author-year level rather than the author level, we allow for changes in a scientist's research focus over time. 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[8]. Using the latest 5-year impact factor of *Science* (54.5)

[8]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

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 as working in the subfield of molecular and structural biology[9]. After adjusting for *Nature*'s impact factor (60.9) and the average citation count per year since publication ($1647/12 \approx 137$), both Samuel Sternberk and Jennifer A. Doudna (and their affiliated institutions) are given an effective contribution of 6.6 articles for this publication. To put these magnitudes of research productivity into context, Table 1 shows that the average author in an elite set of scientific papers has an annual productivity of 0.72 articles and a lifetime productivity of 1.64 articles (measured up until the end of our time frame in 2023).

# II. Productivity Variation in Fundamental Science Discovery

## A. Geographic Variation in Research Productivity

Figure 1 shows the aggregate share of output produced by each country since 1945[10]. The data reveal that the US has consistently dominated fundamental science production, accounting for more than half of the global output and exceeding other countries by a substantial margin.[11] Thus, given the US' large role in fundamental

her research to molecular and cellular biology and genetics and genomics

[9]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.

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

[11]While this fact highlights the pivotal role of the US in producing fundamental science research, also note the emergence of new entrants and the decline of others. As depicted in Figure 1, the decline in the US' share of production from 70% in the early 1990s to 50% today is partially the

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, we analyze the cross-sectional variation in our productivity measure at the author-level.[12] Table 2 presents the top 20 cities by average researcher productivity. Our measure of research productivity is interpreted as the "effective" number of publications for an author. Thus, the average researcher in the Boston metropolitan area is effectively publishing over three times as much as those in the DC metropolitan area. Table 3 further highlights the top five institutions in the ten most productive cities. These results show that the average researcher at NIH is effectively publishing twice as many publications per year than the average researcher at the University of Pennsylvania. Moreover, we see that there is balanced presence of highly productive researchers at universities, research organizations, government agencies, and biotechnology companies within a city. Appendix Table A4 provides the list of the top 50 institutions by average researcher productivity.

# 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 (Lerner et al., 2024) 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.

We motivate our estimating equation for an individual scientist's log productiv-

consequence of China's growing presence in fundamental science production. Since the early 2000s, when China first entered the set of large producers of science, China has made remarkable strides in producing fundamental science research— even surpassing the UK as the second largest producer in recent years. Also noteworthy is the shrinking of the United Kingdom's role from the post-WWII era when it produced about 25% of global science to the most recent period where its share is less than 10%.

[12]When calculating average productivity across author-years within a given geography, we 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.

ity, $y_{ijt}$, at institution $j$ in year $t$ as:

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

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

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 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, $\gamma_j$ 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 $\alpha_i$ and $\gamma_j$ are additively separable in log productivity, which implies that the level of place and person characteristics affects 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 $\gamma_j$ 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. Scientist Movers

We define movers ("wandering scientists") as researchers 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 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 before the first observed publication with the destination institution. Figure 3 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 31,627 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 1 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, $\delta_i$, 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 $\delta_i$ 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 $\delta_i$ 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 $\delta_i$ 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 $\tau_{ft}$ and $\rho_{r(i,t)}$ are field by year and relative year fixed-effects, respectively. We combine $\overline{y_{o(i)}}$ and $\alpha_i$ 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, $\delta_i$. 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 on average captures how much place-specific factors influence productivity. $\theta_{r(i,t)} = 1$ would imply that the mover's productivity has fully converged to that at 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, and 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 standard deviation is 1.33, with a significant number of large moves in either direction. 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 ($\hat{\delta}_i < 0$). To put the -0.11 into perspective, the typical "wandering scholar" in our sample is moving from the University of Southern California to UT-Austin (see Appendix Table A4).

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 4 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 $\alpha_i$ and $\gamma_j$ in equation 1.

We report event study results from equation 2 in Figure 5a 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 scientists.

The figure shows that there is no significant trend in $\theta_{r(i,t)}$ in the eight years 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 4 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 5b 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.

## 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 A2. 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 $\delta_i$ to represent the difference between the average productivity of stars in the destination and origin institutions. Appendix Figure A1 illustrates the distribution of this revised move size, while Figure 6a 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 and everything associated with them, are driving around 50% of the place effect we see in our baseline results.

Motivated by prior research on the life-cycle productivity of researchers and agglomeration effects (Dietz and Bozeman, 2005; Jones, 2010, 2009; Levin and Stephan, 1991; Myers, 2020; Carlino and Kerr, 2015; Kerr and Robert-Nicoud, 2020; Moretti, 2021), we also examine heterogeneity in place effects by mover age and cluster size. These results show similar magnitudes of place effects, regardless of whether researchers move before or after the age of 40 or whether they relocate to larger or smaller clusters. This suggests that neither age nor cluster size significantly impacts the role of place effects in changing researcher productivity.

Having explored the heterogeneity in place-effects, we follow the standard practice in movers designs and examine correlates of these place-based effects. We project the institution-level fixed effects on various institution characteristics from the annual Higher Education Research and Development (HERD) survey conducted by the National Science Foundation. Appendix Figure A3 presents the results of these bivariate regressions, which, despite limited statistical power, show a weakly positive correlation between institutional R&D expenditure and place effects.

## C. Fundamental Science and Commercial Relevance

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

We plot the city-level relationship between fundamental science research productivity and "papers in patents" productivity in Figure 2. 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.[13]

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

## D. Robustness

We assess the sensitivity of our results by constructing the move size at different levels of aggregation. Rather than defining $\delta_i$ at the institution-level, we define an alternative move size at the city-level excluding 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. Appendix Figure A1 presents the distribution of this move size and Figure 6b 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.

# 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 40-50% of the observed variation in productivity. These effects have not diminished over time, even as across-institution collaborations have become easier. Our findings underscore the significant role of institutional factors in shaping the landscape of fundamental life-sciences research. By focusing limited public and philanthropic resources on more productive institutions, policymakers and philanthropists can enhance the impact of funding.

Our results contribute to a larger research effort that measures various aspects of the production function that generates innovation. We focus on one determinant of fundamental science research– the location where it happens, but 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.

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. 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.

# 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.

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.

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: Summary Statistics

| | Mean | Std. Dev. |
|---|---|---|
| **A. Individual-Level** | | |
| **Everyone (N = 182,921)** | | |
| Years Actively Publishing | 2.04 | 2.68 |
| Average Team Size | 1.98 | 0.13 |
| Annual Cluster Size | 7174.40 | 7157.17 |
| Annual Productivity | 0.72 | 3.78 |
| Lifetime Productivity | 1.64 | 9.79 |
| Number of Moves | 0.24 | 0.61 |
| **Non-Movers (N = 151,294)** | | |
| Years Actively Publishing | 1.44 | 1.49 |
| Average Team Size | 1.98 | 0.13 |
| Annual Cluster Size | 7202.66 | 7402.22 |
| Annual Productivity | 0.69 | 3.93 |
| Lifetime Productivity | 1.01 | 5.92 |
| **Movers (N = 31,627)** | | |
| Years Actively Publishing | 4.92 | 4.55 |
| Average Team Size | 1.96 | 0.12 |
| Annual Cluster Size | 7039.19 | 5842.69 |
| Annual Productivity | 0.86 | 2.96 |
| Lifetime Productivity | 4.68 | 19.38 |
| **One-time Movers (N = 19,086)** | | |
| Years Actively Publishing | 3.94 | 3.40 |
| Average Team Size | 1.97 | 0.12 |
| Annual Cluster Size | 7078.44 | 5965.63 |
| Annual Productivity | 0.82 | 2.62 |
| Lifetime Productivity | 3.48 | 15.24 |
| **B. City-Level** | | |
| Number of Institutions | 79.01 | 80.14 |
| Annual Cluster Size | 7110.81 | 7635.02 |
| Annual Productivity | 0.86 | 4.60 |
| Lifetime Productivity | 132.52 | 958.16 |

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

Table 2: Cities with Most Productive Researchers on Average

| Rank | City | Average Author Productivity |
|---|---|---|
| 1 | Boston-Cambridge-Newton, MA-NH | 3.02 |
| 2 | Bay Area, CA | 1.95 |
| 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 | 0.49 |
| 8 | Seattle-Tacoma-Bellevue, WA | 0.43 |
| 9 | Houston-The Woodlands-Sugar Land, TX | 0.41 |
| 10 | Philadelphia-Camden-Wilmington, PA-NJ-DE-MD | 0.41 |
| 11 | New Haven-Milford, CT | 0.36 |
| 12 | Research Triangle Park, NC | 0.36 |
| 13 | Chicago-Naperville-Elgin, IL-IN-WI | 0.34 |
| 14 | St. Louis, MO-IL | 0.28 |
| 15 | Dallas-Fort Worth-Arlington, TX | 0.25 |
| 16 | Ithaca, NY | 0.23 |
| 17 | Ann Arbor, MI | 0.21 |
| 18 | Pittsburgh, PA | 0.16 |
| 19 | Trenton, NJ | 0.15 |
| 20 | Atlanta-Sandy Springs-Roswell, GA | 0.13 |

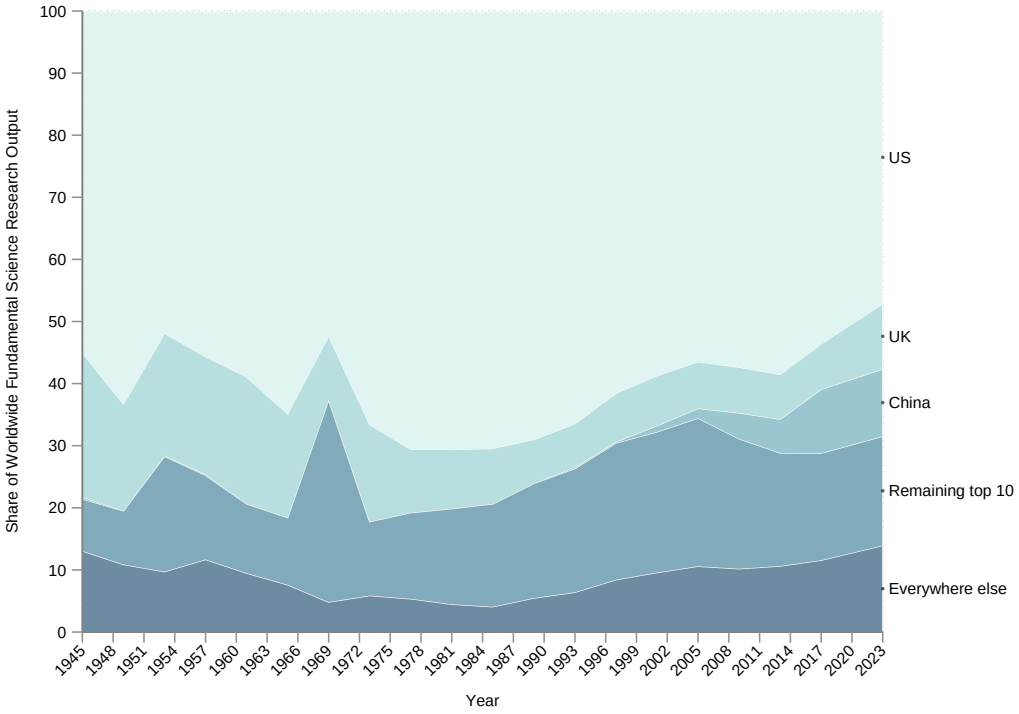
Notes: Table shows the top 20 cities by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the total number of researchers in each location. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics,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

Table 3: Leading Institutions in the Ten Most Productive Cities by Average Researcher Productivity

| Rank | Institution | Average Researcher Productivity |
|---|---|---|
| **I. Boston-Cambridge-Newton, MA-NH** | | |
| 1. | Harvard University | 4.61 |
| 2. | Massachusetts Institute of Technology | 2.80 |
| 3. | Mass General Brigham | 1.77 |
| 4. | Dana-Farber Cancer Institute | 0.86 |
| 5. | Boston Children's Hospital | 0.38 |
| **II. Bay Area, CA** | | |
| 1. | Stanford University | 3.02 |
| 2. | University of California, San Francisco | 1.93 |
| 3. | University of California, Berkeley | 1.14 |
| 4. | Lawrence Berkeley National Laboratory | 0.17 |
| 5. | Google (United States) | 0.17 |
| **III. New York-Newark-Jersey City, NY-NJ-PA** | | |
| 1. | Columbia University | 1.35 |
| 2. | Memorial Sloan Kettering Cancer Center | 0.94 |
| 3. | Rockefeller University | 0.81 |
| 4. | New York University | 0.77 |
| 5. | Icahn School of Medicine at Mount Sinai | 0.60 |
| **IV. Washington-Arlington-Alexandria, DC-VA-MD-WV** | | |
| 1. | National Institutes of Health | 2.09 |
| 2. | University of Maryland, College Park | 0.21 |
| 3. | Carnegie Institution for Science | 0.12 |
| 4. | Georgetown University | 0.08 |
| 5. | National Aeronautics and Space Administration | 0.06 |
| **V. San Diego-Carlsbad, CA** | | |
| 1. | University of California, San Diego | 1.36 |
| 2. | Scripps Research Institute | 0.56 |
| 3. | Salk Institute for Biological Studies | 0.53 |
| 4. | La Jolla Institute For Allergy & Immunology | 0.13 |
| 5. | Sanford Burnham Prebys Medical Discovery Institute | 0.11 |
| **VI. Los Angeles-Long Beach-Anaheim, CA** | | |
| 1. | University of California, Los Angeles | 0.79 |
| 2. | California Institute of Technology | 0.59 |
| 3. | University of Southern California | 0.35 |
| 4. | University of California, Irvine | 0.21 |
| 5. | Cedars-Sinai Medical Center | 0.11 |
| **VII. Baltimore-Columbia-Towson, MD** | | |
| 1. | Johns Hopkins University | 1.66 |
| 2. | University of Maryland, Baltimore | 0.09 |
| 3. | Lieber Institute for Brain Development | 0.01 |
| 4. | Osiris Therapeutics (United States) | 0.01 |
| 5. | Avidea Technologies (United States) | 0.01 |
| **VIII. Seattle-Tacoma-Bellevue, WA** | | |
| 1. | University of Washington | 1.00 |
| 2. | Fred Hutch Cancer Center | 0.25 |
| 3. | Allen Institute | 0.10 |
| 4. | Institute for Systems Biology | 0.09 |
| 5. | Seagen (United States) | 0.04 |
| **IX. Philadelphia-Camden-Wilmington, PA-NJ-DE-MD** | | |
| 1. | University of Pennsylvania | 1.05 |
| 2. | Children's Hospital of Philadelphia | 0.10 |
| 3. | The Wistar Institute | 0.07 |
| 4. | Thomas Jefferson University | 0.06 |
| 5. | Fox Chase Cancer Center | 0.04 |
| **X. Houston-The Woodlands-Sugar Land, TX** | | |
| 1. | The University of Texas MD Anderson Cancer Center | 0.66 |
| 2. | Baylor College of Medicine | 0.60 |
| 3. | The University of Texas Medical Branch at Galveston | 0.09 |
| 4. | The University of Texas Health Science Center at Houston | 0.07 |
| 5. | Rice University | 0.06 |

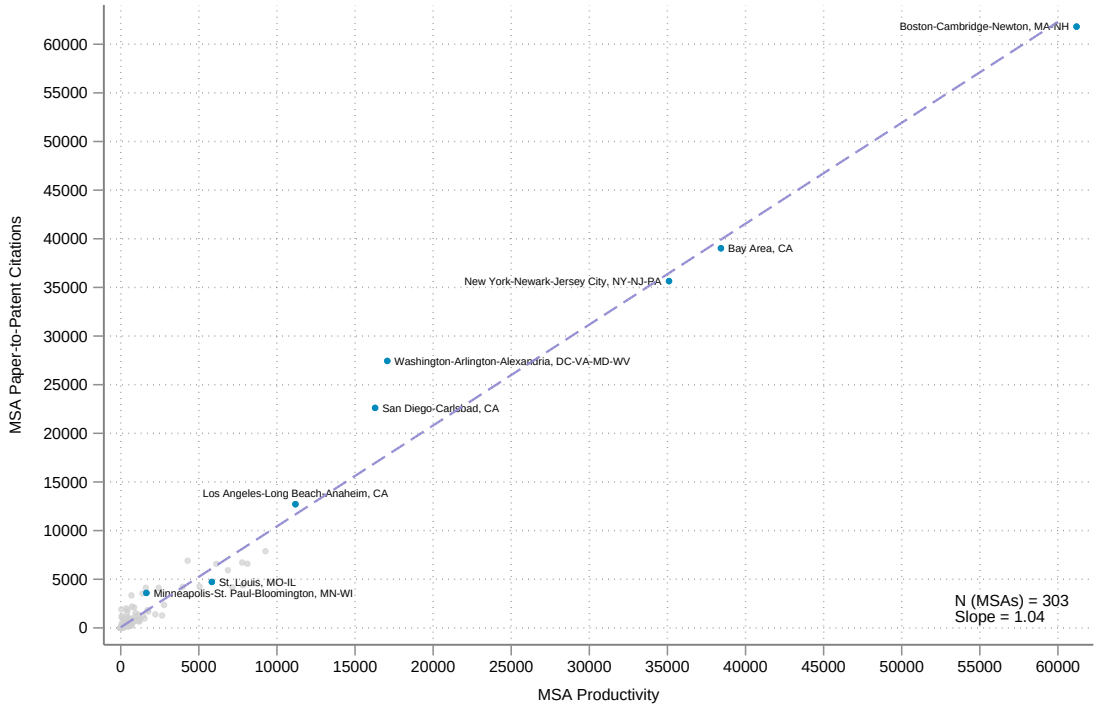
Notes: Table shows the leading institutions in the ten most productive cities by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers in each geography. Journals included are: *Cell, Nature, Science, 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)



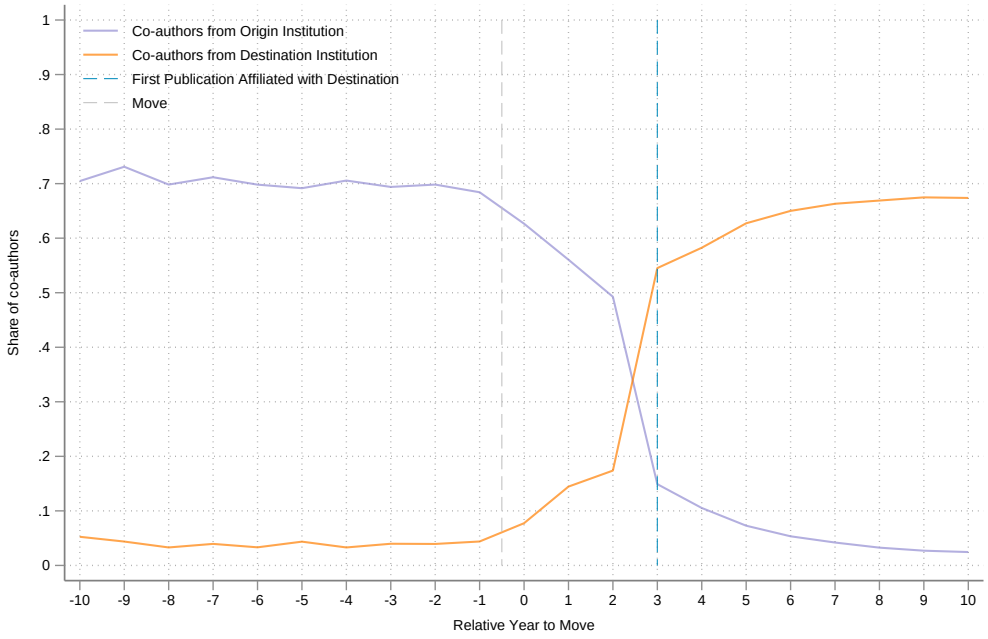

Notes: Figure shows the trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical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



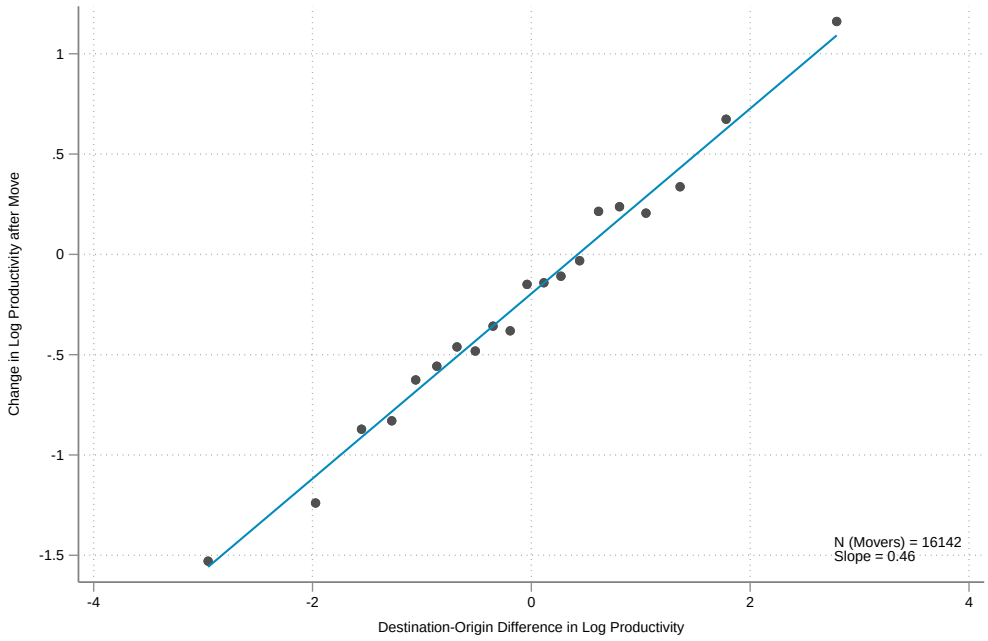

Notes: Figure shows a scatter plot of the number of papers in patents citations against the citation-adjusted measure of fundamental science research produced at the MSA-level. Around 15.5% of papers in our sample have ever been cited in a patent. There are N = 304 MSAs in our sample.

Figure 3: 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.

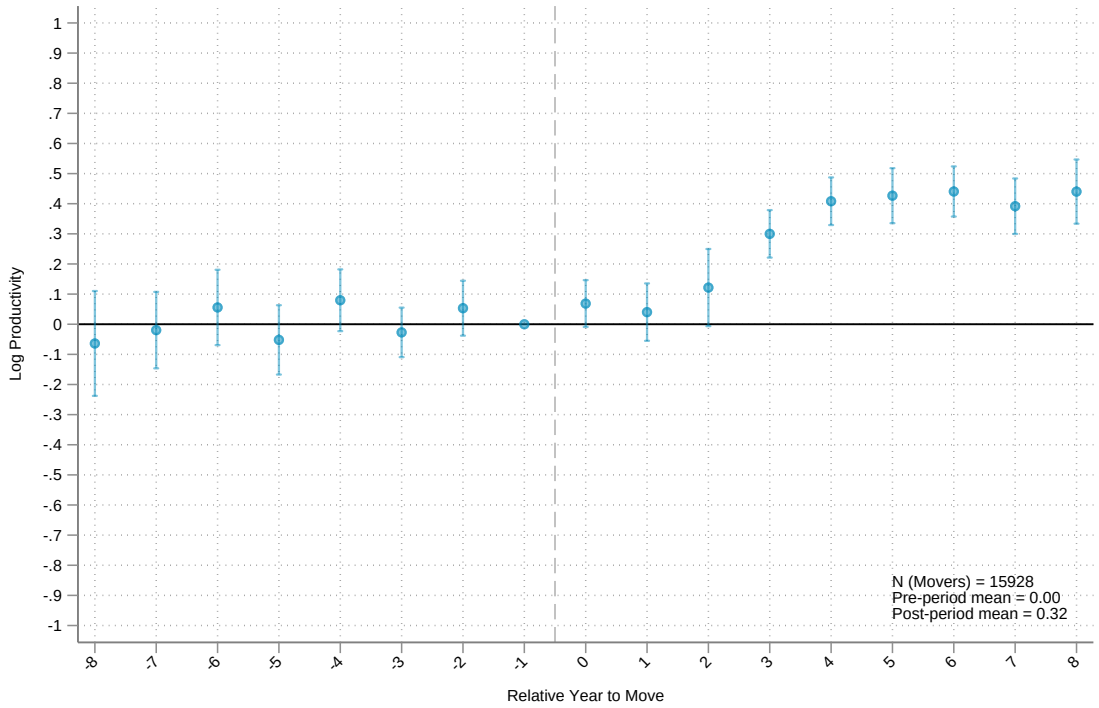
Figure 4: Change in Log Productivity by Move Size



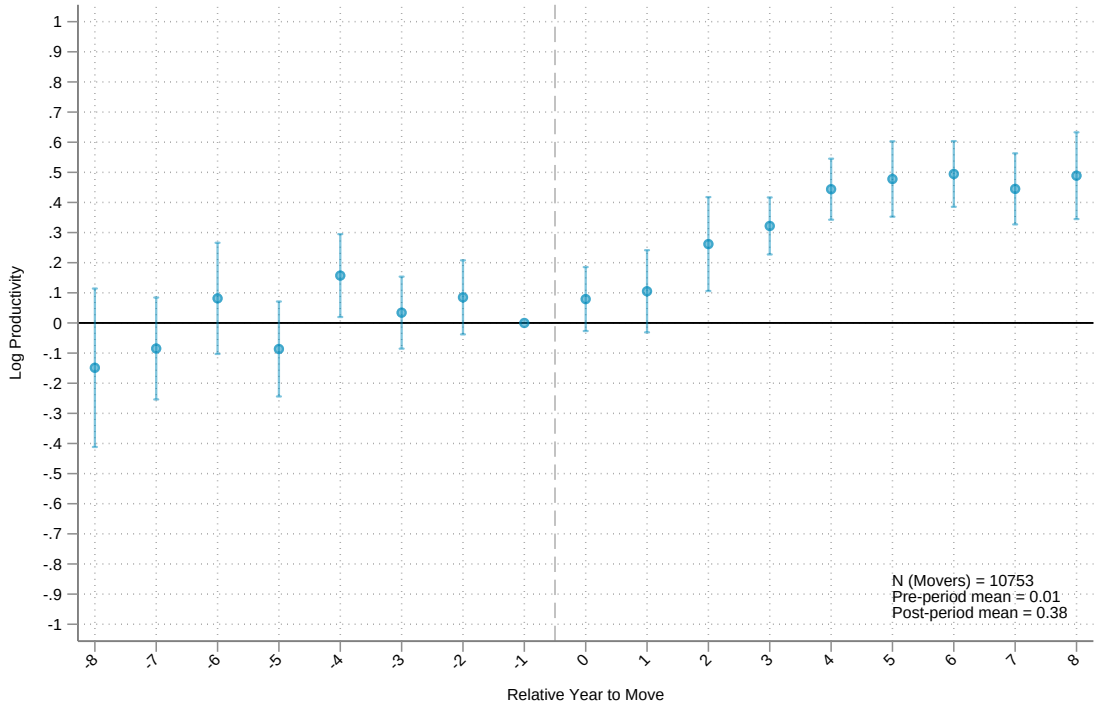

Notes: Figure shows the change in log productivity relative to a mover's productivity move size. Move size is defined as the difference in average produtivity between a mover's origin and destination institution.

## Figure 5: Productivity Change in Response to Move

(a) 1945-2023




(b) 1995-2023

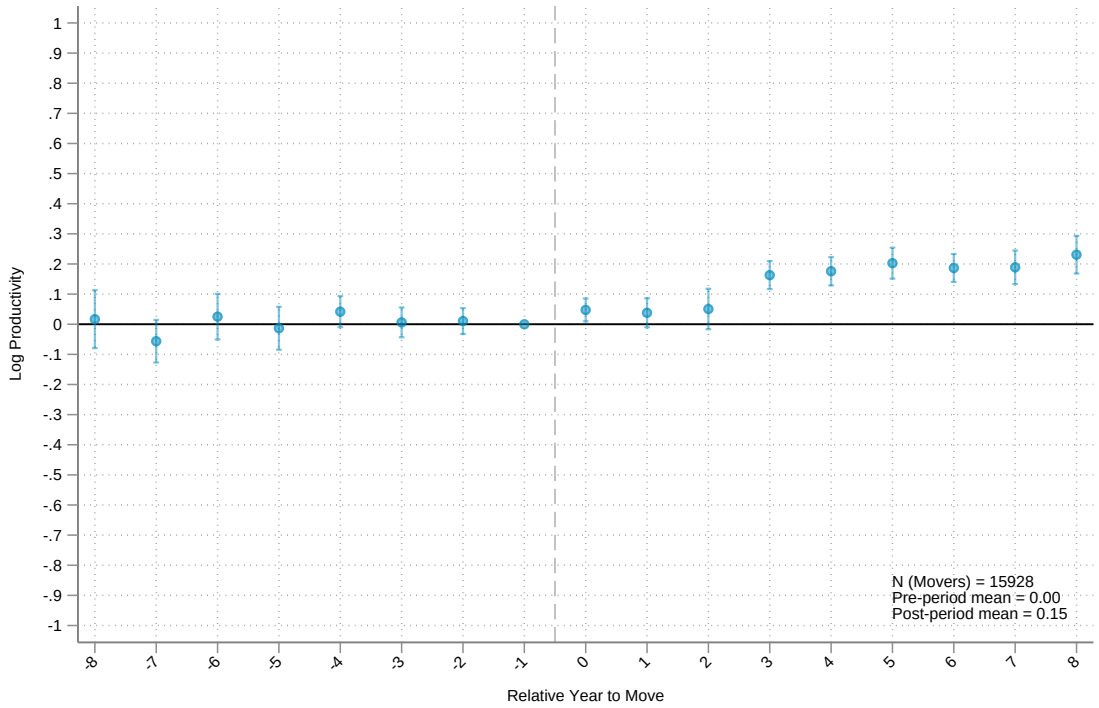




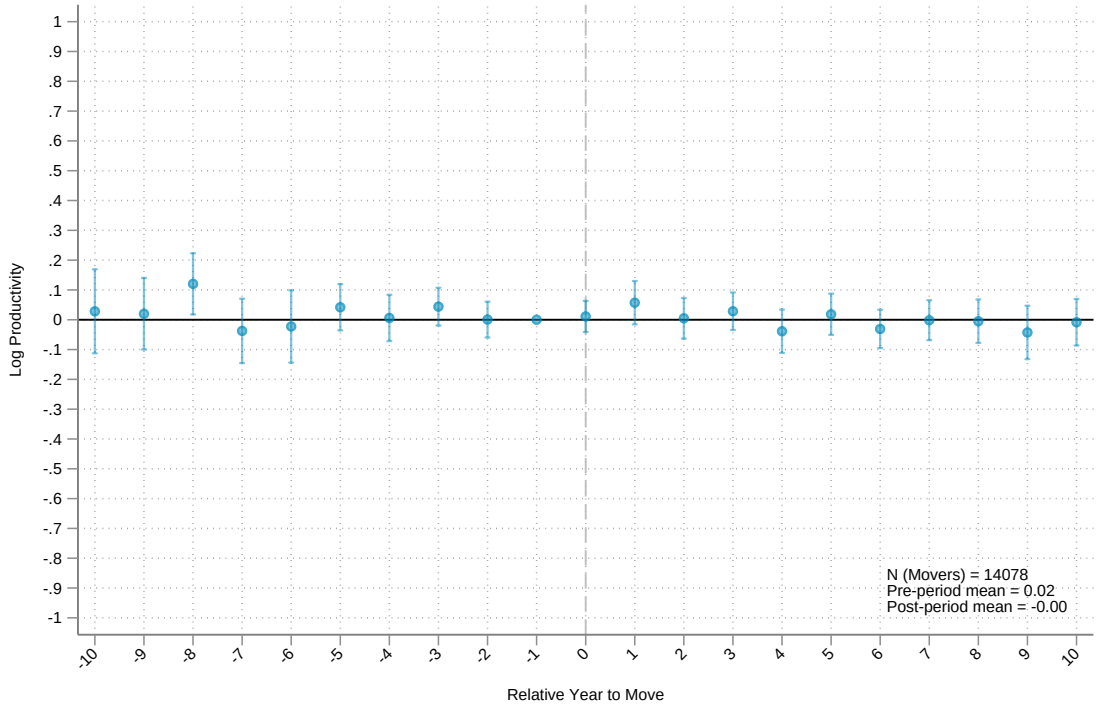

Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2. The coefficient for relative year -1 is normalized to 0. The dependent variable is log productivity and we control for a set of year, field, field x year, and author fixed effects in our specification. Standard errors are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

## Figure 6: Place Effect Mechanisms

### (a) Role of Star Researchers




### (b) Role of Surrounding Institutions in a City




Notes: The first subfigure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers between 1945-2023 using $\delta_i$ defined off the top 95th percentile of researchers in an institution. The second subfigure shows the coefficients $\hat{\theta}$ estimated for movers by the change in their origin and destination cluster size between the time period 1945-2023 using the city-level $\delta_i$ excluding the author's own institution productivity. In both, the dependent variable is log productivity and we control for a set of year, field, field x year, and author fixed effects. Standard errors are clustered at the institution-level.

# Appendix

## A. Additional Exhibits

Appendix Table A1: Journal Sample Sizes

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

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

Appendix Table A2: Measure Weighting Robustness

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

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

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

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

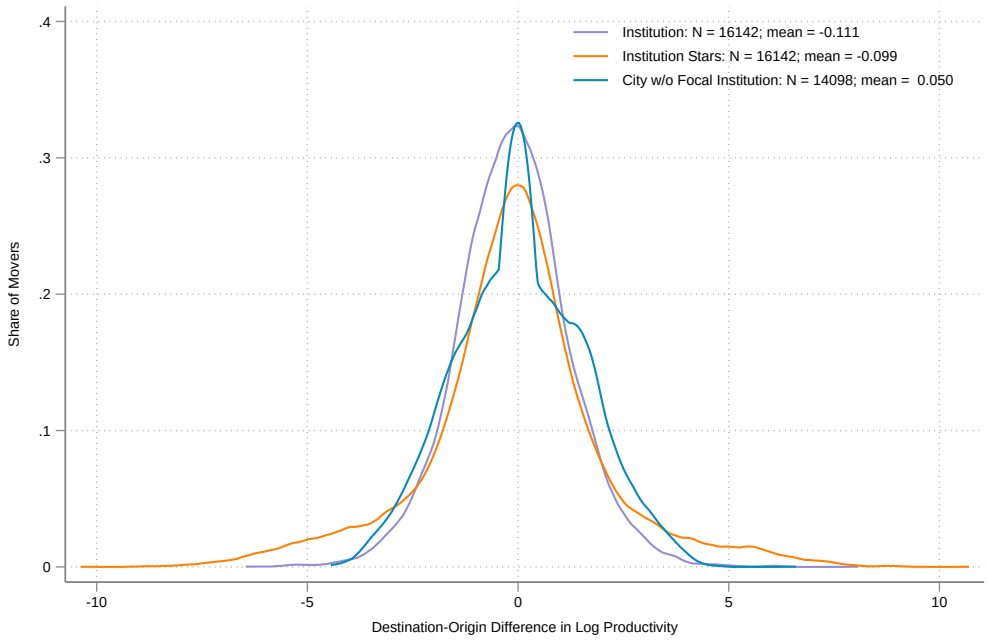
Notes: Table shows the correlation between our main measure of fundamental science productivity and different versions of papers-in-patents citation counts at the state, city, and institution level for articles published in a set of top 15 life-sciences journals *(Cell, Nature, Science, Nature Cell Biology,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

## Appendix Table A4: Top 50 Institutions by Average Researcher Productivity

| Rank | Institution | Researcher Productivity |
|---|---|---|
| 1 | Harvard University | 4.61 |
| 2 | Stanford University | 3.02 |
| 3 | Massachusetts Institute of Technology | 2.80 |
| 4 | National Institutes of Health | 2.09 |
| 5 | University of California, San Francisco | 1.93 |
| 6 | Mass General Brigham | 1.77 |
| 7 | Johns Hopkins University | 1.66 |
| 8 | University of California, San Diego | 1.36 |
| 9 | Columbia University | 1.35 |
| 10 | Yale University | 1.33 |
| 11 | University of California, Berkeley | 1.14 |
| 12 | University of Pennsylvania | 1.05 |
| 13 | University of Washington | 1.00 |
| 14 | Memorial Sloan Kettering Cancer Center | 0.94 |
| 15 | Washington University in St. Louis | 0.87 |
| 16 | Cornell University | 0.87 |
| 17 | The University of Texas Southwestern Medical Center | 0.87 |
| 18 | Dana-Farber Cancer Institute | 0.86 |
| 19 | Rockefeller University | 0.81 |
| 20 | University of California, Los Angeles | 0.79 |
| 21 | New York University | 0.77 |
| 22 | University of Michigan–Ann Arbor | 0.75 |
| 23 | Duke University | 0.67 |
| 24 | The University of Texas MD Anderson Cancer Center | 0.66 |
| 25 | Baylor College of Medicine | 0.60 |
| 26 | Icahn School of Medicine at Mount Sinai | 0.60 |
| 27 | California Institute of Technology | 0.59 |
| 28 | Scripps Research Institute | 0.56 |
| 29 | University of North Carolina at Chapel Hill | 0.55 |
| 30 | University of Chicago | 0.54 |
| 31 | Salk Institute for Biological Studies | 0.53 |
| 32 | Princeton University | 0.51 |
| 33 | New York Genome Center | 0.49 |
| 34 | University of Wisconsin–Madison | 0.44 |
| 35 | Northwestern University | 0.41 |
| 36 | Cold Spring Harbor Laboratory | 0.41 |
| 37 | Boston Children's Hospital | 0.38 |
| 38 | University of Pittsburgh | 0.38 |
| 39 | University of Southern California | 0.35 |
| 40 | Beth Israel Deaconess Medical Center | 0.34 |
| 41 | University of Minnesota | 0.32 |
| 42 | Emory University | 0.31 |
| 43 | Vanderbilt University | 0.30 |
| 44 | University of Massachusetts Chan Medical School | 0.30 |
| 45 | Boston University | 0.30 |
| 46 | St. Jude Children's Research Hospital | 0.27 |
| 47 | Albert Einstein College of Medicine | 0.26 |
| 48 | University of Utah | 0.26 |
| 49 | Fred Hutch Cancer Center | 0.25 |
| 50 | The University of Texas at Austin | 0.24 |

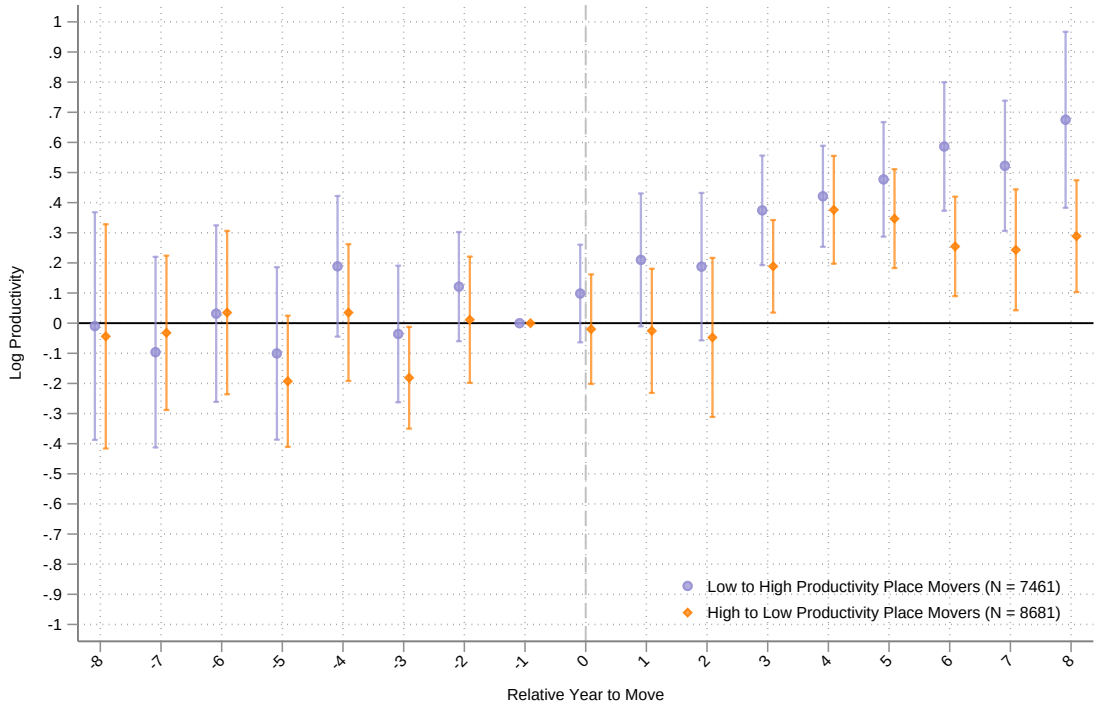
Notes: Table shows the top 50 institutions by average author-year productivity in fundamental science research in life sciences journals. The productivity measure is weighted by the number of researchers at each institution. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature Genetics, 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 Figure A1: Distribution of $\delta_i$



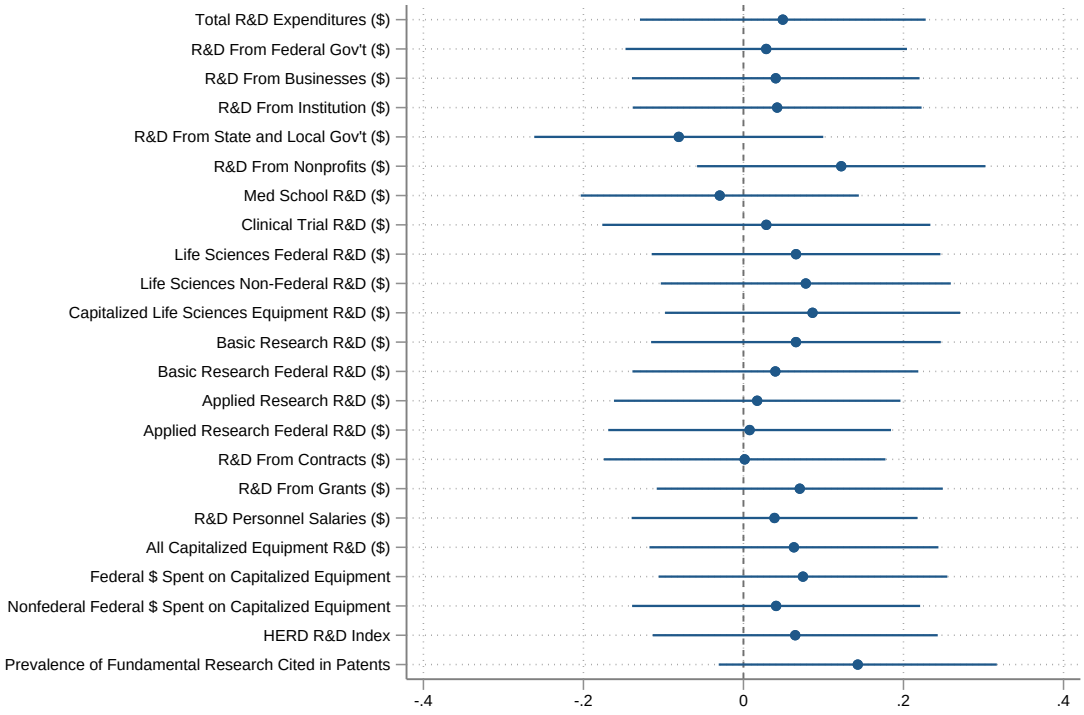

Notes: Figure shows the distribution of $\delta_i$ computed at various levels for scientist-movers. Here, $\delta_i$ measures the size of the move and is constructed as the average difference in productivity between the destination and origin location. We show the distributions for $\delta_i$ at institution-level in purple, for stars at the institution-level in orange, and for cities without the mover's focal institution in blue.

## Appendix Figure A2: 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 A3: 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

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

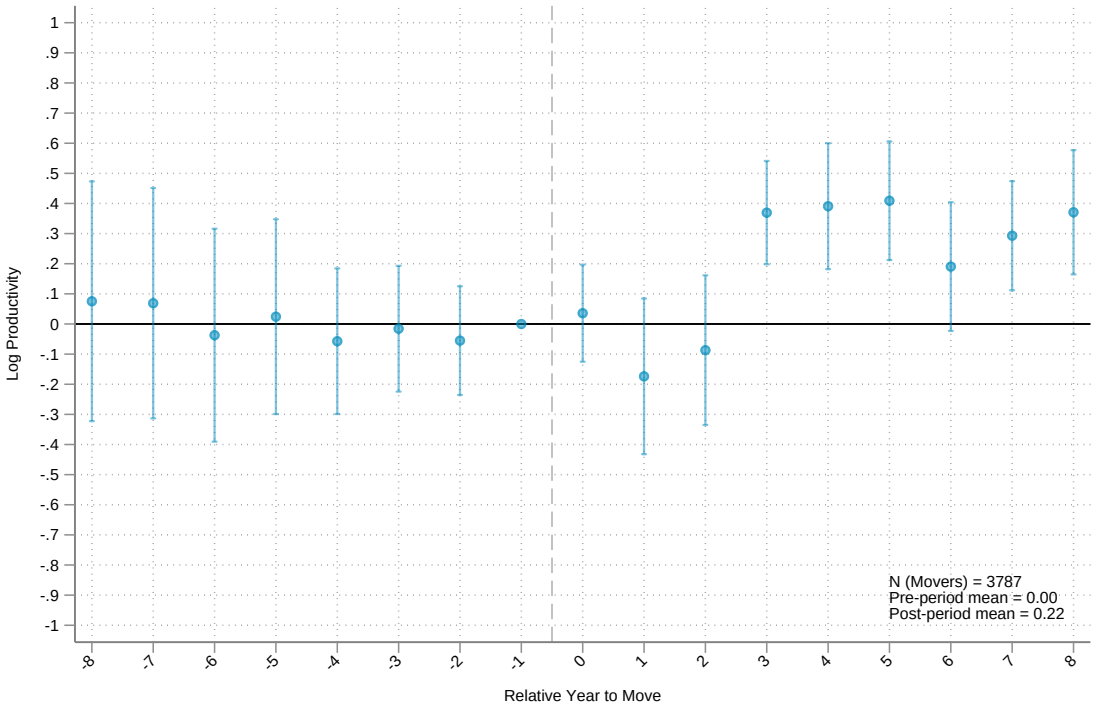
# C. Constructing Life-Sciences Subfields

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

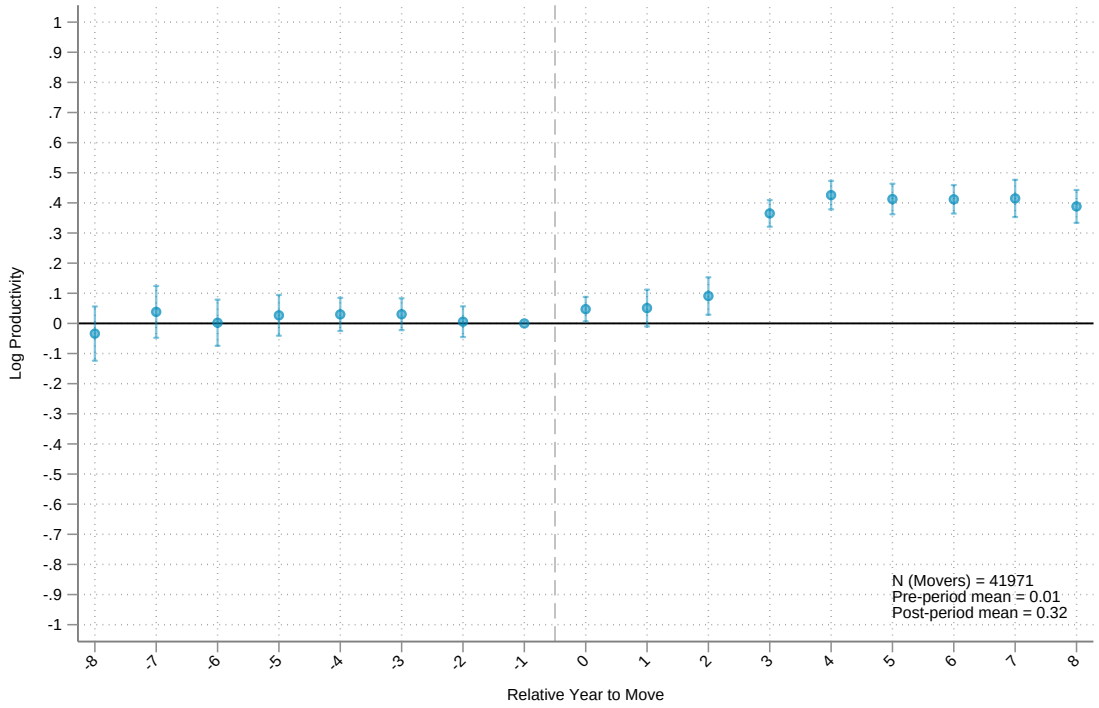
Appendix Figure A4: Productivity Change in Response to Move




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 A5: 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).