# Person and Place Effects in Scientific Discovery *

Amitabh Chandra† Connie Xu‡

March 6, 2026

**Abstract**

How much of a scientist's research output is attributable to individual talent, and how much to the institution where they work? We study this question using publications in leading life-science journals and a movers design that exploits variation in the output of nearly 40,000 scientists who change institutions. We find that 50 to 60 percent of the cross-institutional variation in researcher output is attributable to institutional factors. After correcting for limited mobility bias, we document positive assortative matching between scientists and institutions (a correlation of 0.3 between person and place effects), which is a necessary condition for maximizing discovery. Around two-thirds of the institutional effect is explained by the presence of star researchers. These effects have not diminished despite technologies facilitating remote collaboration and do not increase with the size of the local research cluster beyond the scientist's own institution.

*We thank, without implicating, Alan Garber, George Daley, Kyle Myers, Bitsy Perlman, Bhaven Sampat, Carolyn Stein, Mark Shepard, Josh Angrist, and seminar participants at the Harvard Kennedy School, Harvard Medical School, the NBER, USC, the University of Pennsylvania, the University of Wisconsin–Madison, and the Beth Israel Deaconess for helpful comments.

†Harvard University and NBER, amitabh_chandra@harvard.edu

‡Harvard University, conniexu@g.harvard.edu

# I. Introduction

Fundamental science, the discovery of basic principles that advance human understanding of the natural world, has generated enormous social returns across domains ranging from semiconductors and materials science to agriculture and medicine. Because these discoveries often reveal natural phenomena for which patents are impermissible, or that are too removed from commercial applications to satisfy patentability criteria (Kesselheim et al., 2013; Budish et al., 2015), and because publishing them in journals creates positive externalities that other researchers freely build upon, markets systematically under-invest in fundamental research (Nelson, 1959). The burden of producing this knowledge therefore falls on individual scientists working within universities, government laboratories, and research hospitals that provide equipment, mentors, and students (Furman and Stern, 2011). A central question for science policy is how much of a scientist's research output is attributable to their own talent and how much to their institutional environment, and whether productive scientists are matched to productive institutions in a way that maximizes the production of knowledge.

We examine this question in the life sciences, where breakthroughs such as the discovery of DNA, genes, mRNA, antibodies, and CRISPR have enabled new medicines, improved clinical practice, and deepened our understanding of biology and physiology (Zerhouni, 2005). The life sciences are of interest in their own right, but also offer research advantages over other fields: the complete life-cycle of R&D, from academic research to patents to clinical trials to new drug approvals, is observed, which would not be true in industries that rely on trade secrets or where fundamental knowledge is not published in journals. Using publications in leading life-science journals, weighted by citations, to measure fundamental discoveries, we ask how much of the variation in research output across scientists reflects their individual characteristics and how much reflects the institution where they work, including the presence of talented peers.

The answer to this question has immediate implications for scientific discovery. If institutional environments substantially increase research output, then resources directed to a more productive institution will generate more science than the same resources at a less productive institution, even if the individual scientists at each are equally talented. Moreover, if there is positive assortative matching between scientists and institutions, meaning that the most talented scientists tend to work at the most productive institutions, then this allocation is a necessary condition for maximizing aggregate scientific discovery. The relative role of scientists versus institutions also bears on how the scientific community evaluates individual researchers, particularly in high-stakes decisions such as tenure and promotion. If a large share of a scientist's observed output reflects their institutional environment rather than their individual contribution, then evaluation committees that treat publication records at face value risk over-crediting scientists at strong institutions and under-crediting those who produce at comparable rates despite less favorable environments. Disentangling the individual from the institutional component of research output is therefore important not only for directing funding but also for assessing scientific talent.

Understanding the relative contributions of institutions and individual scientists parallels a core question in labor economics: how much of the variation in worker output reflects individual ability versus the workplace environment and peers? A large literature in personnel economics documents substantial and persistent productivity differences across workers performing the same tasks and

shows that the quality of one's coworkers and supervisors is a first-order determinant of individual output (Mas and Moretti, 2009; Lazear et al., 2015; Sandvik et al., 2020; Hoffman and Stanton, 2024). Our paper asks the analogous question for fundamental science. As in settings studied in labor economics, a key empirical challenge is that the matching between scientists and institutions is endogenous: more talented scientists may sort into better institutions based on personal or professional preferences, such as where they trained, dual-career opportunities, or amenities, making some places appear more productive because of who chooses to work there (Zucker et al., 1998). Our paper addresses these identification challenges by disentangling institution effects from researcher selection and contributes to the literature on the role of institutions in scientific discovery (Allison and Long, 1990; Belenzon and Schankerman, 2013; Azoulay et al., 2011; Hvide and Jones, 2018).

We use publications in leading life-science journals, weighted by citations, to measure fundamental discoveries. The body of work by Azoulay et al. (2010) and coauthors uses citation-weighted publication counts to measure scientific output, demonstrating that these metrics capture meaningful variation in the rate and direction of knowledge creation. A natural concern is that the ability to publish and accumulate citations reflects bias and networks rather than knowledge. Three pieces of evidence weigh against this. First, we show a tight association between producing fundamental science research and receiving citations in patents, which would not hold if publications mostly reflected bias rather than knowledge. Second, if editorial or network advantages accrued disproportionately to prestigious institutions, we would expect a larger institutional share when comparing scientists at the top 5% of institutions to the bottom 5%; the stability of this share across quantiles is consistent with place effects reflecting genuine differences in research environments rather than publication bias. Finally, Card and DellaVigna (2020) show that editors at leading economics journals behave as if they maximize the expected quality of accepted papers and that referees hold papers by more prolific authors to a *higher* standard, implying that famous scientists do not receive preferential treatment in the publication process. Together with the broader literature that relies on impact-weighted citations to measure scientific output (Azoulay et al., 2009; Waldinger, 2016; Azoulay et al., 2019; Lerner et al., 2026), these findings give us reasonable confidence in our measure.

To identify the causal effect of institutions on research output, we implement a movers design, exploiting variation in the research output of scientists who change institutions. This approach, widely used to separate person and place effects in other contexts (Finkelstein et al., 2016; Chetty and Hendren, 2018), was pioneered in the study of academic science by Allison and Long (1990), who tracked job changes among chemists, biologists, physicists, and mathematicians and found that upwardly mobile scientists showed substantial increases in publication rates, an early finding that departmental environments causally affect scientific productivity.[1]

Our contribution is closest to Lerner et al. (2026), who use the same movers design in life sciences and estimate that university-level factors account for one-fifth of the variation across universities in the production of commercially relevant knowledge, as measured by the degree to which patents cite academic research. Their survey of moving scientists also provides reassurance for identification: the modal reasons for moving are better salary, resources, and quality of life, none of

[1]Long and McGinnis (1981) present earlier evidence on the relationship between organizational context and the productivity of biochemists.

which are mechanically related to the destination institution's recent trajectory in research output. Our identification strategy would be threatened if moves were systematically timed to coincide with improving conditions at the destination, and these survey responses suggest they are not. We extend their work in two ways. First, we study a different outcome– scientific discovery. Even when research eventually influences patents, clinical practice, or therapeutics, these impacts often emerge over decades through complex non-linear pathways (Myers and Lanahan, 2022), suggesting that estimates based on direct patent citations provide a conservative measure of institutional effects on fundamental discovery. Second, we estimate the degree of positive assortative matching between scientists and institutions, correcting for the limited mobility bias that attenuates this correlation in standard two-way fixed effects models.

We estimate an event study specification of an individual's research output around job moves, exploiting variation in the "move size," defined as the difference in the average per-scientist research output of the origin and destination institutions.[2] If institutional factors fully determine a scientist's output, then moving to an institution with 20% higher average output should increase a scientist's output by 20%; if research output is entirely an individual-level phenomenon, relocation will have no effect. Since the movers design links changes in an individual scientist's output to changes in the output of their location, we can control for time-invariant individual characteristics and overall time trends, separating institutional effects from the sorting of talented scientists into top institutions and explicitly testing for pre-move trends that would otherwise confound identification. For this analysis, we track nearly 300,000 authors and their research output over the period 1945–2023, including the change in output of those moving (almost 40,000 authors) before and after moving between institutions.

The movers design also allows us to test whether high-output scientists sort into high-output institutions—positive assortative matching (PAM) in the sense of Becker (1973). Under the multiplicative production structure implied by our model, PAM is a necessary condition for maximizing aggregate knowledge production. However, estimating the degree of assortative matching is complicated by limited mobility bias (Andrews et al., 2008): with a finite number of movers per institution, measurement error in the estimated fixed effects biases the person-place correlation *toward zero*, understating the true degree of sorting. We address this using the grouping estimator of Bonhomme et al. (2019), which clusters institutions based on their output distributions before estimation, substantially reducing this bias.

Empirically, we find that 50 to 60 percent of the variation in researcher output across institutions is attributable to institutional factors. After correcting for limited mobility bias, we find a positive correlation (0.3) between researcher and institution effects, confirming that PAM is present in the scientific labor market. Around two-thirds of the institutional effect is explained by the presence of star researchers, defined as scientists at the top 5% of their institution's output distribution, and the infrastructure and students they attract. This aligns with previous research: Waldinger (2016) finds that departments that lost star scientists during Nazi Germany suffered persistent declines in research output, and Azoulay et al. (2019) finds a lasting decline in publication rates for collaborators of deceased stars, a result we replicate. These institutional effects have not diminished over time despite technologies facilitating remote collaboration, do not increase

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

with the size of the local research cluster beyond the scientist's own institution (Gruber et al., 2022; Guzman et al., 2023), and are not concentrated in particular fields. The institutional share increases from 40% among lead authors to 60% when all authors are included, consistent with less-established researchers being more responsive to institutional environments. Consequently, productive institutions in smaller locations not co-located with other large research centers are just as effective at producing fundamental science as equally productive institutions in larger hubs, suggesting that science policy should target institutional strength rather than cluster size. Even across the top 50 institutions, there are fourfold differences in average per-scientist research output, and our estimates imply that over half of these differences are causal.

# II. Measuring Fundamental Science

## A. Definitions and Data

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

To operationalize this definition, we query the Medical Subject Headings (MeSH) assigned to every publication in *PubMed*, a standardized taxonomy for biomedical literature maintained by the NIH National Library of Medicine. We query specifically for the MeSH terms *Anatomy, Chemicals and Drugs, Organisms, Phenomena and Processes, Diseases, Anesthesia and Analgesia,* and *Therapeutics*. The predominant MeSH terms represented in our sample are *receptors, DNA, RNA, transcription factors* and *neurons*, which is consistent with our goal of systematically measuring fundamental science research.

Our dataset consists of journal articles published between 1945 and 2023, excluding comments, case reports, technical notes, letters, and reviews (1945 is the earliest year in which MeSH terms are reliably assigned to articles). We refine the sample 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*.[4] Including

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

[4]We determined this list by asking 5 leading scientists at Harvard Medical School, including the Dean, to list the 10 leading fundamental science journals in the life sciences. There was unanimous agreement on the inclusion of 9 journals including *Cell*, *Nature*, and *Science*, and most disagreement on the inclusion of *PLOS One* and *Oncogene*, which were endorsed by 1 scientist. We exclude clinical journals such as *Journal of the American Medical Association* (JAMA) or the *New England Journal of Medicine* because the NAM definition of fundamental science does not consider clinical studies such as advanced Phase 3 trials. An alternative approach could use the methods in Angrist

journals beyond a narrow set such as *Cell*, *Nature*, and *Science* allows us to generate a sufficiently large sample of movers to precisely estimate place effects. At the same time, this restriction excludes researchers who publish exclusively outside these journals. To the extent that these missing researchers are earlier in their careers and would benefit more from institutional factors, our estimates should be interpreted as a lower bound on the role of institutions in a scientist's research output.[5] This selection process yields a core sample of 563,620 publications, authored by 1,433,962 scientists. The proportion of fundamental science articles in these journals, measured using our definition, ranges from 37% (Science) to 98% (Oncogene), with the remaining articles related to the physical and social sciences (see Appendix Table A1).

After obtaining the relevant set of fundamental science papers from *PubMed*, we gather detailed metadata on the authors of these articles using *OpenAlex*, a comprehensive and openly accessible catalog of the global research system developed by the nonprofit *OurResearch* (Priem et al., 2022).[6] *OpenAlex* indexes over 250 million works by 95 million authors affiliated with 109,000 institutions, aggregating and standardizing information from sources including *PubMed*, Microsoft Academic Graph, *Crossref*, and *ORCID*. It disambiguates authors using an algorithm based on names, publication records, citation patterns, and (where available) ORCID identifiers, and parses raw affiliation strings using a machine learning algorithm to standardize institutional information.

We also use *OpenAlex* to impute the age of fundamental science researchers to understand how institutions affect life-cycle output, identifying the first time a researcher appears in a paper indexed by *OpenAlex* and assigning an age of 25 in that year.[7]

## B. Measuring Scientists' Output

We construct an individual-level measure of research output by quantifying each author's "effective" contribution to a given paper (Azoulay et al., 2010, 2009). We first allocate credit by dividing each paper equally among the number of authors. We then adjust for research impact using paper-specific weights that account for both journal quality and citation counts: we assign weights based on the journal's most recent five-year impact factor from Clarivate (2023) and the average annual citation count since publication. The weighting scheme ensures that the total sum of research contributions across all authors equals the total number of papers in our dataset, so our output measure remains interpretable in units of effective papers. The mathematical details are provided in Appendix Section B, and Appendix Table A2 compares various weighting methods, demonstrating

et al. (2017), who implement a Google PageRank algorithm to assign influence to journals; we do not adopt this methodology because we have already conditioned on a reasonably select group of journals.

[5]We also consider the possibility that results are driven by strategic journal targeting, whereby scientists at higher-ranked institutions disproportionately submit to higher-impact journals. We find this concern to be minimal: the share of papers in top journals remains stable at 43% for movers from lower- to higher-ranked institutions. For movers from higher- to lower-ranked institutions, we observe only a modest decline (43% to 35%).

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

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

that all approaches yield highly correlated results at the country, city, and institution levels.

For our main analysis, we focus on researchers listed among the first two or the last two authors on a publication. In the life sciences, the first authors are typically ordered by contribution, while the last positions usually represent the principal investigator(s). This restriction caps credited authors at four per paper, which helps control for varying coauthorship norms across institutions and fields. This approach likely provides a conservative estimate of place-based effects, since first and last authors tend to be more established researchers who may be less affected by local institutional factors compared to middle authors. In our robustness section, we re-estimate our main specification using both a restricted sample of first and last authors and a broader sample of all listed authors. Consistent with this interpretation, place-based effects increase by approximately 10 percentage points when we include all authors, suggesting that middle authors who may be earlier in their careers are more responsive to institutional characteristics.

To illustrate our approach, consider the paper "*A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity*," published in *Science* in 2012 and cited 13,832 times. We allocate an effective contribution of 0.25 to each of the four credited authors: Martin Jinek (first author) at UC Berkeley, Krzysztof Chylinski (second author) at Umeå University, Jennifer A. Doudna (second to last author) at UC Berkeley, and Emmanuelle Charpentier (last author) at Umeå University. Using the 5-year impact factor of *Science* (54.5) and the average citation count per year since publication ($13{,}832/12 \approx 1{,}152$), we reweight each author's contribution to 28.62 effective articles. For context, Table 2 shows that the average author publishes 0.39 articles annually in our set of journals with a lifetime output of 0.79 effective articles.

## C. Measuring Institutional Output

Identifying the home institution where scientists conduct their research is challenging due to the prevalence of multiple simultaneous affiliations. To address this, we infer each author's principal institution by analyzing their full publication history and identifying the most consistent affiliation for any given year. We maintain hospital affiliations as distinct from their affiliated universities when they are separately branded employers (e.g. Mass General Brigham remains distinct from Harvard).[8] If a unique primary affiliation exists for that year, we assign it. In cases of ties or alternating patterns, we assume institutional continuity. For example, if an author is affiliated with institution A in 2008 and 2010 but not in 2009, we designate institution A as the principal affiliation for all years. If an author lists two affiliations (A and B) in the same year, we pick A if A was the primary affiliation in the previous year; otherwise, we select the one that appears as the primary affiliation in the subsequent year. If a tie remains, we randomly assign one of the institutions, affecting around 5% of author-years.

Having assigned each author to a unique institution for each year, we aggregate output to broader geographic levels such as cities and countries. For areas within the US, we use Metropolitan Statistical Areas (MSAs) and, where appropriate, combine major MSAs into integrated labor

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

markets. For example, we define the Bay Area as the combination of San Francisco-Oakland-Hayward and San Jose-Sunnyvale-Santa Clara. Similarly, we define the Research Triangle Park as a combination of Durham-Chapel Hill and Raleigh-Cary. In the rest of the paper, we refer to these classifications as "cities." We focus on US institutions because our ability to disambiguate US names was better than for other countries; the US also accounts for over 50% of the world's output.

To account for differences in output trends across life-sciences subfields, we categorize each author-year into one of fifteen research fields (e.g. cell signaling, stem cells) within every 20-year period in our sample. This controls for scientists selecting into fields advancing at different rates, such as genetics versus structural biology following the human genome mapping. We define these fields using a k-means textual clustering algorithm applied to paper titles, MeSH terms, and abstracts at the author-year level, separately for each 20-year period so that field definitions evolve over time. Appendix Section C details the tokenized features that define each field. We incorporate these field effects as an alternative regression specification and show that our main results are unchanged in Appendix Figure A4.

# III. Empirical Framework

We now turn to our central question of disentangling place-based effects from the sorting of productive scientists into particular institutions. Place-based components capture local institutional factors such as R&D expenditures, while person-based components capture individual characteristics such as educational background and intrinsic motivation.

We model an individual scientist's log output (that is, the log of citation-weighted publications), $y_{ijt}$, at institution $j$ in year $t$ with a two-way fixed effects specification

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

where $\alpha_i$ 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 include observable researcher characteristics $x_{it}\beta$, in which case the researcher-level component becomes $\alpha_i + x_{it}\beta$ rather than $\alpha_i$ alone. For expositional convenience, we describe the motivating theory with the simplified model, but in alternative specifications, we incorporate research field fixed effects under $x_{it}\beta$. Identification of $\gamma_j$ requires the sample to include scientists who move between institutions. Without movers, $\gamma_j$ would be collinear with the average individual effect $\bar{\alpha}_j$, making it impossible to disentangle whether a high-output institution is productive because of its resources or because of the scientists it hires.

In the above estimating equation, inclusion in the estimating sample is conditional on publication, so a dependent variable with a value of zero reflects the rare case of zero-citation papers in a given year rather than no publications, and occurs in only 0.4% of observations, with over half appearing in the final sample year due to truncation. This minimizes concerns regarding a large probability mass at zero (Mullahy and Norton, 2024; Chen and Roth, 2024).

A causal interpretation of place effects requires several identifying assumptions. First, we assume that the timing of the move is exogenous to unobserved time-varying shocks in researcher

output. For example, this assumption would be violated if "rising stars" systematically relocated to high-output institutions just as their own output was naturally increasing. We assess the validity of this assumption by examining pre-trends in an event study specification described in Section B.

Second, we assume additive separability of $\alpha_i$ and $\gamma_j$ in log output. This functional form implies that in levels, person and place effects are multiplicative: a scientist with person effect $\alpha_i$ at an institution with place effect $\gamma_j$ produces $\exp(\alpha_i+\gamma_j+\tau_t)$ in expectation. The institutional component thus acts as a proportional shifter that scales the research output of all researchers by a common factor. While the *relative gain* from moving to a higher-output institution is uniform across scientists, the *absolute gain* is larger for researchers with higher baseline output. By constraining the institution effect to be constant ($\gamma_j$) across individuals, this assumption rules out idiosyncratic match-specific complementarities ($\gamma_{ij}$) between scientists and institutions.

This multiplicative structure has an important implication for how scientists should be allocated across institutions. Because the exponential function is supermodular in $(\alpha_i, \gamma_j)$, total knowledge production $\sum_i \exp(\alpha_i + \gamma_{j(i)} + \tau_t)$ is maximized when higher-$\alpha$ scientists are matched to higher-$\gamma$ institutions, when the assignment exhibits positive assortative matching (PAM) in the sense of Becker (1973). Any reallocation that breaks this rank-order correspondence strictly reduces aggregate output.[9] Whether this condition holds in practice — that is, whether the scientific labor market sorts talent in a way that maximizes knowledge production — is an empirical question that we address below.

Finally, we assume that the place effects identified by movers are generalizable to non-movers. If this assumption is violated, our estimates would apply only to the set of researchers who relocate rather than the entire population. Table 2 shows that movers tend to publish more and for a longer time than non-movers. Our estimates are a lower bound on the institutional effect if movers are more established and less sensitive to institutional characteristics. On the other hand, if movers are selected on their adaptability, our estimates will overstate the effect of institutions.

## A. Decomposition and Assortative Matching

With the estimated parameters from Equation 1 in hand, we want to answer three questions. First, if we could equalize institutional environments, how much of the observed gap in research output across institutions would disappear? Second, if we could instead equalize the talent of researchers across institutions, how much would disappear? Third, are productive scientists matched to productive institutions in a way that amplifies total knowledge production? The first two questions speak to resource allocation: if institutional environments account for a large share of output differences, then funding directed to more productive institutions generates more science than the same funding at less productive institutions. The third question speaks to the efficiency of the scientific labor market and helps determine whether the current allocation of scientists is close to the knowledge-maximizing assignment described above.

[9] A simple example illustrates this concept. Suppose there are two scientists with person effects $\alpha_H = 2$ and $\alpha_L = 1$ and two institutions with place effects $\gamma_H = 2$ and $\gamma_L = 1$. Under PAM, total output is $\exp(2+2)+\exp(1+1) = e^4+e^2 \approx 62.0$. Under the alternative assignment, total output is $\exp(2 + 1) + \exp(1 + 2) = 2e^3 \approx 40.2$. The assortative allocation produces over 50% more total knowledge.

We address the first two questions with two decomposition methods. The *additive decomposition* focuses on the output gap between specific pairs of institutions. Let $\overline{y}_j$ denote the time-averaged mean of $y_{it}$ across researchers at institution $j$, and $\overline{\alpha}_j$ the corresponding average of person effects. The difference in average log output between institutions $j$ and $j'$ is

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

so the institutional share of the output gap is $S_{institution}(j, j') = (\gamma_j - \gamma_{j'})/(\overline{y}_j - \overline{y}_{j'})$ and the researcher share is $S_{researcher}(j, j') = (\overline{\alpha}_j - \overline{\alpha}_{j'})/(\overline{y}_j - \overline{y}_{j'})$. Under additive separability, $S_{researcher}(j, j')$ is the proportion by which the output gap would diminish if researchers were randomly reallocated between the two institutions. In Section D, we apply this decomposition at various quantiles of the institutional output distribution. It also doubles as a test of whether editorial or network advantages accrue disproportionately to the most prestigious institutions, for this would result in a larger institutional share when comparing the top 5% of institutions to the bottom 5%; stability across quantiles, which is what we find, is consistent with place effects reflecting genuine differences in research environments rather than publishing outcomes that reflect a preference for scientists at elite institutions.

The *variance decomposition* takes a broader view, asking what share of the cross-institution variance in average log output is attributable to each component. If all place-specific factors were equalized, the remaining variation would reflect only compositional differences in researcher talent, so the institutional share is $S^{var}_{institution} = 1 - Var(\bar{\alpha}_j)/Var(\bar{y}_j)$. Symmetrically, the researcher share is $S^{var}_{researcher} = 1 - Var(\bar{\gamma}_j)/Var(\bar{y}_j)$. Unlike the additive decomposition, these two shares need not sum to one. The gap between their sum and one is governed by the covariance between $\bar{\alpha}_j$ and $\gamma_j$: when the covariance is positive, productive scientists are sorting into productive institutions and the shares sum to more than one; when it is zero, there is no systematic sorting (Card et al., 2013).

This covariance is the empirical counterpart of PAM. Its sign tells us whether the scientific labor market allocates talent in the direction that maximizes aggregate knowledge production. A positive correlation between $\hat{\alpha}_i$ and $\hat{\gamma}_{j(i)}$ indicates sorting toward the optimum; a zero or negative correlation would imply that reallocation could increase total output at no additional cost.

Estimating this correlation from a two-way fixed effects model is complicated by limited mobility bias (Andrews et al., 2008). With a finite number of movers per institution, sampling error in the estimated fixed effects generates a mechanical negative covariance between $\hat{\alpha}_i$ and $\hat{\gamma}_{j(i)}$: any estimation error that inflates a particular institution's fixed effect simultaneously deflates the person effects of scientists observed there, biasing the measured correlation toward zero. This bias is especially severe when institutions have few movers, precisely the setting where fixed effects are least precisely estimated.[10]

To correct for this, we follow Bonhomme et al. (2019) and group our set of over 4,000 institutions into 100 clusters based on the empirical distribution of researcher output before estimating the two-way fixed effects model. By averaging across many institutions within each cluster, the grouping estimator reduces the noise in estimated place effects and recovers a bias-corrected

[10]While researcher fixed effects also contain measurement error, this noise inflates researcher variance and biases the resulting institution share downward, making our estimates a lower bound.

estimate of the person-place correlation. We explore the sensitivity of this estimate to alternative bias-correction methods, including Empirical Bayes shrinkage and sample restrictions based on the number of movers per institution, in Appendix Section D.

## B. Event Study Design

The decompositions above quantify the relative importance of institutions and researchers but rely on cross-sectional variation. We complement them with an event study that exploits within-researcher variation over time, which allows us to control for time-invariant individual characteristics through person fixed effects and to directly inspect pre-trends that would otherwise confound identification. We view this as our preferred specification.

We first scale each mover's log output to account for differing output levels across origin and destination institutions. Without scaling, moves from $j$ to $j'$ would be canceled out by moves from $j'$ to $j$. We define mover $i$'s "move size" as the difference in average log output between institution $d(i)$ (destination) and institution $o(i)$ (origin), excluding mover $i$:

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

Here, $\overline{y}_{d(-i)}$ averages across all author-years (including non-movers but excluding mover $i$) at the destination, and $\overline{y}_{o(-i)}$ is computed analogously for the origin. Since we restrict to researchers with one move, each mover is associated with a single value of $\delta_i$. Institution averages are computed across all author-years, so researchers with the same destination-origin pairing share the same $\delta_i$.[11]

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

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

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

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

the coefficients $\theta_{r(i,t)}$ represent the extent to which researchers have converged toward their destination's productivity by relative year $r(i,t) = t - t_i^*$, where $t_i^*$ is the move year. Thus $r(i,t) = -1$ is the last year at the origin and $r(i,t) = 0$ is the first year at the destination.

In practice, because $\delta_i$ is often close to zero (Figure 4a), we do not estimate the scaled model directly. Rearranging to avoid dividing by $\delta_i$ and including fixed effect controls, our baseline event study specification is:

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

[11] $\delta_i$ is time-invariant, calculated as the difference in the average outcome across all periods between the destination and origin. If place averages were calculated for every period to create a time-varying $\delta_{it}$, this would likely introduce endogeneity: such a measure could capture contemporaneous shocks correlated with both the timing of the move and the individual's outcome, rather than reflecting the stable difference in place effects.

where $\tau_t$ are calendar year fixed effects, $\rho_{r(i,t)}$ are relative year fixed effects, and $\widetilde{\alpha}_i$ combines the origin institution mean $\overline{y_{o(i)}}$ and the individual effect $\alpha_i$ into a single fixed effect. The relative year fixed effects allow for a common output trend across all movers regardless of the size and direction of the move, capturing life-cycle patterns such as a decline in output immediately following a move or a burst preceding one. In alternative specifications, we include field-by-year fixed effects to account for differential output trends across research fields.

The key parameters are the relative year coefficients $\theta_{r(i,t)}$, normalized to zero at $r(i,t) = -1$. These measure the change in log output around the move, scaled by move size $\delta_i$. We expect $\theta_{r(i,t)}$ to be indistinguishable from zero for $r(i,t) < 0$ (no pre-trends) and non-zero following a move. A post-move value of $\theta_{r(i,t)} = 1$ indicates complete convergence to the destination's average—output is entirely place-determined; $\theta_{r(i,t)} = 0$ indicates no convergence and no role for place. We report robust standard errors clustered at the institution level.[12]

# IV. Evidence on Person and Place Effects

## A. Descriptive Facts on Institutional Variation

We illustrate the variation in institutional effects before controlling for researcher composition in Table 1, which reports the top institutions by average per-researcher output in the ten cities with the largest share of fundamental science output. Our measure of research output is interpreted as the "effective" number of publications for an author. The average researcher at Dana-Farber Cancer Institute, for example, effectively produced four times as many publications per year as the average researcher at Lawrence Berkeley National Laboratory.[13] While elite institutions like Harvard and Stanford dominate worldwide scientific production because of their size, many smaller institutions house highly productive researchers, including universities, research organizations, government agencies, and biotechnology companies. Figure 1 illustrates the persistence of these places in producing a large share of fundamental research output.[14]

[12] If there is classical measurement error in $\delta_i$, as there likely is for smaller institutions, this biases our estimates toward zero. As a robustness check, we estimate our main specification on "larger institutions" (at least 100 authors and 10 movers) and find that estimated place effects increase to 60–70%.

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

[14] Our focus is on US institutions, but the analysis could in principle be extended to other countries. Appendix Tables A4 and A5 provide the top 50 institutions by average researcher output for US and non-US institutions, respectively. These tables reinforce the distinction between aggregate research output, which is dominated by a small number of institutions, and researcher-level output, which is distributed across a diverse set of institutions. Several international institutions such as the Wellcome Sanger Institute, several Dutch hospitals and universities, and the Weizmann Institute of Science in Israel have per-capita output comparable to leading US institutions.

## B. Commercial Relevance of Fundamental Science

We next establish the correlation between fundamental science research and downstream commercial activity. Building on work by Bryan et al. (2020) and Marx and Fuegi (2020), we measure the propensity of fundamental science research to be cited in patents. We construct a parallel measure of "papers in patents" that replaces citation weights with the number of times each paper is cited in patent body text. 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, which prior research finds are more likely to reflect inventors' actual knowledge base; front-page citations, by contrast, are typically added by patent attorneys and are less informative about knowledge flows (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 institution-level relationship between fundamental science research output and "papers in patents" output in Figure 2. The slope is 0.86, indicating that cities producing more fundamental science research also produce research that is more frequently cited in patents.[15] This underscores the commercial and welfare implications of fundamental science research. At the same time, our analysis does not depend on this correlation. Fundamental science has value regardless of its commercial relevance.[16]

## C. Characteristics of Scientist Movers

Using the sample of authors described above, we now focus on scientists who moved between institutions (this is the principal reason we considered a broader set of journals, because it enlarges the pool of movers). We define a scientist to be a "mover" if they are affiliated with different institutions in consecutive years. Since this geographic panel is constructed from publication history, publication lags could affect the accuracy of identifying 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. Under the assumption that coauthorship reflects physical location, the number of coauthors from the destination institution should not increase prior to the move. We find that movers begin accumulating coauthors from their destination institution around three years before their first publication officially affiliated with the new institution. We therefore define the move year as three years before the first observed publication with the destination institution.[17] Figure 3 illustrates the effect of this adjustment, showing that the share of coauthors from the destination institution only increases after our newly defined move year.

We validate this imputed move date using a hand-audited random sample of 100 star movers. Of these, we located detailed employment histories for 77 authors using CVs and faculty bios and accurately identified the origin and destination for 55 (a 71% accuracy rate). Among these 55

[15] We see a similar result at the city-level with a correlation of 0.93.

[16] We also examined the sensitivity of our results to using front-page vs. patent body citations. Regardless of the chosen measure, Appendix Table A3 shows a strong correlation between producing fundamental science research and having it cited in patents at the state, city, and institution level.

[17] As a robustness check, we restrict our analysis to movers observed at distinct universities in two consecutive years and find that our main results remain qualitatively similar.

correctly identified movers, we accurately impute the move within one year 45% of the time, and within two years 80% of the time. Our imputed dates tend to be slightly conservative, coding the move one or two years earlier than the actual date on average. The remaining 20% have larger errors, often due to complex career paths or extended dual affiliations. Measurement error in move timing will attenuate the sharpness of the event study discontinuity, biasing our post-move estimates toward zero. Our estimates should therefore be interpreted as lower bounds on the true institutional effect.

Table 2 provides summary statistics for US-based movers and non-movers. Given the US's large role in fundamental science research, we do not view our restriction to US-based scientists as a major limitation (in our original sample of 563,620 papers, 294,782 were authored by 331,846 US-based researchers). Among US-based researchers, we identify 37,809 scientist-movers, of whom 25,845 move exactly once. We focus on one-time movers to avoid complications from multiple treatments and to minimize potential noise from author disambiguation and affiliation imputation. Movers tend to come from slightly smaller cities and produce more output on average than non-movers. These one-time movers represent about two-thirds of the moving sample and are representative of the larger moving population.

Figure 4a presents the distribution of our key independent variable, $\hat{\delta}_i$, the difference in average log output between a mover's destination and origin institution. The distribution has a standard deviation of 0.77, with a substantial number of large moves in either direction. It is centered around $-0.091$ (meaning the typical move is to an institution whose output is approximately 9 percent lower than the origin) and is relatively symmetric, indicating an approximately equal number of moves to higher- and lower-output institutions. Institutions with lower per-capita output can offer higher salaries and amenities, and the symmetry of the distribution in Figure 4a confirms that moves are not concentrated toward more productive institutions. To contextualize this average move size, the average mover in our sample is moving from the equivalent of the University of Chicago to Baylor College of Medicine.[18]

## D. Estimates from Two-Way Fixed Effects

We estimate the two-way fixed effects model in Equation 1 on our sample of movers and non-movers. As discussed in Section A, the raw fixed effects suffer from limited mobility bias, so we apply the grouping estimator of Bonhomme et al. (2019) throughout. Appendix Section D explores alternative bias-correction methods.

Panel A of Table 3 presents the variance decomposition. Equalizing institutional effects would eliminate 84% of the observed cross-institution variance in research output. Equalizing researcher effects would eliminate 61%.

These two shares sum to more than 100%, confirming the presence of positive assortative

[18]We define move size as the difference in average log output of two institutions, which is not the same as the log of the ratio of average output in levels. Table A4 shows average research output in levels; logging these values produces the log of an average and will not equal move size. This table has been restricted to institutions with more than 100 researchers and 10 movers between 2015 and 2023 to provide a clearer sense of the variation driving our estimates. When estimating regressions, we include all institutions.

matching. After bias correction, the correlation between researcher and institution effects is 0.3, indicating that the scientific labor market sorts productive scientists into productive institutions. As shown in Section A, this is a necessary condition for maximizing aggregate knowledge production under the multiplicative structure of our model.

Panel B of Table 3 presents the additive decomposition, examining whether the relative importance of place varies across the distribution of institutional quality. This decomposition analyzes the output gap between top and bottom institutions at various thresholds (top and bottom 5%, 10%, 25%, and 50%). For each threshold, we decompose the total output gap ($\hat{y}_{top} - \hat{y}_{bottom}$) into a component attributable to researcher composition ($\hat{\alpha}_{top} - \hat{\alpha}_{bottom}$) and a component attributable to place effects ($\hat{\gamma}_{top} - \hat{\gamma}_{bottom}$).

Approximately 73% of the output gap between top and bottom institutions is driven by place characteristics, with the remaining 27% attributable to differences in researcher composition.[19] The stability of this share across partitions implies that top institutions do not rely disproportionately on talent sorting to generate their output. Place-specific factors uniformly dominate across the entire distribution of institutional quality.

The two decompositions yield institutional shares of 84% and 73%, respectively. Both are larger than the event study estimates we present next (50–60%). We view the event study as our preferred specification because it isolates the within-researcher change in output following a move, as discussed in Section B.

## E. Estimates from Event Study

As an initial exploration, we plot the post-move change in scientist log output against the size of the move in Figure 4b. We define a scientist's change in log output as their average post-move log output minus their average pre-move log output. If place fully explained output differences, the slope would be 1; if individual factors fully explained the variation, the slope would be 0. The figure shows a slope of 0.75, suggesting that 75% of the variation can be attributed to institutional factors. The linear and symmetric relationship above and below a move size of zero supports our assumption of additive separability between $\alpha_i$ and $\gamma_j$ in Equation 1: moves to more productive institutions are quantitatively similar, but with opposite sign, to moves to less productive institutions.

Figure 5a presents the event study estimates from Equation 2 for the period 1945–2023, plotting $\theta_{r(i,t)}$ with 95% confidence intervals and normalizing $r(i,t) = -1$ to zero. Standard errors are clustered at the institution level.[20] The figure shows no significant pre-trend in $\theta_{r(i,t)}$ in the ten years before relocation, supporting our identifying assumption that unobserved person-based output factors do not drive the decision to move. Log output remains stable until about two years after the move, at which point it increases to approximately 0.5, peaking at around 0.6. These

[19] We run the same set of results using a non-citation-weighted measure of research output and find similar results. The correlation between institutional fixed effects from the two measures is 0.916.

[20] The number of movers declines slightly toward relative year −10, which explains the somewhat larger confidence intervals in that period. We do not observe a substantial decline at relative year 10, which explains the relatively stable confidence interval widths there.

findings suggest that institutional factors account for roughly 50–60% of the cross-institutional variation in research output.[21] This estimate is lower than the 75% from Figure 4b because the event study controls for time-invariant individual characteristics through person fixed effects.

Figure 5b shows the same analysis for the more recent 1995–2023 period. The similarity of estimates across the full and recent samples suggests that technologies facilitating remote collaboration have not substantially diminished institutional effects, though we cannot rule out more gradual changes.

## F. Heterogeneity by Move Direction and Researcher Characteristics

To understand the mechanisms through which place influences research output, we examine whether our effects vary by the type of move and the characteristics of the researcher.

We first explore the role of move direction, categorizing moves as positive ($\hat{\delta}_i > 0$) if a researcher moves to a higher-output institution and negative ($\hat{\delta}_i < 0$) otherwise. Higher-output institutions likely have more resources and facilities that enable scientists to work at the frontier. If researchers incorporate these benefits when moving to higher-output institutions but do not unlearn skills when moving to lower-output ones, we would expect larger place effects for positive moves. Appendix Figure A1 separates the estimated effect by direction. The point estimates are slightly higher for scientists moving to higher-output institutions, but the differences are not statistically significant.

Given the complementary research on the role of "stars" in discovery (Azoulay et al., 2012, 2019; Waldinger, 2016), we examine the influence of star researchers in explaining our place effect estimates. We define "stars" as researchers in the top 5 percent of their institution's output distribution and modify $\delta_i$ to represent the difference between the average output of stars in the destination and origin institutions. Figure 4a illustrates the distribution of this revised move size, while Figure **??** presents the estimated place effect. The results indicate that stars significantly contribute to the output of movers, with the estimated place effect around 0.33. Star scientists, together with the research laboratories and students they attract, explain roughly two-thirds of the baseline place effect. Panel A of Table A6 compares the average post-period estimates between our main specification and the star-based definition of move size.

Motivated by prior research on the life-cycle output of researchers and agglomeration effects (Dietz and Bozeman, 2005; Jones, 2010, 2009; Levin and Stephan, 1991; Myers, 2020; Carlino and Kerr, 2015; Kerr and Robert-Nicoud, 2020; Moretti, 2021), we examine heterogeneity by mover age and city size. Figures A2 and A3 show similar magnitudes of place effects regardless of whether researchers move before or after the age of 40, or whether they relocate to larger or smaller cities as measured by the number of scientists excluding their own institution. Neither age nor the size of the local scientific agglomeration significantly affects the role of place in changing researcher output.

Finally, Table A6 (Panel B) reports event-study elasticities. A 1% increase in non-federal life-sciences funding is associated with a 0.08% increase in a scientist's output; the corresponding

[21] We run the same specification using research output without citation weighting and obtain similar results.

elasticities with respect to federal and total life-sciences funding are 0.11 and 0.10, respectively. With respect to an institution's total output, a 1% increase is associated with a 0.13% increase in individual output and a 0.13 increase in unique US coauthors. These small magnitudes suggest that funding and collaboration networks alone may not fully explain the institutional effects we estimate, though the imprecision of these estimates warrants caution.

## G. Robustness

### G.1. Alternative Movers Design Specification

We extend our baseline event study to incorporate field-specific fixed effects, accounting for the possibility that researchers select into subfields with rising output or benefit from field-level breakthroughs through knowledge spillovers. The results, presented in Appendix Figure A4, demonstrate that our core findings are robust to the inclusion of field effects. We also show in Appendix Figure A4 that removing the relative year fixed effects $\rho_{r(i,t)}$ from our main specification has little effect on the results.

### G.2. Alternative Author Samples

We replicate our analysis for (1) only first and last authors and (2) all authors. For first and last authors, the estimated institutional share is slightly smaller, around 40% (Appendix Figure A5). When we expand to all listed authors, the institutional share increases to approximately 60%. The gradient across author positions, 40% for lead authors, 50% for our primary sample, and 60% for all authors, is consistent with less-established researchers being more responsive to institutional environments.

### G.3. Publication Bias

A natural concern is that observed changes in output reflect editorial preferences for authors at certain institutions rather than changes in knowledge production. As shown in Panel B of Table 3, the institutional share of the top-bottom output gap is stable across quantiles, which is inconsistent with publication bias concentrated at elite institutions. We also re-estimate our outcomes without citation-weighting research output. If publication bias were the primary mechanism, estimated place effects should shrink when citations are removed. Appendix Figure A6 shows that the magnitudes are essentially unchanged, indicating that citation accumulation based on institutional prestige is not a first-order driver.

# V. Conclusion

By tracking the research output of nearly 40,000 scientists who move between institutions, we estimate that 50 to 60% of the cross-institutional variation in researcher output is attributable to place-specific factors. The remaining 40 to 50% reflect the sorting of productive scientists into

productive institutions: after correcting for limited mobility bias, we find a positive correlation of 0.3 between person and place effects, confirming that positive assortative matching is present in the scientific labor market. Under the multiplicative production structure implied by our model, this sorting pattern is a necessary condition for maximizing aggregate knowledge production. Around two-thirds of the institutional effect is explained by the presence of star researchers and the infrastructure and students they attract, consistent with Waldinger (2016), who finds that the dismissal of star scientists in Nazi Germany led to persistent declines in departmental output, and Azoulay et al. (2019), who finds lasting declines in the output of a star's collaborators following the star's death. These institutional effects have not diminished over time despite technologies facilitating remote collaboration, do not increase with the size of the local research cluster beyond the scientist's own institution, and are not concentrated in particular fields.

These findings have direct implications for the two policy questions raised in our introduction. First, regarding funding allocation: if institutional environments causally increase research output, then resources directed to a more productive institution will generate more science than the same resources at a less productive institution, even holding researcher talent constant. The potential magnitudes are large. Appendix Tables A4 and A5 show 3x–4x variation in average researcher output even among the most productive institutions, a truncated portion of the true distribution, and our estimates imply that over half of this variation is causal. Second, regarding the evaluation of individual scientists: because a substantial share of observed output reflects the institutional environment, tenure and promotion committees that treat publication records at face value risk over-crediting scientists at strong institutions and under-crediting those who produce at comparable rates in less favorable environments.

Our results also clarify the role of agglomeration. Productive institutions in smaller locations not co-located with other large research centers are just as effective at producing fundamental science as equally productive institutions in larger hubs, echoing a point made by Ellison et al. (1997) about manufacturing agglomerations. This suggests that place-based science policy should target institutional strength rather than cluster size. At the same time, investing in the resources that attract and retain star researchers may be one path to increasing institutional output, though such investments are likely to require substantial resources (Myers, 2020).

Our work complements the broader literature on the production function of innovation (Lach and Schankerman, 2008; Azoulay et al., 2013; Bloom et al., 2020; Myers, 2020; Arora et al., 2021; Acemoglu, 2023; Hill and Stein, 2023, 2024; Tham et al., 2024) and on the sources of patenting, a downstream outcome of fundamental science (Jaffe, 1989; Trajtenberg et al., 1997; Belenzon and Schankerman, 2013; Williams, 2017; Bell et al., 2019; Hausman, 2022; Moretti et al., 2023; Lerner et al., 2026). A key question we leave open is causally identifying which institutional characteristics make some institutions more productive than others. Our finding that R&D expenditures correlate with institutional effects validates the estimates but does not establish a mechanism. Understanding how institutions structure their research environments, how funding translates into output (Myers and Lanahan, 2022), and whether early-career interventions at productive institutions yield compounding returns (Azoulay et al., 2011, 2019, 2021) would bring us closer to an answer with direct consequences for how societies organize and fund the production of knowledge.

# References


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

Allison, P. D. and J. S. Long (1990). Departmental Effects on Scientific Productivity. *American Sociological Review 55*(4), 469–478.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008). High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias? *Journal of the Royal Statistical Society. Series A (Statistics in Society) 171*(3), 673–697.

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. F. Lu (2017, May). Economic Research Evolves: Fields and Styles. *American Economic Review 107*(5), 293–297.

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., W. Ding, and T. Stuart (2009, December). THE IMPACT OF ACADEMIC PATENTING ON THE RATE, QUALITY AND DIRECTION OF (PUBLIC) RESEARCH OUTPUT. *The Journal of Industrial Economics 57*(4), 637–676.

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.

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.

Azoulay, P., J. S. Graff Zivin, and J. Wang (2010, May). Superstar Extinction. *The Quarterly Journal of Economics 125*(2), 549–589.

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

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy 81*(4), 813–846.

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

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

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

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica 87*(3), 699–739. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA15722.

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

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

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

Card, D. and S. DellaVigna (2020, March). What Do Editors Maximize? Evidence from Four Economics Journals. *The Review of Economics and Statistics 102*(1), 195–217.

Card, D., J. Heining, and P. Kline (2013, August). Workplace Heterogeneity and the Rise of West German Wage Inequality*. *The Quarterly Journal of Economics 128*(3), 967–1015.

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

Chen, J. and J. Roth (2024, May). Logs with Zeros? Some Problems and Solutions*. *The Quarterly Journal of Economics 139*(2), 891–936.

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

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

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

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

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

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

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

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

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

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

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

Hoffman, M. and C. T. Stanton (2024, August). People, Practices, and Productivity: A Review of New Advances in Personnel Economics.

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

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

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.

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

Lazear, E. P., K. L. Shaw, and C. T. Stanton (2015, October). The Value of Bosses. *Journal of Labor Economics 33*(4), 823–861.

Lerner, J., H. J. Manley, C. Stein, and H. L. Williams (2026). The Wandering Scholars: Understanding the Heterogeneity of University Commercialization. *Econometrica: Journal of the Econometric Society*.

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.

Long, J. S. and R. McGinnis (1981). Organizational Context and Scientific Productivity. *American Sociological Review 46*(4), 422–442.

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.

Mas, A. and E. Moretti (2009, March). Peers at Work. *American Economic Review 99*(1), 112–145.

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.

Morris, C. N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association 78*(381), 47–55.

Mullahy, J. and E. C. Norton (2024). Why Transform Y? The Pitfalls of Transformed Regressions with a Mass at Zero. *Oxford Bulletin of Economics and Statistics 86*(2), 417–447. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/obes.12583.

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.

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].

Sandvik, J. J., R. E. Saouma, N. T. Seegert, and C. T. Stanton (2020, August). Workplace Knowledge Flows*. *The Quarterly Journal of Economics 135*(3), 1635–1680.

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

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

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

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

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.

Table 1: Leading Institutions in Ten Cities with Largest Fundamental Science Output, 2015-2023

| Rank | Institution | Average Researcher Output | Share of World Output (%) |
|---|---|---|---|
| **I. Boston-Cambridge-Newton, MA-NH** | | | **9.79** |
| 1. | Dana-Farber Cancer Institute | 2.81 | 1.01 |
| 2. | Massachusetts Institute of Technology | 2.33 | 2.09 |
| 3. | Harvard University | 1.64 | 3.63 |
| 4. | Mass General Brigham | 1.52 | 1.56 |
| 5. | Boston Children's Hospital | 1.15 | 0.39 |
| **II. Bay Area, CA** | | | **7.67** |
| 1. | Stanford University | 1.86 | 3.38 |
| 2. | Gladstone Institutes | 1.77 | 0.16 |
| 3. | University of California, Berkeley | 1.33 | 1.00 |
| 4. | University of California, San Francisco | 1.31 | 1.87 |
| 5. | Lawrence Berkeley National Laboratory | 0.67 | 0.07 |
| **III. Trenton, NJ** | | | **0.54** |
| 1. | Princeton University | 1.42 | 0.51 |
| **IV. Dallas-Fort Worth-Arlington, TX** | | | **0.86** |
| 1. | The University of Texas Southwestern Medical Center | 1.35 | 0.83 |
| **V. New York-Newark-Jersey City, NY-NJ-PA** | | | **6.67** |
| 1. | Memorial Sloan Kettering Cancer Center | 2.24 | 1.20 |
| 2. | Novartis (United States) | 1.72 | 0.12 |
| 3. | Icahn School of Medicine at Mount Sinai | 1.69 | 0.90 |
| 4. | Cold Spring Harbor Laboratory | 1.42 | 0.22 |
| 5. | Rockefeller University | 1.34 | 0.60 |
| **VI. San Diego-Carlsbad, CA** | | | **2.25** |
| 1. | Salk Institute for Biological Studies | 1.73 | 0.32 |
| 2. | University of California, San Diego | 1.00 | 1.17 |
| 3. | Scripps Research Institute | 0.91 | 0.33 |
| 4. | Sanford Burnham Prebys Medical Discovery Institute | 0.51 | 0.03 |
| **VII. Seattle-Tacoma-Bellevue, WA** | | | **1.68** |
| 1. | Allen Institute | 2.40 | 0.22 |
| 2. | Fred Hutch Cancer Center | 1.42 | 0.24 |
| 3. | University of Washington | 0.85 | 0.95 |
| **VIII. New Haven-Milford, CT** | | | **1.20** |
| 1. | Yale University | 1.01 | 1.17 |
| **IX. St. Louis, MO-IL** | | | **1.39** |
| 1. | Washington University in St. Louis | 1.06 | 0.96 |
| 2. | Saint Louis University | 0.40 | 0.07 |
| **X. Worcester, MA-CT** | | | **0.28** |
| 1. | University of Massachusetts Chan Medical School | 0.97 | 0.28 |

Notes: Table shows the leading institutions in the ten most productive cities by average researcher output in fundamental science research in life sciences journals. Average researcher output is weighted by the number of researchers in each geography. The rightmost number in the city header row shows the *city*-level share of total world output over the period in percent, while subsequent rows show the institution-level share. Here, world output refers to the total output in our set of 15 journals. 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 2: Movers Design: Summary Statistics by Sample

| | Primary sample | All authors | First & last authors |
|---|---|---|---|
| **Everyone** | | | |
| Number of researchers | 296,757 | 452,574 | 189,832 |
| Years actively publishing | 1.7927 | 1.8902 | 1.7266 |
| | (2.0296) | (2.0595) | (1.9786) |
| Average number of coauthors | 3.6234 | 11.2325 | 1.9728 |
| | (0.6597) | (13.0194) | (0.1460) |
| Annual output | 0.3906 | 0.2226 | 0.6853 |
| | (2.0792) | (0.9405) | (3.8594) |
| Annual output (75th pctl.) | 0.1689 | 0.1740 | 0.2465 |
| Annual output (95th pctl.) | 0.9169 | 0.5667 | 1.5359 |
| Lifetime output | 0.7908 | 0.5011 | 1.2969 |
| | (4.5859) | (2.4780) | (8.1346) |
| Number of moves | 0.1274 | 0.1334 | 0.1231 |
| | (0.3334) | (0.3400) | (0.3285) |
| **Non-Movers** | | | |
| Number of researchers | 258,948 | 392,183 | 166,472 |
| Years actively publishing | 1.4844 | 1.5749 | 1.4280 |
| | (1.5053) | (1.5747) | (1.4597) |
| Average number of coauthors | 3.6322 | 11.3000 | 1.9740 |
| | (0.6721) | (13.3883) | (0.1485) |
| Annual output | 0.3766 | 0.2119 | 0.6615 |
| | (2.1578) | (0.9604) | (3.9561) |
| Annual output (75th pctl.) | 0.1286 | 0.1484 | 0.1875 |
| Annual output (95th pctl.) | 0.8523 | 0.5362 | 1.4259 |
| Lifetime output | 0.5951 | 0.3745 | 0.9736 |
| | (3.5693) | (1.8946) | (5.8643) |
| **Movers** | | | |
| Number of researchers | 37,809 | 60,391 | 23,360 |
| Years actively publishing | 3.9046 | 3.9377 | 3.8546 |
| | (3.4207) | (3.2931) | (3.3862) |
| Average number of coauthors | 3.5635 | 10.7941 | 1.9642 |
| | (0.5643) | (10.2971) | (0.1270) |
| Annual output | 0.4859 | 0.2923 | 0.8553 |
| | (1.4248) | (0.7958) | (3.0785) |
| Annual output (75th pctl.) | 0.4332 | 0.3159 | 0.6690 |
| Annual output (95th pctl.) | 1.2612 | 0.7359 | 2.1883 |
| Lifetime output | 2.1317 | 1.3232 | 3.6005 |
| | (8.7035) | (4.6827) | (16.9296) |
| **One-time Movers** | | | |
| Number of researchers | 25,845 | 40,948 | 15,871 |
| Years actively publishing | 4.1203 | 4.1683 | 4.0567 |
| | (3.5833) | (3.4582) | (3.5380) |
| Average number of coauthors | 3.5610 | 10.5577 | 1.9657 |
| | (0.5555) | (9.8290) | (0.1228) |
| Annual output | 0.4861 | 0.2962 | 0.8476 |
| | (1.3166) | (0.7728) | (3.1243) |
| Annual output (75th pctl.) | 0.4547 | 0.3251 | 0.6973 |
| Annual output (95th pctl.) | 1.2743 | 0.7429 | 2.1991 |
| Lifetime output | 2.2441 | 1.4220 | 3.7500 |
| | (9.2125) | (4.9753) | (18.4036) |

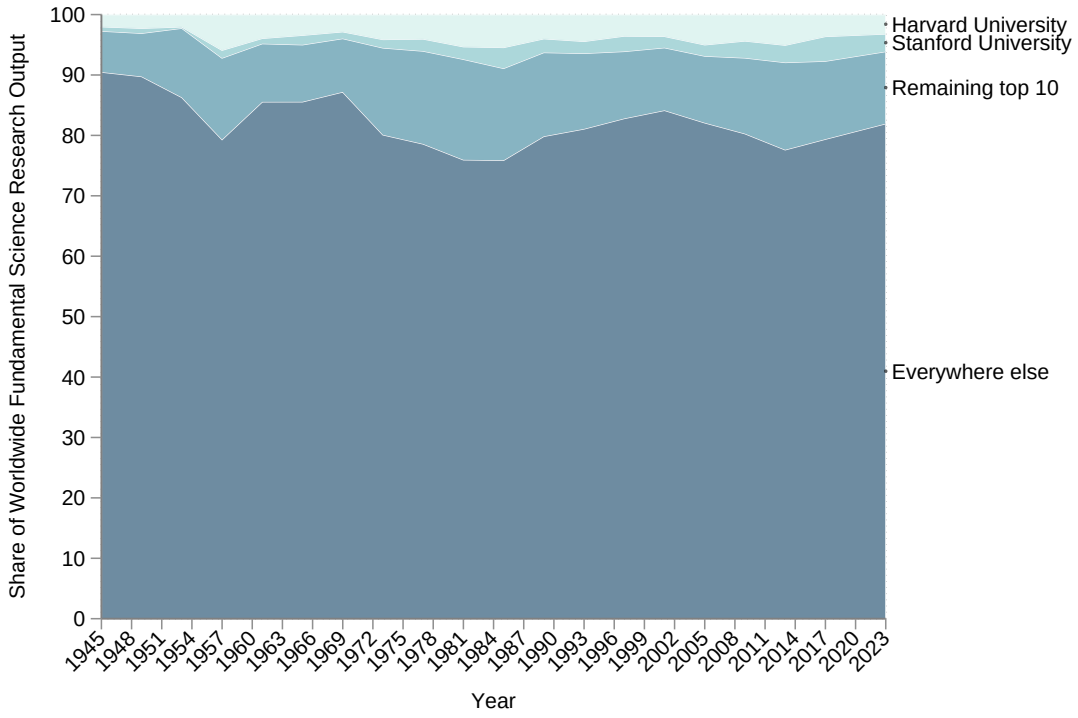
Notes: Table shows individual- and city-level summary statistics for 1945–2023 across three sets of author samples: (1) our primary sample with first, second, second to last, and last authors (2) all authors, (2) and first & last authors. Standard deviations for averages are shown in parentheses below the estimates.

## Table 3: Decompositions of Institutional Output

| **Panel A: Variance Decomposition of Institutional Output** | | | | |
|---|---|---|---|---|
| **Cross-institution variance of mean:** | | | | |
| Institutional output | | | | 2.93 |
| Researcher effects | | | | 1.15 |
| Institution effects | | | | 0.47 |
| **Correlation of average researcher and institution effects** | | | | 0.30 |
| **Share of institutional output variance that would be eliminated if:** | | | | |
| Researcher effects were made equal | | | | 0.61 |
| Institution effects were made equal | | | | 0.84 |
| **Panel B: Additive Decomposition of Institutional Output** | | | | |
| **Top & Bottom** | **5%** | **10%** | **25%** | **50%** |
| **Difference in Research Output** | | | | |
| Overall | 6.39 | 5.22 | 3.46 | 2.07 |
| Researchers | 1.79 | 1.39 | 0.85 | 0.55 |
| Institutions | 4.60 | 3.84 | 2.61 | 1.52 |
| **Share of difference attributable to** | | | | |
| Researchers | 0.28 | 0.27 | 0.24 | 0.26 |
| Institutions | 0.72 | 0.73 | 0.76 | 0.74 |

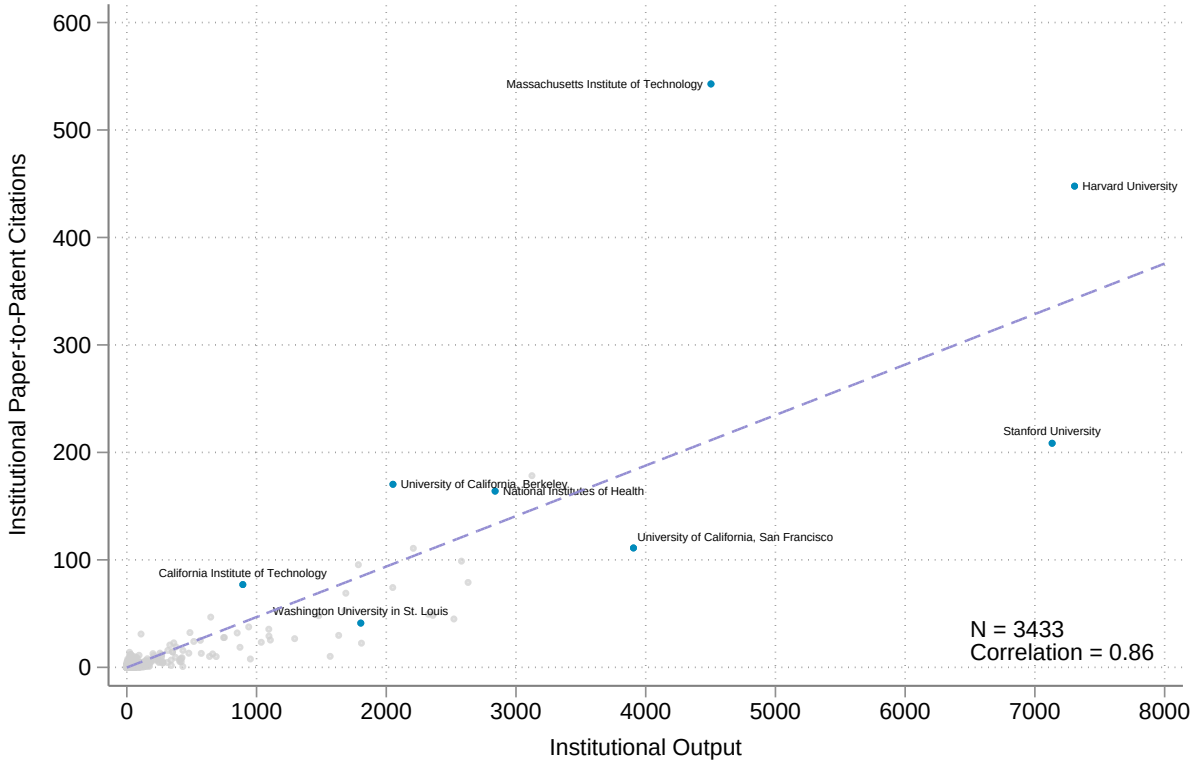
Notes: Panel A reports the variance decomposition of research output across institutions. The first part reports the overall variance and its components, while the second part illustrates the share of variance that would be reduced if researcher or institution effects were made equal. Panel B decomposes the difference in researcher output across top and bottom percentile groups. The first part presents the absolute difference in output, while the second part shows the share of this difference attributable to individual researchers versus institutions. Sample includes all authors (movers and non-movers) from 1945–2023. We reduce noise in our institution fixed effects estimates by grouping our set of over 4000 institutions into 100 clusters based on their empirical research output distribution. In total we have 100 institution-cluster fixed effects and 87,664 unique researcher fixed effects before averaging to the institution-level.

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



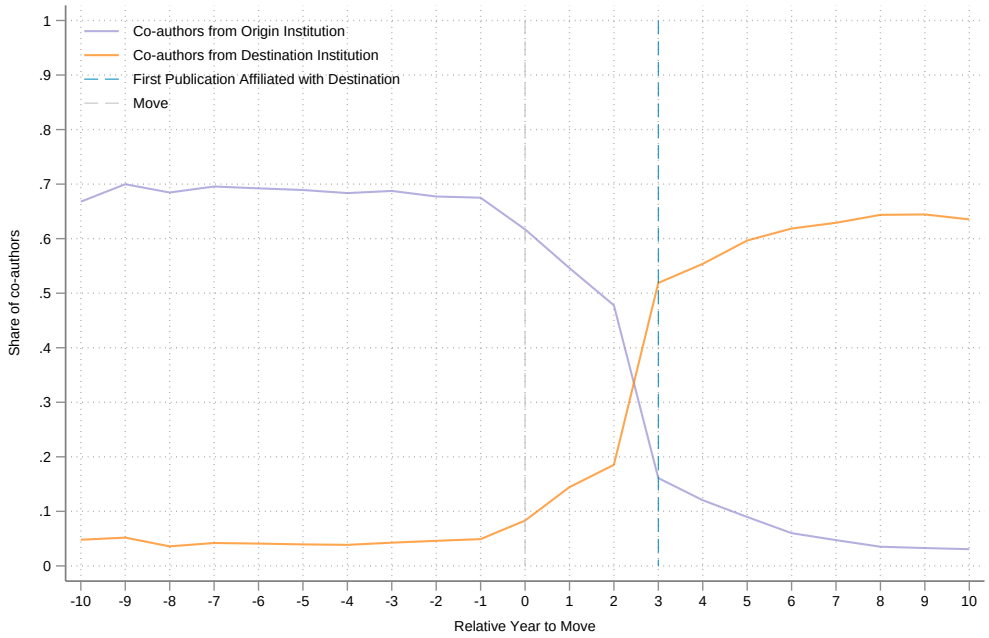

Notes: Figure shows the institution-level trends in the share of total fundamental science research output from 1945 to 2023 using three-year bins and publications in a set of 15 top life-sciences journals *(Cell, Nature, Science, Nature Cell Biology, Nature Genetics, Nature Medicine, Nature Biotechnology, Nature Chemical Biology, Nature Neuroscience, Neuron, Cell Stem Cell, PLOS One, Oncogene, Journal of Biological Chemistry, and the FASEB Journal).* Our measure of output 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 output 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, 2015–2023




Notes: This figure shows the analogous relationship at the institution level. There are $N = 3433$ institutions in our sample. Around 15.5% of papers in our sample have ever been cited in a patent.

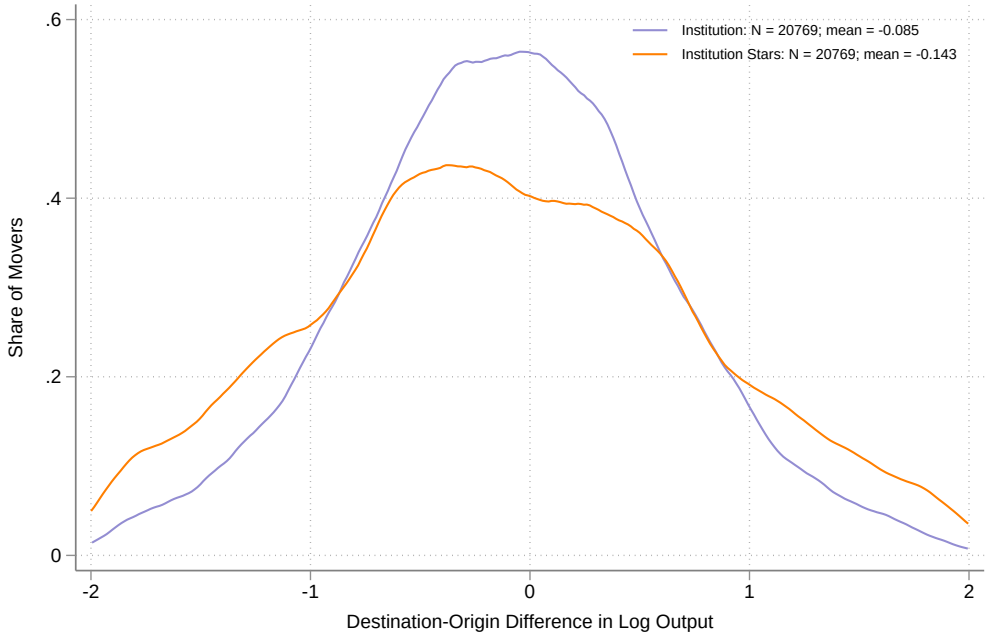
Figure 3: Share of Fundamental Life Science Coauthors from Origin vs. Destination Institution



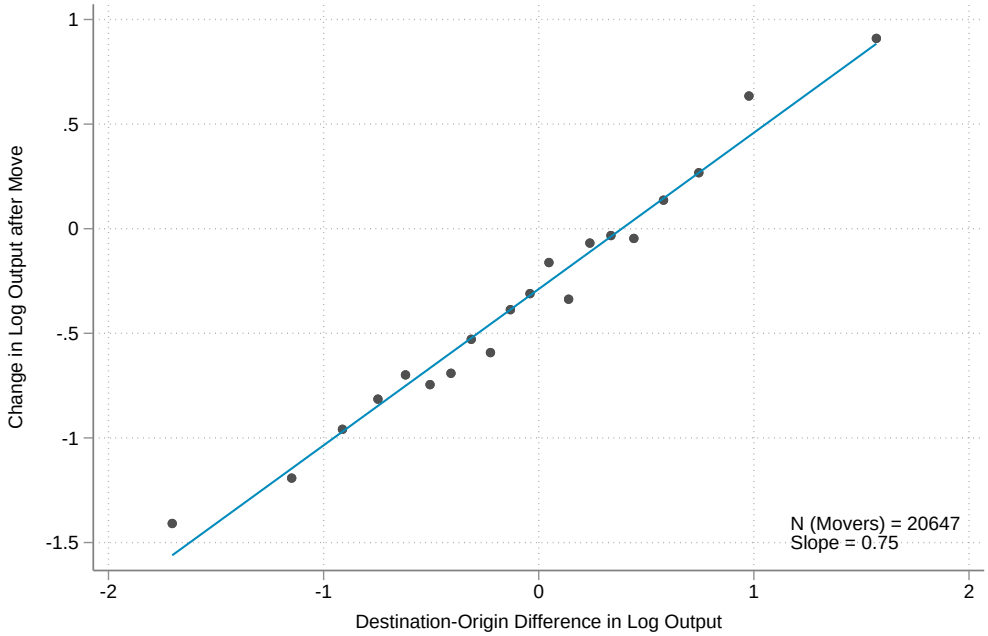

Note: Figure plots the share of coauthors that are affiliated with the mover's origin and destination institution over time. Move year is mechanically defined as three years prior to the first publication at the mover's destination.

Figure 4: Move Size: Distribution and Effect on Output

(a) Distribution of $\delta_i$



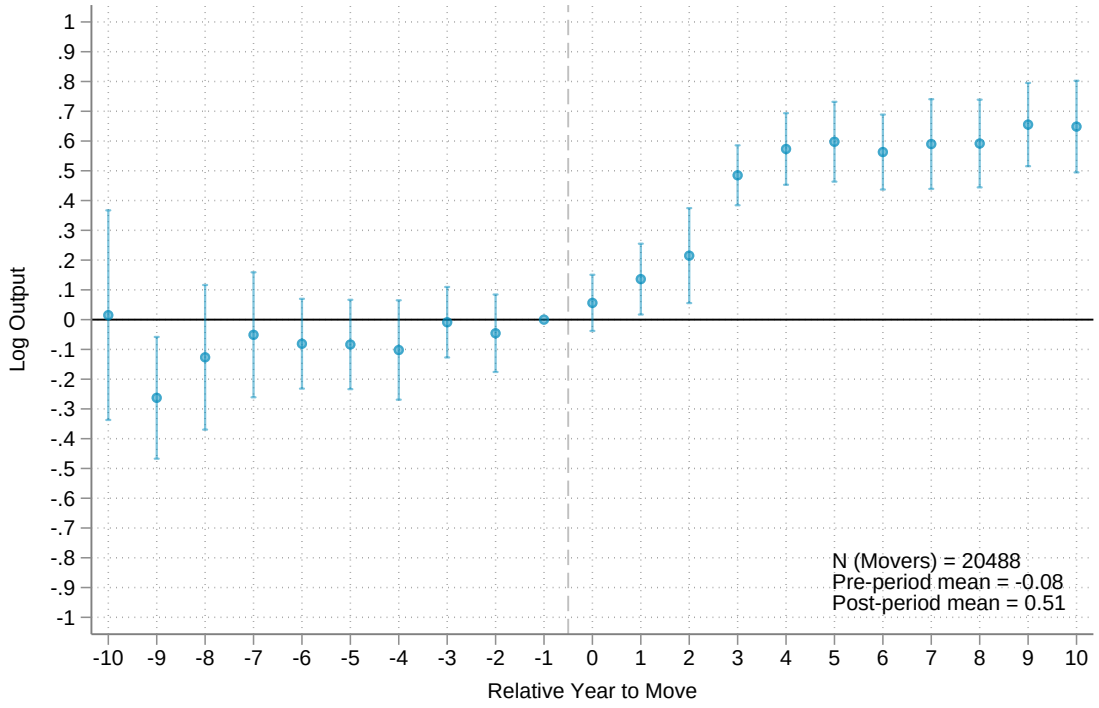

(b) Change in Log Output by Move Size



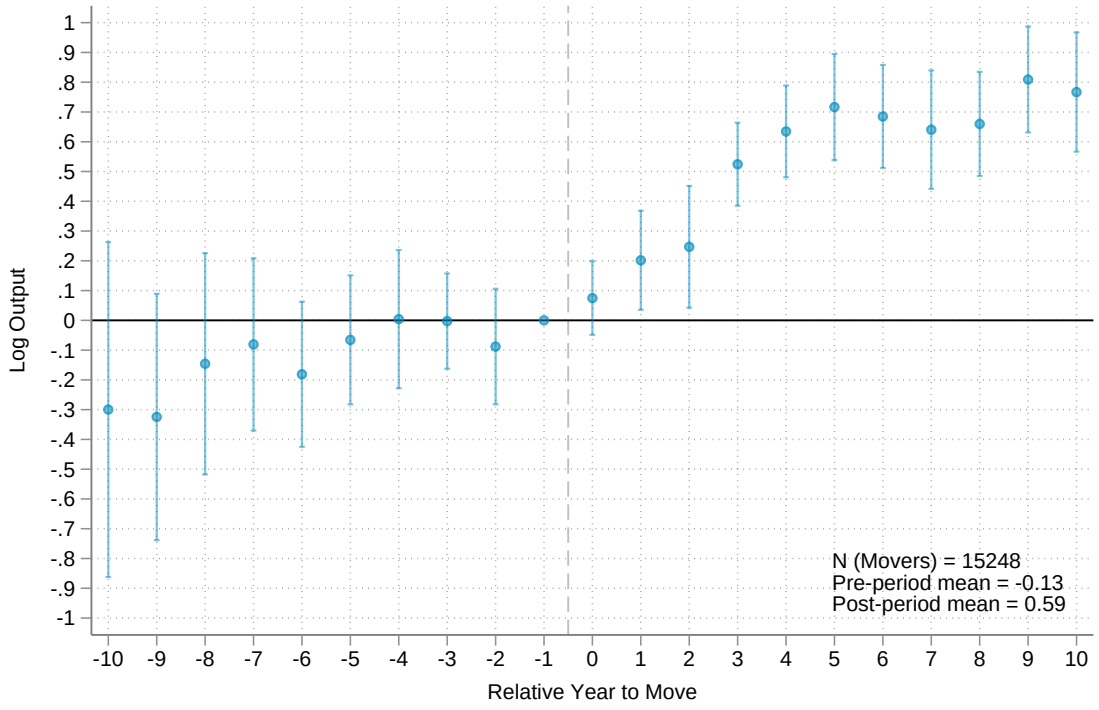

Notes: Panel (a) shows the distribution of $\delta_i$ for scientist-movers, where $\delta_i$ measures the difference in average output between the destination and origin institution. Distributions are shown at the institution level (purple) and for star researchers at the institution level (orange). Panel (b) plots the change in log output against move size. Move size is defined as the difference in average output between a mover's origin and destination institution.

Figure 5: Output Change in Response to Move

(a) 1945-2023

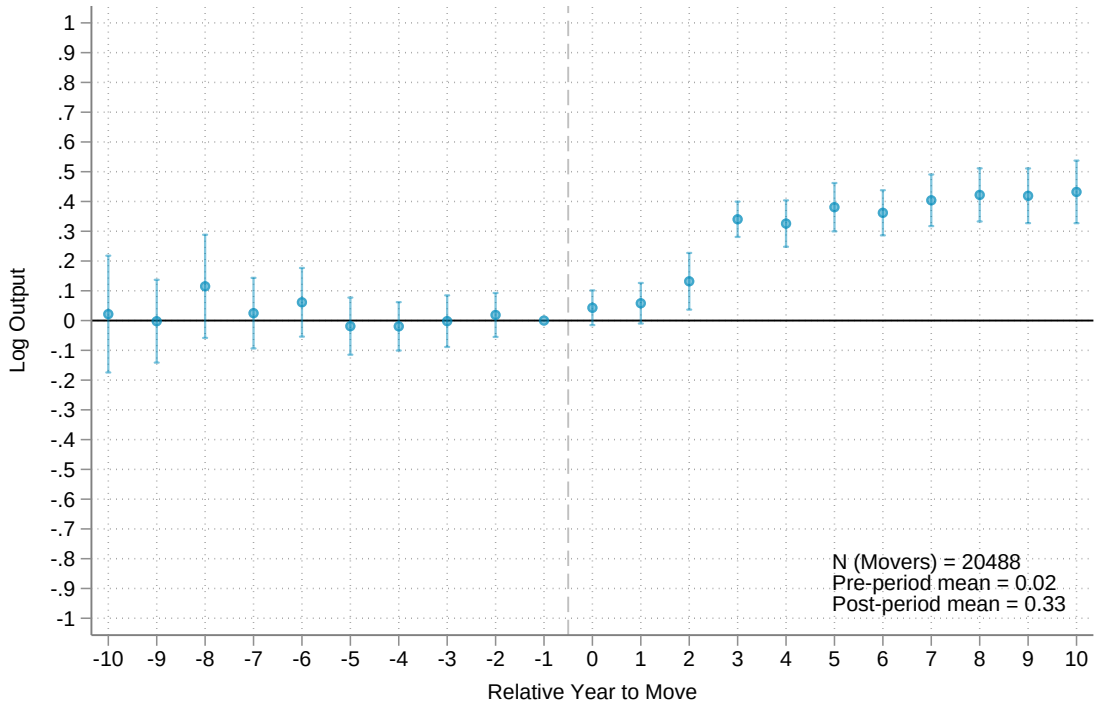



(b) 1995-2023




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

Figure 6: Impact of Star Researchers




Notes: This figure reports coefficients from a canonical event–study that tracks how the log output of the star's coauthors evolves in the years surrounding the move. For both, the coefficient for relative year -1 is normalized to 0. The dependent variable is log output and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

# Appendix

## A. Additional Exhibits

Appendix Table A1: Journal Sample Sizes

| Sample | Journal | % Fundamental | # Fundamental |
|---|---|---|---|
| CNS | Science | 37.10 | 39,454 |
| | Nature | 64.36 | 64,034 |
| | Cell | 93.89 | 16,163 |
| Leading Life-Sciences Journals | Neuron | 93.14 | 8,464 |
| | Nature Genetics | 90.59 | 5,775 |
| | Nature Medicine | 57.11 | 4,625 |
| | Nature Biotechnology | 43.14 | 3,445 |
| | Nature Neuroscience | 87.08 | 4,252 |
| | Nature Cell Biology | 84.97 | 3,182 |
| | Nature Chemical Biology | 78.45 | 2,235 |
| | Cell Stem Cell | 89.24 | 1,642 |
| Other Life-Sciences Journals | PLOS One | 81.44 | 210,334 |
| | Journal of Bio Chem | 95.07 | 171,693 |
| | Oncogene | 97.92 | 15,604 |
| | The FASEB Journal | 69.06 | 12,718 |

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

Appendix Table A2: Measure Weighting Robustness

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

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

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

| | State-Level | City-Level | Institution-Level |
|---|---|---|---|
| **Research Output vs. Papers-in-Patents Citations** | | | |
| Impact Factor & Citation Weighted vs. All Cites | 0.9950 | 0.9916 | 0.9598 |
| Impact Factor & Citation Weighted vs. Front Page Cites | 0.9948 | 0.9929 | 0.9682 |
| Impact Factor & Citation Weighted vs. Body Cites | 0.9939 | 0.9857 | 0.9337 |

Notes: Table shows the correlation between our main measure of fundamental science output 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: US Institutions Ranked by Average Researcher Output, 2015-2023

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

Notes: Table shows the leading 50 institutions by average author-year output in fundamental science research in life sciences journals. The output 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: International Institutions Ranked by Average Researcher Output, 2015 - 2023

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

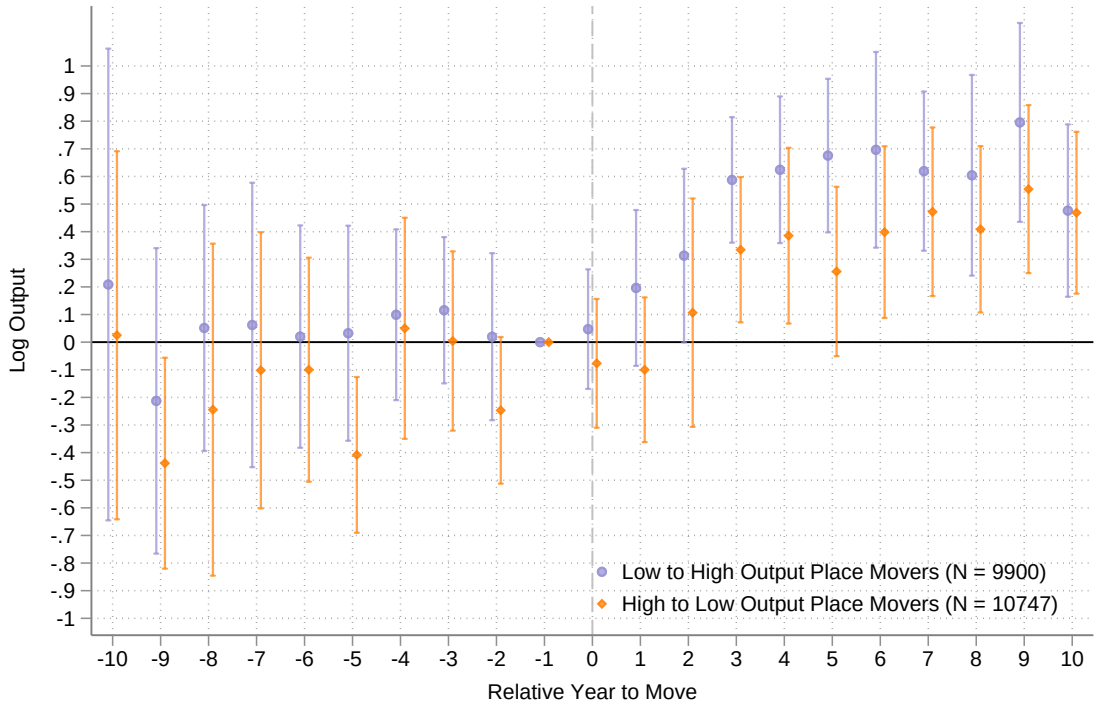
Notes: Table shows the leading 50 international institutions by average researcher output in fundamental science research in life sciences journals. The output measure is weighted by the number of researchers at each institution. Journals included are: *Cell, Nature, Science, Nature Cell Biology, Nature 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 A6: Post-period Averages and Elasticities

| Effect | Estimate |
|---|---|
| **Panel A: Movers Design, Average Effect for Post Period** | |
| Primary specification | 0.5100 |
| | (0.1811) |
| Move size defined by star output | 0.3300 |
| | (0.1288) |
| **Panel B: Event study – Elasticities (per 1% change)** | |
| Researcher output, with respect to non-federal life-sciences funding | 0.0800 |
| | (0.0432) |
| Researcher output, with respect to federal life-sciences funding | 0.1100 |
| | (0.0481) |
| Researcher output, with respect to total life-sciences funding | 0.1000 |
| | (0.0485) |
| Researcher output, with respect to total output | 0.1300 |
| | (0.0471) |
| Unique US coauthors, with respect to total output | 0.1300 |
| | (0.0583) |

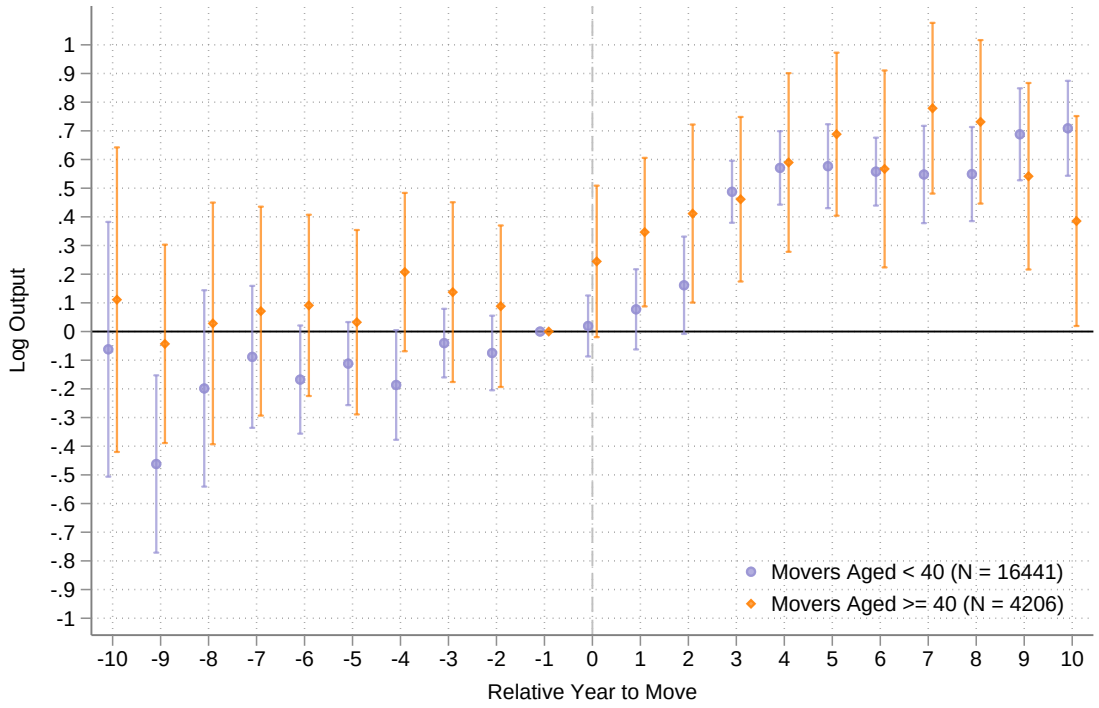
Notes: Panel A reports post-period averages from the movers design for the main specification and a variant that defines move size using star output. Panel B reports event-study elasticities: for logged outcomes (researcher output), coefficients are elasticities per 1% change in the covariate; for coauthor counts (levels), coefficients are percent-to-level semi-elasticities. Life-sciences funding is from the National Science Foundation's Higher Education Research and Development (HERD) Survey; institution-level averages are computed over the available survey years, 2010–2022.

## Appendix Figure A1: Output Change based on Move Size



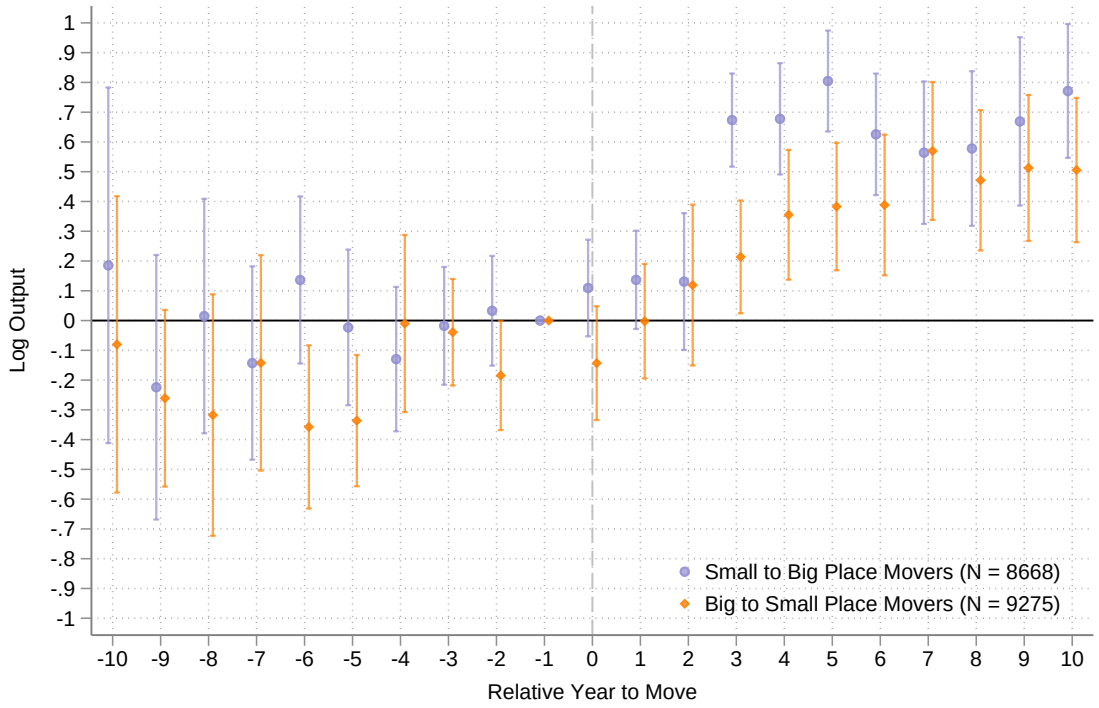

Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers based on move size for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log output and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

## Appendix Figure A2: Output Change based on Age



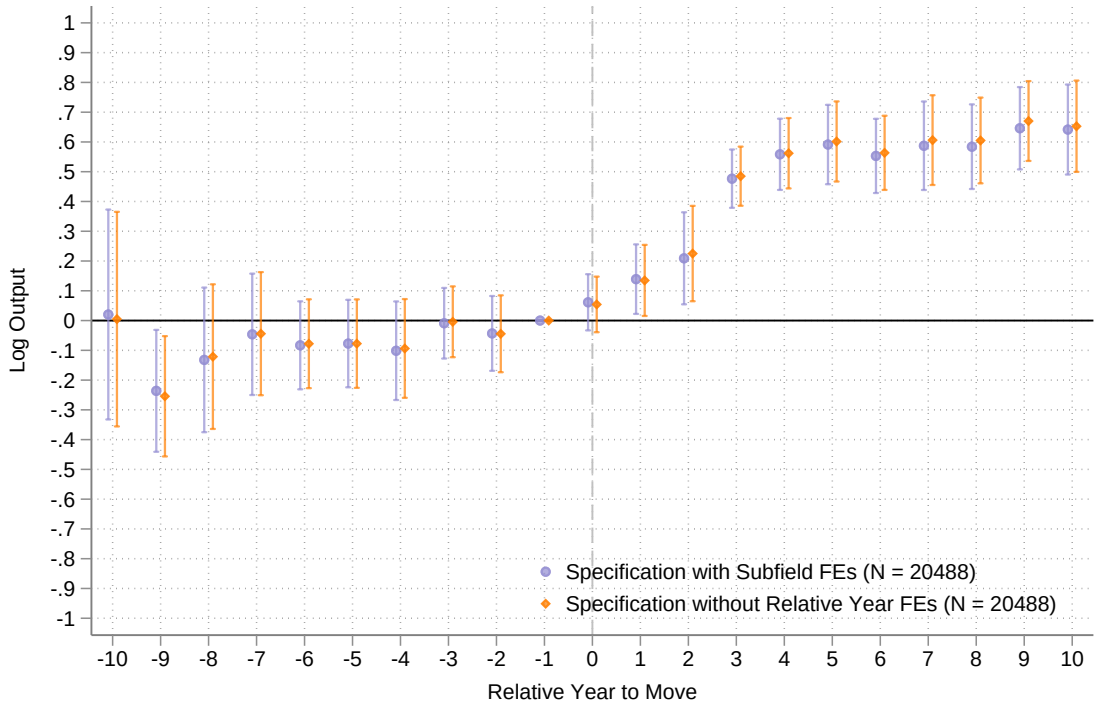

Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers split by age for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log output and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

## Appendix Figure A3: Output Change based on City Size



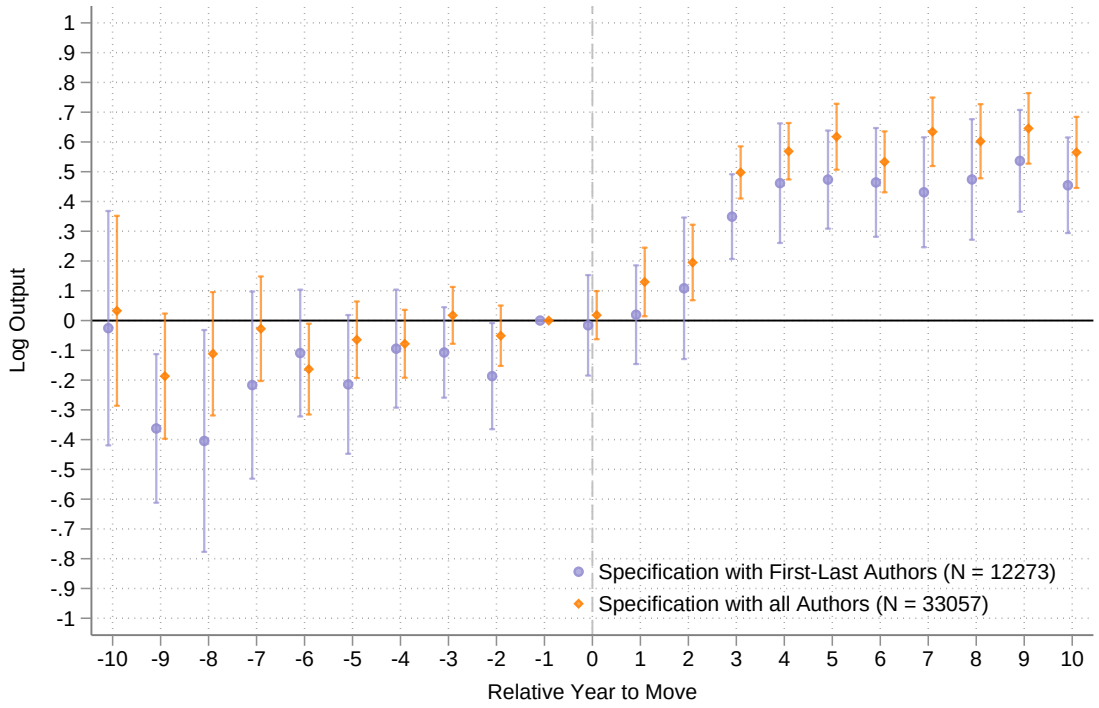

Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for movers based on city size changes for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. In calculating city size, we ignore the researchers located at the mover's origin and destination location. The dependent variable is log output and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

## Appendix Figure A4: Alternative Regression Specifications



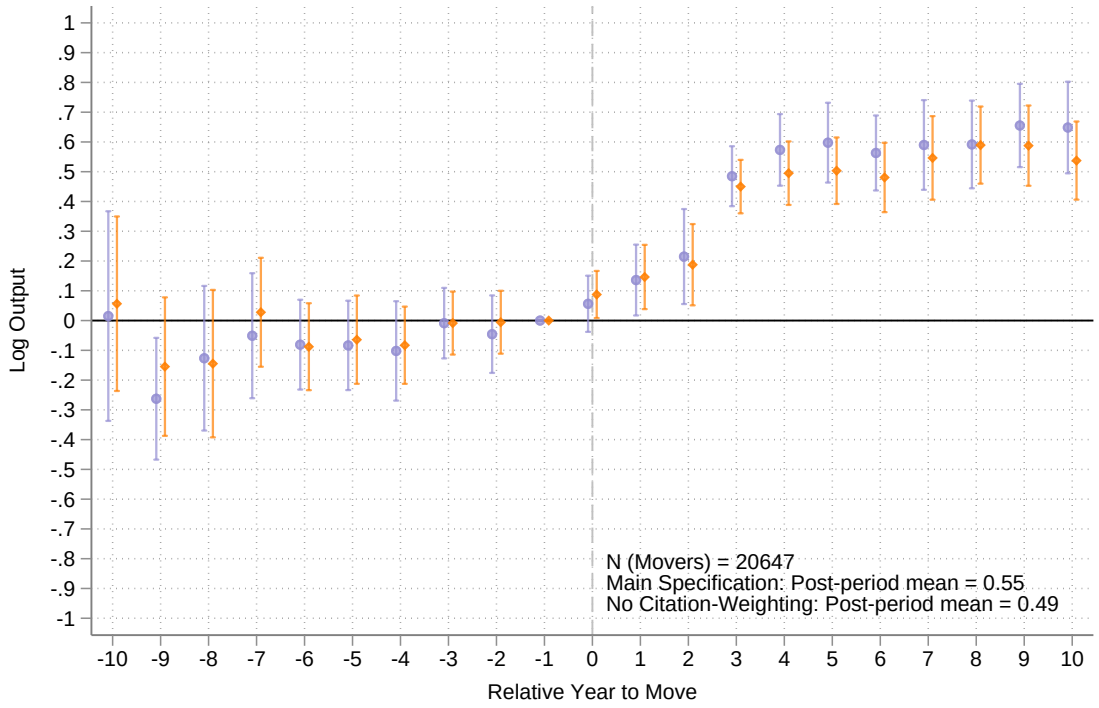

Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 with the inclusion of subfield fixed effects and without the inclusion of relative year fixed effects for the time period between 1945-2023. The coefficient for relative year -1 is normalized to 0. The dependent variable is log output and we control for a set of year and author fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

## Appendix Figure A5: Alternative Author Samples




Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for the sample of first-last authors and all authors publishing in the top 15 journals. The coefficient for relative year -1 is normalized to 0. The dependent variable is log output and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

## Appendix Figure A6: Removing Citation Weighting




Notes: Figure shows the coefficients $\hat{\theta}$ estimated from equation 2 for our main specification and for output being defined without citation-weighting. The coefficient for relative year -1 is normalized to 0. The dependent variable is log output and we control for a set of year, author, and relative-year fixed effects in our specification. Standard errors are clustered at the institution-level. We focus on scientists who only moved once between 1945-2023.

# B. Weighting Publications for Impact

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

Letting $j$ denote journals and $i$ denote papers, define

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

We first adjust the distribution of journal representation in our sample according to their impact factors. This redistribution results in a sample that shifts weight toward journals with higher impact factors, regardless of their original number of papers. This helps adjust for the fact that higher impact journals may publish less often than lower impact journals. For instance, PLOS One and Journal of Biological Chemistry have raw journal sample sizes that each represent 31% to 37% of our full sample but have relatively 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 its paper count $W_j$ as

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

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

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

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

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

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

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

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

factor and citation weighted output measure is equal to

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

# C. Constructing Life-Sciences Subfields

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

## Year Bin: 1945–1964

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

## Year Bin: 1965–1984

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

## Year Bin: 1985–2004

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

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

## Year Bin: 2005–2024

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

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

# D. Robustness to Limited Mobility Bias

As discussed in Section IV.D, estimates of institutional effects ($\hat{\gamma_j}$) derived from a sample with finite movers are subject to limited mobility bias (Andrews et al., 2008). This measurement error inflates the estimated variance of place effects and biases the correlation between researcher and place effects downward. In the main text, we address this by following (Bonhomme et al., 2019) (BLM) and grouping institutions into clusters based on their empirical distribution.

In this appendix, we demonstrate the robustness of our results by estimating the variance decomposition on a restricted sample of institutions with at least 25 movers. This ensures that the variance of the estimation error is sufficiently small relative to the variance of the true effects, as suggested by (Andrews et al., 2008).

In addition, we also employ a non-parametric Empirical Bayes correction (Morris, 1983) to shrink noisy estimates toward the average. This approach pools information across institutions to improve our estimated fixed effects, shrinking unreliable estimates (those with large standard errors) more aggressively. We assume that our estimated fixed effect $\hat{\gamma}_j$ is an unbiased estimate of the true latent effect $\gamma_j$ observed with additive measurement error $\eta_j$:

$$\hat{\gamma}_j = \gamma_j + \eta_j \tag{3}$$

where $\eta_j \sim N(0, s_j^2)$ and $s_j^2$ is the squared standard error of the fixed effect estimate for institution $j$. We further model the distribution of true institutional effects as normal with mean $\mu$ and variance $\sigma_\gamma^2$:

$$\gamma_j \sim N(\mu, \sigma_\gamma^2) \tag{4}$$

Under these assumptions, the best linear unbiased predictor of the true institution effect $\gamma_j$ is the posterior mean, calculated as a precision-weighted average of the observed estimate and the prior mean:

$$\hat{\gamma}_j^{EB} = \frac{\hat{\sigma}_\gamma^2}{\hat{\sigma}_\gamma^2 + s_j^2}\hat{\gamma}_j + \frac{s_j^2}{\hat{\sigma}_\gamma^2 + s_j^2}\hat{\mu} \tag{5}$$

where $\hat{\sigma}_\gamma^2$ is the estimated variance of the true effects (signal) and $s_j^2$ is the error variance (noise). The term $\frac{\hat{\sigma}_\gamma^2}{\hat{\sigma}_\gamma^2 + s_j^2}$ represents the signal-to-noise ratio. For institutions with precise estimates (small $s_j^2$), this weight approaches 1, preserving the original estimate. Conversely, for institutions with noisy estimates, the weight decreases, and the estimate is shrunk aggressively toward the sample mean $\hat{\mu}$.

We estimate the hyperparameters $\hat{\mu}$ and $\hat{\sigma}_\gamma^2$ using the method of moments on the restricted sample:

$$\hat{\mu} = \frac{1}{J}\sum_{j=1}^{J}\hat{\gamma}_j, \qquad \hat{\sigma}_\gamma^2 = \frac{1}{J}\sum_{j=1}^{J}\left[(\hat{\gamma}_j - \hat{\mu})^2 - s_j^2\right] \tag{6}$$

The subtraction of $s_j^2$ in the variance equation serves as a bias correction for the sampling variance of $\gamma_j$. We then use these Empirical Bayes adjusted estimates ($\hat{\gamma}_j^{EB}$) to re-calculate the variance decomposition shares.

Table A7 reports the results. Column (1) presents the baseline estimates (BLM). Column (2) restricts the sample to institutions with at least 25 movers. Column (3) applies the Empirical Bayes shrinkage to the restricted estimates.

## Appendix Table A7: Decompositions of Institutional Output (Robustness)

| Noise Reduction Method | BLM | $N_{\text{movers}} \geq 25$ | $N_{\text{movers}} \geq 25$ + EB |
|---|---|---|---|
| **Cross-institution variance of mean:** | | | |
| Institutional output | 2.93 | 0.57 | 0.57 |
| Researcher effects | 1.15 | 0.25 | 0.02 |
| Institution effects | 0.47 | 0.28 | 0.28 |
| **Correlation of average researcher and institution effects** | 0.29 | 0.06 | 0.11 |
| **Share of institutional output variance that would be eliminated if:** | | | |
| Researcher effects were made equal | 0.61 | 0.55 | 0.97 |
| Institution effects were made equal | 0.84 | 0.50 | 0.50 |

Notes: Table reports variance decompositions under three approaches to addressing limited mobility bias. Column 1 applies the grouped fixed effects method of Bonhomme et al. (2019), clustering institutions into 100 groups. Column 2 restricts the sample to institutions with at least 25 movers, reducing noise in the estimated fixed effects. Column 3 applies Empirical Bayes shrinkage to the restricted sample, pulling imprecise institution and researcher estimates toward their respective grand means.