The Mobility of Scientists and Inventors: Patterns and Determinants

PhD Program in Economics, Management and Data Science

XXXI Cycle

By

Luca Verginer

2018

Contents

List of Figures

List of Tables

Abstract

Life Scientists working both at Universities and private institutions are very mobile. This fact is reflected both in their tendency to move globally, from city to city, as well as from institution to institution. This thesis addresses several questions regarding the mobility patterns of these scientists and tests possible determinants for their relocation choice.

We develop a novel dataset tracking the mobility of 3.7 million scientists across 9,745 cities over two decades. We show that mobility is marked by national borders and shared languages and that the mobility network is dominated by a small set of "global cities". We also find that only a few countries clearly benefit from international exchange. Moreover, we find that young and prolific researchers gravitate towards these "global cities".

We use the mobility data to show how state and federal Stem Cell funding restrictions in the US have affected the spatial distribution of scientists as well as their propensity to leave the country. In fact we find that differential state and federal approaches to Stem Cell research has had the overall effect of geographically concentrating scientists and averting an exodus of these researchers.

Finally, we analyze the impact of M&As in the Pharmaceutical Sector to establish if these shocks cause a higher than average turnover. High turnover of R&D personnel in this R&D intensive sector is an undesirable outcome. We do in fact find that turnover is higher following an acquisition. However, as noted elsewhere, acquired companies experience often financial distress before the event and so defection starts even before the deal takes place.

Chapter 1

Introduction

Mobility of scientists and inventors — highly skilled individuals — is an important modern phenomenon. Their mobility has caused alarm with fears that "brain drain" is an inevitable consequence of the mobility of skilled labor. However, as with any complex real world phenomenon international mobility is a multifaceted issue, which requires careful analysis to assess its impact beyond the simple notion of mobility as a zero-sum game. The analysis of individual level mobility remains challenging, given the difficulties to obtain individual level mobility data. The authoritative manual on the "Global Mobility of Research Scientists" (Geuna, 2015, Ch.5, p.24) notes that research "on the mobility of researcher scientists is scarce because of a lack of reliable data to trace scientists along their careers". This is reflected in the relative small number of works looking at global mobility patterns and its characterization. It is here that we want to contribute to the literature on the mobility of researchers, by presenting a characterization of the global mobility network of scientists (see Chapter 2). We highlight the importance of geography and provide estimates of the global "brain circulation".

This mobility analysis offers interesting and relevant findings in and off itself, however more importantly, individual level global mobility data can be used to evaluate the impact of research policies. Specifically, for impact assessment of research policies it is common practice to proxy

impact by publications and patenting output aggregating the effect at either firm or geographic area. However, scientific output as a response to policy can be argued is the sum of several effects. Scientists might (1) reduce their scientific output in numbers and quality even stopping their career in the given field completely, (2) potential scientists in anticipation of reduced career or funding prospects opt to engage in other research and (3) vote with their feet and move to more supportive places. A similar argument can be made for the response of inventors, working on new technologies, when confronted with either sectoral or firm level shocks. Specifically, we look at scientists and inventors as they responded to restrictive Stem Cell regulations in the US (see Chapter 3) and if inventors working for acquired firms reacted by leaving the company. By looking at the response of affected researchers working on Stem Cells, we can assess how differential approaches at US state and federal level have impacted geographic preferences. Switching from authors of papers (i.e. scientists) to producers of patents, we look at how inventors respond to shocks to their working environment (See Chapter 4). Here again, inventors can vote with their feet and leave the company, adversely affecting the patenting potential of the firm.

In Chapter 2 we reconstruct the mobility paths of scientists through their publication history. Tracking scientist through their paper trail, an "activity based " approach, is best suited for estimating mobility patterns and brain circulation phenomena. The use of papers as direct signal of production and location alleviates problems one might encounter in surveys and scraping[1] of job listing services (e.g. LinkedIn). Most importantly, however publications are the actual output of interest when studying scientific output. In this work, we focus primarily on the mobility network extracted from the 1999 to 2009 period. We find that the mobility is heavily influenced by national borders, which confirms similar results in the literature regarding collaboration networks. The community detection indicates also that this "border effect" is alleviated by a shared language. For example, Spain has more exchanges of scientist with Mexico then with Portugal and Portugal more with Brazil, even

[1] A method to download data from the open web by simulating human interaction.

though they share a border. Similar examples can be found in Belgium, where part of it is placed in the French community while the other in the Netherlands, reflecting the French (language), Dutch (language) divide. This result suggests that national borders matter and must be considered when analyzing scientist mobility at global scale. An analysis of inter-city movements taking productivity into account offers an overview on which moves are associated with the highest gain in productivity and which offer below average returns. Here we observe that for most pairs of cities there is an imbalance suggesting that there is preferred direction, however we do not find a correlation between the intensity of the flow and the gain. This suggest that while there is preferred direction mobility should take place this does not affect the intensity of the flow. At national level, we obtain interesting results regarding, which countries are most likely to directly (i.e. we do not take higher order effects into account) benefit from the international exchange. For example, USA, Switzerland, and Spain in the given period experience a substantial boost to their national innovation systems, however Argentina is one of the major losers in this comparison. Additionally, we find evidence that the scientists that do move to the most central cities (i.e. global cities) are those that are most prolific before the move. Similarly, to analyze at individual city level, how well they can cope with the turnover of scientists, we measure if the output coming from those that leave the city can be covered by new arrivals. Again, we find that US cities can manage the turnover well, with continental European cities (e.g. Berlin and Paris) not clearly able to replace lost scientific output by attracting new scientists. While there are a lot of dimensions and higher order effect which might reasonably affect the quality of scientific output at scientist, city and country level, this result suggests clearly that international circulation does not benefit all countries unambiguously.

In Chapter 3 we look at Human Embryonic Stem Cell (hESC) research. HESC is a controversial research topic, which has lead various countries to either heavily regulate or ban it summarily. With the extensive dataset obtained in the first part, we consider the relocation choice of hESC scientist in the US over the period 1998 to 2008. This period cov-

ers the beginnings of the field up to the Bush years, which saw a ban on federal funding. We look specifically at the US, because we have a large resident scientist population, which allows us to identify the affected scientists as well as constructing a believable control group. An additional reason to choose the US is that the US is marked by a significant within mobility. This means that relocation is not uncommon and we can exploit this variability to study relocation choices. US states can be classified into three camps regarding their stance on the issue: "supportive", "baseline" and "restrictive". With restrictive states strictly regulating funding and permissible experiments, which go above and beyond the federal norm. Permissive states on the other hand actively support research with additional funding and fewer regulations. The baseline states reflect the federal guidelines without major deviations. This dimension of heterogeneity beyond changes in Federal funding regulations, allows us to separate the effect of federal and state legislation. On the one hand, we find that the stance of the various US states does have a significant effect on the distribution of hESC researchers, i.e. supportive states have on average a higher portion of hESC researchers, on the other hand we do not find any indication that the mobility itself is significantly affected. We do however find evidence to suggest that a multi-billion dollar funding initiative in California has contributed to avert an exodus of hESC researchers following strict federal funding restrictions (i.e. the Bush Ban).

And finally, in Chapter 4 we consider the propensity of Inventors in the Life Sciences to leave a company following a takeover (i.e. outright acquisition by other firm) Specifically we consider acquisitions in the Pharmaceutical insdustries in the period 1998 to 2008. The motivations for the acquisition are varied, but given the R&D intensive nature of this sector it stands to reason that when a company is acquired the acquiring company wants to retain not only the explicit Intellectual Property but also the tacit knowledge embodied in the Inventors carrying out the research. For this reason, we want to assess, if there is a significant effect on turnover due to an acquisition event and if so how strong it is. We use patent data to reconstruct the employment history through patenting activity using a dataset of disambiguated Assignees and Inventors.

To have a good control group and estimate the effect of an acquisition on the turnover, using the Difference in Difference approach, we match treated and control companies on their size (no. of patents), patenting rate (patents per year) and IPC similarity. Given this match we proceed to find controls for the treated employees working for the control companies 4 years before the event. We do in fact find that acquired companies after the event loose on average 30% of their employees to third parties above and beyond the comparable firms.

Chapter 2

Brain-Drain Network: The International Mobility of the Life Scientists

An earlier version of this chapter has been published as a working paper on REPEC.[1] *My co-author on this project is Prof. Massimo Riccaboni, also listed in the REPEC preprint.*

2.1 Introduction

Scientists are known to be highly mobile intellectuals, especially in the early phase of their careers. This has been true in the past (Cardwell, 1972; Mokyr, 2016; Serafinelli and Tabellini, 2017), but the size of the phenomenon has drastically increased in a globalized market for advanced human capital (Culotta, 2017; Geuna, 2015; OECD, 2017). Modern economies require a highly skilled labor force to maintain their competitive advantage and grow (Chambers et al., 1998; Solimano, 2008; Ozden and Rapoport, 2018; Zucker and Darby, 2007). Which makes it important to understand what determines this mobility. The authoritative manual on the "Global Mobility of Research Scientists" (Geuna,

[1] https://ideas.repec.org/p/ial/wpaper/4-2018.html.

2015, Ch.5, p.24) gives an overview of the current state of the research on the mobility of scientists and notes that research "on the mobility of researcher scientists is scarce because of a lack of reliable data to trace scientists along their careers". We contribute to this literature by constructing and analyzing a large scale and global scientist mobility dataset of 3.7 Million scientists working in 189 Countries and 9,745 cities.

Previous research on the mobility of scientists has used, among other approaches, large-scale surveys (Franzoni, Scellato, and Stephan, 2012; Franzoni, Scellato, and Stephan, 2014; Franzoni, Scellato, and Stephan, 2018), and more recently massive bibliographic databases (Bohannon and Doran, 2017; Deville et al., 2014; Graf and Kalthaus, 2018). There are other sources of mobility information (e.g. Job search portals, social media), however papers offer the most direct and high frequency signal of scientific activity. We take advantage of the fact that scientists, especially in some disciplines, publish regularly in their career, and a lack of publications arguably signals its end. Inspired by bibliographic approaches we use MEDLINE, a large publications repository primarily covering research in the life sciences.

This work focuses the analysis on the level of the most important locations with activity in the life sciences (about 10 thousand populated places). We think that cities and especially global cities are an appropriate level to analyses mobility patterns and their role within the global economy in general (Taylor and Derudder, 2015; Sassen, 2016) and sciences in particular (Catini et al., 2015). We will also discuss implications at national level to complement the discussion on the more granular city level.

In this work, we set out to characterize the geographic determinants of mobility, identify which cities lie at its center and show these "global cities" attract the most prolific scientists. We would like to point out early on that we do not have information on the nationality of the authors and when talking about countries and the moves from, to and within we talk about the origin of the move and not about the nationality of the scientists. This has the important implication, that we talk about *mobility* and not about *migration*.

The rest of the paper is structured as follows. In Section 2.2 we show which data we use for our analysis and how the mobility network has been extracted. Then we characterize the basic properties of the global mobility network of the life-scientists describing topological and geographic features in Section 2.3. In Section 2.4 we present an analysis of "productivity" gains at scientist, city and country level, as a direct result of the observed mobility. We present the findings on the tendency of central cities to attract prolific scientists early in their career in Section 2.5. Finally, in Section 2.6 we summarize the findings, offer an outlook for possible ways to extend the present analysis and discuss how this dataset might be used for different applications.

2.2 Reconstructing the Mobility Network

We reconstruct the mobility paths of life scientists through their publication history. Tracking scientist through their paper trail, an "activity based" approach, is best suited for estimating mobility patterns and brain circulation phenomena. The use of papers as direct signal of production and location alleviates problems one might encounter in surveys and scraping of job listing services (e.g. LinkedIn). Most importantly, however publications are the actual output of interest when studying scientific output.

To reconstruct the mobility paths and estimate productivity we need to merge several sources of information. First, we need a publication repository with enough papers (MEDLINE), proper disambiguation of the authors (AUTHOR-ITY), assignment of these authors to locations (MAPAFFIL) and a proxy for the quality of scientific output and by extension the authors themselves (SCIMAGO).

In this section, we introduce the four datasets, explain how they have been merged, how the mobility networks have been extracted and how we proxy author scientific production.

2.2.1 Data

For the analysis we use four datasets, MEDLINE, AUTHOR-ITY, MAPAFFIL, and SCIMAGO.

MEDLINE provides open access to more than 26 million records of scientific publications, with most of the corpus covering research in the life sciences. The data goes as far back as 1867 (earliest publication in the dataset) and is updated continuously. However, we will focus on papers in the period between 1990 to 2009. We restrict our analysis to this period to have a good coverage and make use of existing high quality disambiguations of scientists (AUTHOR-ITY) and affiliations (MAPAFFIL), which are restricted to this time interval. MAPAFFIL and AUTHOR-ITY have been developed and published by Torvik (2015) and Torvik and Smalheiser (2009).

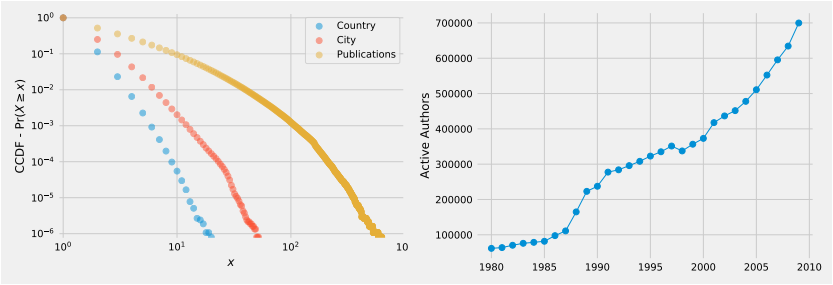
MAPAFFIL lists for a large portion of MEDLINE papers the disambiguated city corresponding to the affiliation of each author as listed on the paper (ca. 37,396,671 author-locations). AUTHOR-ITY developed by Torvik and Smalheiser (2009) contains the disambiguated names of 61,658,514 appearances of names on MEDLINE papers (author-name instances). These author-name instances have been mapped to 9,300,182 disambiguated authors. MAPAFFIL, also developed by Torvik (2015), is a disambiguation of affiliations listed on MEDLINE papers. This dataset allows us to map the affiliation string to the city this affiliation is located in.

By merging MEDLINE with AUTHOR-ITY we obtain the necessary data to uniquely identify an author across publications. This information has been used in the past to reconstruct the global collaboration networks.[2]. The ability to reconstruct mobility comes from merging the previous two datasets with MAPAFFIL. Without this last step, affiliations would not be disambiguated and we would have hundreds of different versions of "Boston University" in our dataset. Fortunately, MAPAFFIL

[2]Examples of co-authorship networks being used for research can be found in Newman (2001), Girvan and Newman (2002), Wagner and Leydesdorff (2005), and Jackson and Rogers (2007)

can accurately[3] map these various strings to a city.

By adding location information to the publication records we obtained for each author-publication pair a date and location. An example of which is available in Table 16 in the Appendix. From MAPAFFIL we obtain as location the center of a city (low resolution), however these are mixed with locations at a higher resolution, which identify a suburb or part of a city. For example, for "London, UK" we have the location (lat=51.5, lon=-0.126) but also 118 districts or city parts (i.e. "Bethnal Green, London, UK", "Goodmayes, Ilford, Redbridge, London, UK"). These have been reduced to the lowest common resolution So "Bethnal Green, London, UK" and "Goodmayes, Ilford, Redbridge, London, UK" would be mapped to "London, UK" at position (lat=51.5, lon=-0.126). And similarly, the Boston neighborhoods "Jamaica Plain, Boston, MA, USA" and "Roslindale, Boston, MA, USA" are mapped to the lower resolution city center "Boston, MA, USA" (lat=42.359, lon=-71.057). By applying this method, we obtain 9,745 urban areas.



(a) (b)

Figure 1: (a) Counter Cumulative Distribution of Countries, Cities and Publications per author. Each data point shows the probability to observe at least x unique Countries, Cities and Publications for a given author (i.e. $\Pr(X \geq x)$). (b) Number of unique active authors identified in AUTHORITY.

To have an appreciation for the number of unique cities, publications

[3]Torvik (2015) give a thorough explanation of their quality checks and provides estimates of the accuracy and precision.

and countries any given author has been to or published in we show in Figure 1 their distribution. We see that all three distributions are highly skewed distributions (hence plotted in log-log) with a sharp decline for all values beyond 1. We see that only 10% of authors have at least two countries or 3 cities on their CV, or published at least 8 papers. Similarly, only 1% of authors have worked in at least 3 different countries or 5 cities, or published at least 38 papers.

We analyze the affiliation path of 3,740,187 individuals, for which geo-location data is available in the period 1990 up to 2009. The coverage over time of these authors is available in Figure 1.

To estimate the quality of the researchers — required for Brain Circulation considerations — we augment the publication history with journal impact scores and research field classifications provided by SCIMAGO. SCIMAGO provides access to yearly "impact factor" scores for a large portion of journals indexed in MEDLINE. We use this dataset to proxy the productivity of a scientist by the impact factor of the journal they publish in. SCIMAGO calculates impact factors for journals starting from 1999 and backfills them. For this reason, we do not use data for the brain circulation part of the analysis (Section 2.5) which reaches back several years before 1999 to reduce problems with deviations from the "true" citations per document in the journal. By considering only the period from 1999 to 2009 we have still 2,456,345 Scientists in our dataset, however only for 1,363,280 do we have complete coverage in SCIMAGO. A detailed discussion on how the productivity indicators are constructed is available in Section 2.4. In addition to impact metrics we also use SCIMAGO's journal classification to assign papers to thematic areas.

2.2.2 Methodology

With the extracted publication, we can reconstruct the path for a given author over time as observed by the affiliations on the papers she publishes. In other words, we have a path for author i over several years indicating where she passed through. It might and does happen, that an author has multiple publications in the same year as well as multiple lo-

cations[4]. Here we define what a move is and how we extract it from the empirically observed publication sequences. To determine a move, and just as importantly a non-move, we define mobility by determining the location of an author within a given time window before a year of interest (t) (i.e. the move year) and assess where she is located, in the window after.

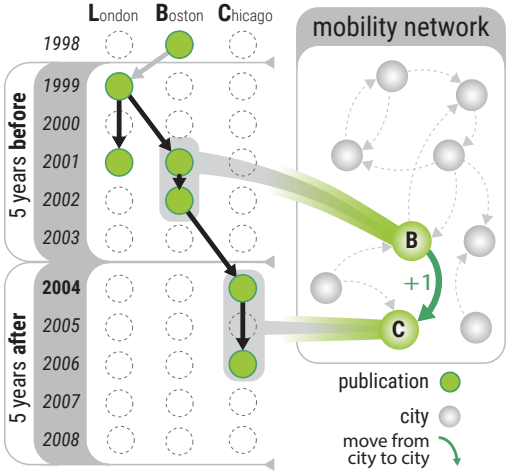


Figure 2: Creating the mobility network from MEDLINE publications. The scientific publications by a single author are illustrated as a sequence of green circles from top to bottom. Each publication has a time (in rows) and location (in columns) associated with it. We take a buffer time (i.e. 5 years) before and after a candidate move from Boston (B) to Chicago (C) in 2004. In this example, we identify Boston as the source, since it is the longest sequence within the window and closest to the end of the move year. Similarly, the destination is Chicago since it is the only observed city in the second window. Each move is tracked in a similar way and added to the mobility network by incrementing the edge weight accordingly.

More specifically to determine the source and destination of a move, for a given time interval we chose a candidate move-year (t) and a number of buffer years (b) around it (see Figure 2). To transform a publication path into a single edge representing a move we proceed e as follows. We

[4]an example of which is available in Table 16 of the Appendix

chose a "move year" t of interest. The move year represents the year around which the decision to move happened. Next, we choose a number of years around t defining two windows: **before** $[t-b, t)$ and **after** $[t, t+b)$. Given these two windows we proceed to determine in which location any given author was before and after. If the locations differ then the author moved, otherwise she stayed.

To determine a unique starting position in window $[t-b, t)$ we choose the longest uninterrupted sequence of locations closest to t. Take for example the observed publication sequence as illustrated in Figure 2. Here we have the publication history $\{B_{1998}, L_{1999}, L_{2001}, B_{2001}, B_{2002}, C_{2004}, C_{2006}\}$, move year 2004 and a buffer of 5 years before and after. The Uppercase letter indicates the city and index the year. To determine the starting location, we take all publications in the interval $[1999, 2004)$ and chose the locations with the longest sequence closest to 2004. In this example, we observe 3 publications in B, but only 2 of these are within the $[1999, 2004)$ window, so we discard B_{1998}. On the other hand, we observe 2 publications in L and one simultaneously with B. According to the aforementioned rule, we chose B as source since it closest to 2004 even though both L and B have 2 observations. As the destination of the move we chose C since in this case it is the only observed location in the window $[2004, 2009)$.

We chose this method, since it discards ambiguous affiliations in publication sequences with spurious affiliations (e.g. multiple affiliations in the same year but either of these appear only once).

This definition allows us to carry out several robustness checks in generating the network. For example, we can increase the number of publications required in each location before and after to reduce the chance that a move was only temporary (e.g. visiting or double affiliations). Similarly, we can restrict the size of the windows, thus requiring that authors have fewer holes in their publication history, however doing so will drop any scientist not publishing at least once in the two periods.

2.3 Descriptive Analysis of Mobility Network

In this section, we offer an overview of several statistic describing the geographic mobility patterns of scientist at international and intercity level as well as an estimation of the centrality of cities within this network. We want to show that mobility does not only have a national component, but that analyzing it at the city level can give important insights into the position of countries within the international innovation system. First, we show, that the most central cities in the international mobility network are US cities, with some minor exceptions. This observation is confirmed by analyzing inflow and outflow patterns. In fact, we find that these super-connected cities source their scientists from a wide range of cities and countries but their outflow is restricted to a smaller set of cities, suggesting that scientists passing through them remain in the core of the network. An analysis of the community structure of the mobility patterns suggests, not only that mobility is significantly influenced by national borders but that shared language can facilitated mobility.

Where we do only provide statistics for one network we refer to the mobility network for the move year 2004 with 5 years of buffer around it. In practice this means that the earliest publications we consider are from 1999. The starting city is determined in the period $[1999, 2004)$ and the destination is determined in the period $[2004, 2009)$. The analysis has been carried out also for 2003 and 2002 with window sizes ranging from 3 to 6, yielding similar results. We use this network because it is the most recent network for which we can be confident to have a good coverage of disambiguated authors and accurate SCIMAGO scores.

2.3.1 City Centralities

Which cities are at the center of the exchange of life scientists and how do different countries fare in this comparison? To answer this question, we look at the 2004 Mobility network. Specifically, we compute several standard network centrality measures to rank the position of cities. We determine which cities are part of highly connected "clubs" (k-core), would be the most likely location to find a scientist moving freely on the network

(i.e. PageRank) and how many cities this city has access to (i.e. in/out-degree)

The k-core is defined as the set of nodes left after removing iteratively all nodes with degree less than k, until the graph is either empty or no more removal is possible. So, for example in the case of an undirected, unweighted graph the 4-core contains all nodes which are connected to at least 4 other nodes which in turn are connected to 4 other nodes with the same property. The procedure filters out nodes which contribute to the degree of other nodes but do not themselves have many connections. This means that at a relatively low k most nuisance nodes (nodes which have few partners overall) are removed. PageRank is a commonly used centrality metric for directed weighted graphs. It estimates how likely a random walker traversing the network is to be found in each node (Page et al., 1998). In the case of a mobility network, the measure can be understood as the stationary probability of a scientist to be found in any given city if she were to move following the strength and direction of the observed moves, with an occasional probability to be "teleported" to a random city.

In Table 1 we report the top 30 Cities as ranked by PageRank centrality along with k-core and degree rankings. The ranking reveals that US cities dominate the mobility network in the life sciences.

Among these top 30 cities only 9 are not US American and only 2 of these are from continental Europe: "Paris, France" and "Berlin, Germany". This ranking does not give a complete picture of the mobility network, but it suggests that cities are an important component. A more detailed analysis of the in- and out-flows (see Appendix, D-core decomposition) highlights the asymmetry in the global intercity exchange. We find that central cities in the US source their scientists from a wide verity of cities but they feed a smaller subset of cities .

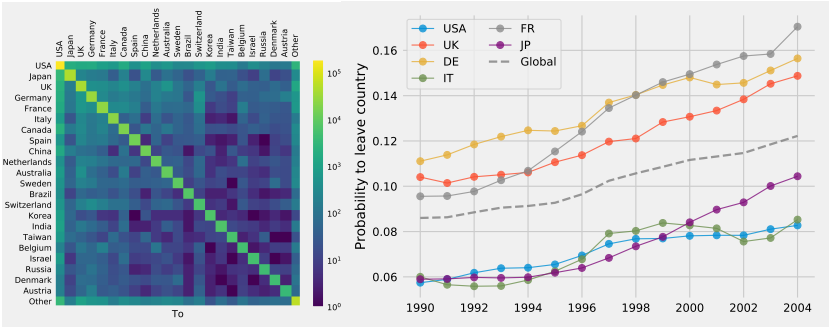
2.3.2 National Border Effects

Co-authorship networks have been found by Hoekman, Frenken, and Tijssen (2010) and Chessa et al. (2013) to be influenced by national borders

	Ranking				
City	k-core	PageRank	in-deg.	out-deg.	deg.
Boston, MA, USA	1	1	1	2	2
London, UK	1	2	2	1	1
New York, NY, USA	1	3	6	4	5
Bethesda, MD, USA	1	4	3	5	4
Paris, France	1	5	5	3	3
Baltimore, MD, USA	1	6	4	7	7
Philadelphia, PA, USA	1	7	7	6	6
Chicago, IL, USA	1	8	9	8	8
San Francisco, CA, USA	1	9	13	18	14
Houston, TX, USA	1	10	8	9	9
San Diego, CA, USA	1	11	11	10	10
Tokyo, Japan	1	12	28	11	16
Atlanta, GA, USA	1	13	10	14	11
Seattle, WA, USA	1	14	12	12	12
Cambridge, MA, USA	1	15	15	15	15
Durham, NC, USA	1	16	18	21	19
Beijing, China	1	17	25	23	22
Toronto, ON, Canada	1	18	16	17	18
Westwood, Los Angeles, CA, USA	1	19	20	33	27
Ann Arbor, MI, USA	1	20	19	20	20
Cambridge, Camb., UK	1	21	16	16	17
Montreal, QC, Canada	1	22	23	28	25
Los Angeles, CA, USA	1	23	25	39	35
Stanford, CA, USA	1	24	22	26	23
Pittsburgh, PA, USA	1	25	23	28	25
New Haven, CT, USA	1	26	28	25	27
Berlin, Germany	1	27	31	31	31
Saint Louis, MO, USA	1	28	21	30	24
Seoul, Korea	1	29	59	70	62
Washington, DC, USA	1	30	35	24	30

Table 1: Ranking of top 30 Cities by centralities sorted by k-core and PageRank for the 2004 mobility network. Members of the EU (except UK) are **bold**

.



(a) (b)

Figure 3: (a) The Country to country mobility flows for the mobility network of 2004 with 5 year of buffer. On the main diagonal, we find the *number* of all scientists who did not leave the country (i.e. the national scientist population). The rows are the source and the columns are the destination, with the color indicating the number. The countries are sorted according the size of their scientist population in the period 1999 to 2004. (b) Probability to leave country for selected countries and global mean (1990 to 2004). Note: the "country" is the country from which the move originates, not necessarily the nationality of the author.

resulting in collaborations being more likely within than across countries. In line with these findings we test the hypothesis, that countries have a stronger within mobility than across.

Figures 3 (a) shows the pattern of cross country mobility in 2004. Clearly most scientists do not leave their country (as indicated by the main diagonal). Note also that certain countries have few exchanges with all other countries, as indicated by having only few off diagonal elements brighter than the rest. This means that while the network is dense (i.e. all major countries have at least one exchange) there are preferences. Note also that the probability to leave the country has increased steadily year by year as can be seen in Figures 3 (b). The global probability to observe a move, i.e. that any given scientist moves abroad if we look at 5 years before and after, has never dipped since 1990. The listed countries fall into two categories, below the global mean and above. With the US, Japan and Italy clearly falling short of the global average, indicating a strong within mobility. Moves originating from the US tend to be mostly within the US. This number has gone from 5% in 1990 to 8.1% in 2004, however compared to France (16.8%) and the global average (12%) it is low. Note however, that scientist based in the US do not leave the country as often as most other countries, but there is a substantial domestic exchange.

The international mobility patterns seen in Figure 3 suggest that international mobility varies by country and that there is more mobility within than across. The notion of "more within" and "less across" is made precise by the measure of *modularity* (Newman and M. Girvan, 2004). At a high-level, modularity is a quality score of how well a given partitioning of nodes (i.e. set of cities) separates nodes which are well connected with each other but have few ties to members of other partitions. More specifically modularity measures the ratio of links falling within a given partition minus the ratio of links we would expect from a random network. A random network in this context is a network, which has the same degree sequence as the observed network, but rewired without regards for any underlying structure (see Newman and M. Girvan (2004) for more detail). Thus, this null model represents a mobility network where scientists move without regard for geographic proxim-

ity or national borders. We estimate the communities by maximizing the modularity of the partition following the Louvain algorithm (Blondel et al., 2008) implemented by Traag (2017).

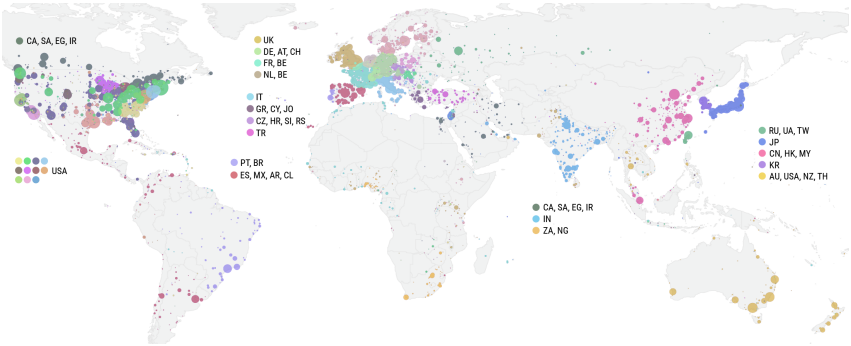


Figure 4: Community structure implied by the 2004 mobility network. Each node is a city and its color indicates to which community it belongs. The size is proportional to the sum of incoming and leaving authors.

If the null hypothesis that scientists move without regard for national borders were true, we should find that the community structure we obtain by maximizing the modularity does not coincide with any geographic or political boundaries.

However, we do find that geography and national borders capture the community structure of the mobility network well (see Figure 4). A breakdown of countries as they fall within the various communities in 2004 is available in the Appendix (Table 17). The communities of the intercity mobility network in continental Europe, is clearly conditioned by national borders. For example, we find that the community to which Italy belongs is composed of 75% Italian cities, 6% US cities and several other minor percentages, the same goes for several other countries, which are the absolute majority within their community. However, the picture changes when looking at North America. Here we also observe a national component in the form of Canada and Mexico being identified as separate communities, but within the US the identified communities are less spatially segregated than in the rest of the world.

Beyond the pure border effect the community structure reveals some additional patterns. We see that countries sharing a language are more likely to fall within the same community. For example, three majority German speaking countries, Germany, Austria and Switzerland are identified as belonging to the same mobility community. Even more strikingly are Spain and Portugal. The two countries share a border but not a language. And we see that they are part of different communities. However, as Table 17 (see Appendix) shows, Portugal and Brazil have a more significant exchange among themselves than Portugal has with Spain even though one is across the ocean and the other a next-door neighbor. Similarly, Spain and Mexico are placed in the same community, both countries share a colonial history and language, as do Portugal and Brazil.

We should note that community detection through modularity maximization may fail to separate communities which are "too small" due to the method's "resolution limit" (Fortunato and Barthelemy, 2007). Ground truth communities, which are not of comparable size to the identified communities may be lumped together with larger communities or split up. In practice this could mean that we have lumped "small" communities together which probably should be kept separate, for example Greece, Cyprus and Jordan are placed in the same community. While Greece and Cyprus share a language the inclusion of Jordan in this community is most likely since Jordan has had an exchange with the other two but was "erroneously" placed in the same community.

2.4 Mapping Brain Circulation

The concept of "Brain Drain", most prominent when discussing the mobility of scientists has been described by Geuna (2015, Ch.1, p.5) as an "unidirectional migration of skilled workers from less developed to more developed countries or regions". However, as Agrawal, Kapur, et al. (2011) and Saxenian (2005) argue, connections between migrant scientists and their home country persist and might facilitate knowledge flows in the opposite direction. Thus, it is more appropriate to talk about brain

circulation.

We present a high-level overview of the flow of "talent/brains" at global scale taking various levels of aggregation into account. Specifically, we want to look at the benefit scientists have from moving along certain paths/dyads and the gain in productivity a city has due to turnover (see Section 2.4.1). Similarly, in Section 2.4.2 we describe at country level how international flows affects the scientific output.

To determine the productivity for any given scientist we use the journal "impact" factor data from SCIMAGO. Specifically, we use the "citations per document in the 2 years before the publication year" of the journal as the measure of quality of scientific output. To avoid inflating the output, we apply a fractional count, whereby any author receives for any paper coauthored with n authors and factor x the fraction x/n. We define several indicators, whose definition and description are summarized in Tables 2, 3 and 5.

To measure productivity, we define two basic measures of scientific output, P_i^θ and r_i^θ, where i is the author and θ identifies in which window (before or after) her publications are aggregated. With $\theta = 0$ indicating the period before the move year t and $\theta = 1$ the period after. Specifically, for every author i we obtain her publication list in the windows $\theta = 0 \equiv [t - b, t)$ and $\theta = 1 \equiv [t, t + b)$ (see Figure 2 window before and after). For each publication authored by i we then obtain the impact of the journal it is published in and divide it by the number of authors on that publication (i.e. fractional count). This yields for each author i a productivity before P_i^0 and after P_i^1. Additionally, to consider that authors might only start their career within the window we normalize this measure, such that it can be interpreted as the impact weighted annual productivity r_i^0. For example, an author with $P_i^0 = 90$ who has started publishing in 1995 when considering the move year 1998 and a 5 year buffer would have a $r_i^0 = 90/\min(5, 1998 - 1995) = 30$ and similarly if the same author had published her first paper in 1990, r_i^0 would be $90/\min(5, 1998 - 1990) = 18$. Similarly, for r_i^1 we divide by the buffer size b since she was by definition active from the beginning of that period (i.e. $90/5 = 18$).

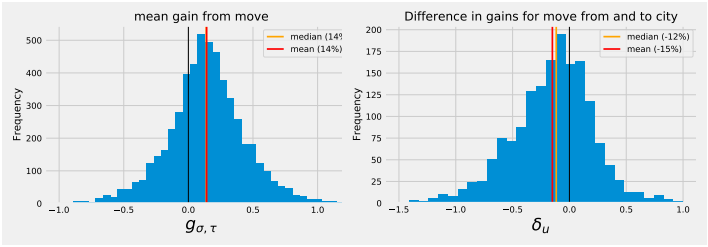
	Definition	Description
t, b, θ		t is the move year, b the number of buffer years around it, $\theta = 0$ is the period before $[t-b, t)$ and $\theta = 1$ after $[t, t+b)$.
$\mathcal{P}_i^\theta$		Set of papers produced by i in period θ
P_i^θ	$\sum_{p \in \mathcal{P}_i^\theta} w(p)$	Impact weighted fractional count of papers for author i in period θ. $w(\cdot)$ returns the impact factor of the journal the paper was published in in that year, divided by the number of authors on the paper
r_i^0	$P_i^0 / \min\{\text{age}_i, b\}$	Annual productivity rate before the move
r_i^1	P_i^1 / b	Annual productivity rate after the move
$\mathcal{S}_{\sigma,\tau}$		Set of authors moving from source city σ to target city τ
$n_{\sigma,\tau}$	$\lvert\mathcal{S}_{\sigma,\tau}\rvert$	Number of scientist moving from σ to τ
$\rho_{\sigma,\tau}^\theta$	$\sum_{i \in \mathcal{S}_{\sigma,\tau}} r_i^\theta / n_{\sigma,\tau}$	Mean productivity rate in period θ for scientists moving from σ to τ
$\Psi_{\sigma,\tau}^\theta$	$\sum_{i \in \mathcal{S}_{\sigma,\tau}} P_i^\theta$	Total output for scientists moving from σ to τ in θ
$\psi_{\sigma,\tau}^\theta$	$\Psi_{\sigma,\tau}^\theta / n_{\sigma,\tau}$	Average output for scientists moving from σ to τ in θ

Table 2: Variables used in Brain Circulation calculations

2.4.1 Intercity mobility Gains

To understand the role and the importance of the cities in the international mobility and brain circulation network we define and compute several indicators of "productivity" gains. We want to quantify which routes/dyads confer the highest productivity gains on the scientist and if cities are able to replace the leaving scientists with incoming scholars.

To quantify and identify the gain a scientist can gain from moving from a given city σ to an other city τ we measure her impact weighted annual productivity before ($\psi^0_{\sigma,\tau}$) and after the move ($\psi^1_{\sigma,\tau}$) and compute the gain (i.e. log of ratio) and obtain $g_{\sigma,\tau}$. This measure represents the average gain scientists moving from σ to τ have experienced. Since a move might be due to productivity considerations and the global scientific output grows year by year we expect the global mean of $g_{\sigma,\tau}$ to be positive. And in fact, we find that on average every move from any city to any other yields a gain of 14% (see Figure 5).



(a) (b)

Figure 5: (a) Distribution of average productivity gains ($g_{\sigma,\tau}$) for observed city to city flows (at least 10 moves, frequencies are dyads). (b) Distribution of gains for incomers net of gains for leavers (δ_u, frequencies are cities with at least 10 in and 10 out moves). Similarly, for δ_u we show the distribution of cities falling within the specified bin.

By plotting $g_{\sigma,\tau}$ on a map (see Figure 7 and 6) and coloring the links according to its distance from the median, we see that most of the moves are green. Since there are as many red edges as green ones on the map, this implies that shorter moves, too small to be seen on the map, are below the median (i.e. red). We also notice that moves from the east to

the west (edge direction is clockwise), especially the US are green, while moves from west to east are red (i.e. gains below the median).

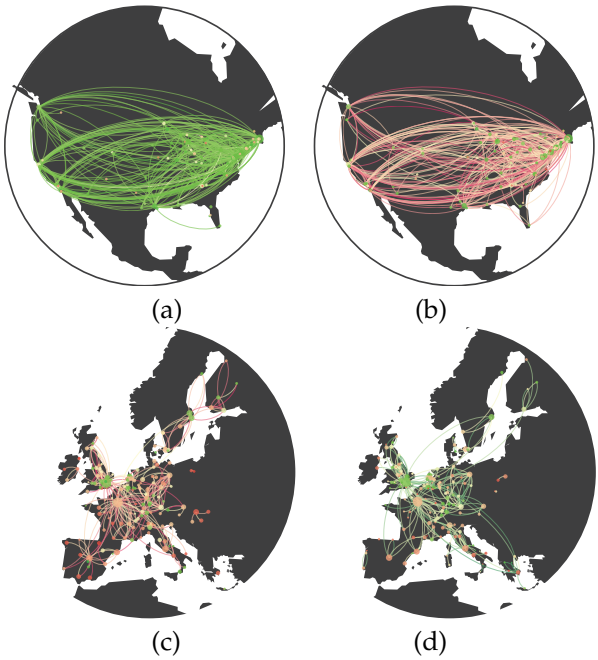


Figure 6: Brain circulation network zoom USA and EU. Here we show only city to city connection within the USA and EU respectively. Each arc represents an observed movement of at least 15 people. Locations with neither in nor outflow or a scientist population of less then 50 are hidden. (a) and (b) show flows where $g_{\sigma,\tau}$ is above the median and (b, d) below. The colors and direction of flows are the same as in Figure 7.

Additionally, we can look if there is an imbalance in the two possible direction the flow could take place, i.e. $\xi_{\sigma,\tau} = ||g_{\sigma,\tau} - g_{\tau,\sigma}||$. Note that we can only compute this value for actually observed dyads. If the gain in any direction would be the same then $\xi_{\sigma,\tau}$ would be 0, however we find that this is not the case as it has a mean of 14%. This fact points to an imbalance in the direction of travel. We would expect that the direction with the higher gain to be chosen more often. However, we do not find that the strength of the flow ($n_{\sigma,\tau}$) is correlated with the mean gain ($g_{\sigma,\tau}$).

This is confirmation of the our visual intuition of red vs green edges on the circulation map (Figure 7). This is an indication that there is not only a supply side (i.e. scientists choose were to go) but also a demand side to scientist intercity mobility. Cities, in the form of universities and research centers, are discerning who they hire or reject.

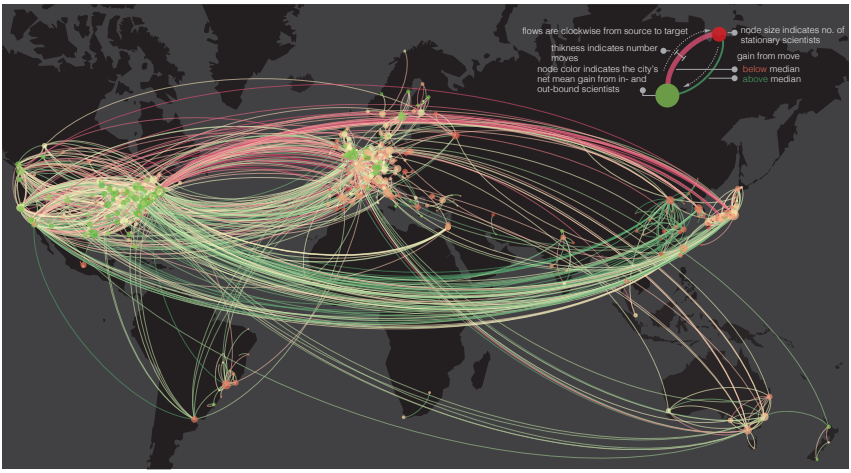


Figure 7: Brain circulation network for the year 2004. The brain circulation map shows the intercity movement for cities (nodes) which have at least 15 authors incoming or 15 leaving and a stationary scientist population of at least 50. The thickness of the edges is proportional to the natural log of number of people moving between two cities. Locations with neither in nor outflow are dropped. Link colors show the average net gain ($g_{\sigma,\tau}$). Red colors indicates moves which are below the median of the shown links (14%) yellow close and green above. Node colors indicates the difference in gain for incoming and leaving scientist δ_u (see Section 2.4.1).

To determine if on average incoming scientists gain more than those they replace, an indicator that working in this city confers on the incoming scientist a large boost we look at δ_u (shown as node colors in Figure 7) and shown for a subset of global cities in Table 4. The global mean of this value is -25%. In other words, on average the gain from moving to any given city is less than leaving it. For example, a move to "Boston, MA, USA" from anywhere confers on the scientists a 28% gain

	Definition	Description
$g_{\sigma,\tau}$	$\log(\rho^1_{\sigma,\tau}/\rho^0_{\sigma,\tau})$	Average productivity gain for scientists moving from σ to τ
$\xi_{\sigma,\tau}$	$\|\|g_{\sigma,\tau} - g_{\tau,\sigma}\|\|$	Absolute Difference in gains for scientists moving from σ to τ and vice versa.
δ_u	$g_{\star,u} - g_{u,\star}$	Difference between the productivity gain from moving to u and leaving it
Γ_u^{move}	$\log(\Psi^1_{\star,u}/\Psi^0_{u,\star})$	Increase in output for incoming relative to leavers
Γ_u^{stay}	$\log(\Psi^1_{u,u}/\Psi^0_{u,u})$	Increase in output for stationary scientists

Table 3: City level Brain Circulation indicators

but anyone leaving for any other city gains only about 1%, which gives us $\delta_{\text{Boston}} = 0.27$. This value is depicted in Figure 7 and Figures 6. Here we see that the US contains several cities, which have a positive δ_u, while Europe has mostly negative δ_u.

These measures (i.e. g and δ) are interesting to the scientist making the decision to relocate. However, cities have other priorities, i.e. increase scientific output. To quantify if cities benefit from the international exchange we look at two indicators Γ_u^{stay} and Γ_u^{move}. Γ_u^{stay} gives us for city u the increase in total output for stationary scientists. In other words, it measures the percentage increase in total scientific output for scientists who do not move. And Γ_u^{move} gives us the growth in total output coming from new scientist in period 1 relative to the output of the scientist who did leave in period 0. If $\Gamma_u^{\text{move}} > \Gamma_u^{\text{stay}}$ for a city u then the mobile scientist where able to produce enough scientific output to cover their predecessors and contributed positively to the total output growth of the city. From the histograms in Figure 8 we see that on average this is true ($\Gamma_u^{\text{move}} = 29\%$ and $\Gamma_u^{\text{stay}} = 23\%$). In fact, we see in Table 4 $\Delta\Gamma_u$, the difference between growth from mobile scientists and the growth due to stationary scientists. The listed cities are the most central cities in the mobility network as identified in Section 1. We see again that US cities are able to manage the turnover better than central European cities, such

as Paris, and Berlin. However, within the US there are differences, with "Bethesda, MD, USA" for example being able to replace their scientific output with new scientists better than "Boston, MA, USA". This does not necessarily mean that they loose out, since these cities have a prolific stationary scientist populations, however it highlights some cities which are able to manage the turnover better than others.

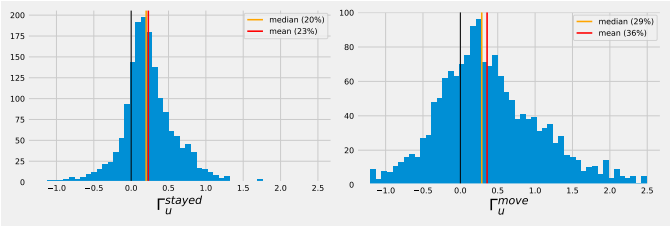


Figure 8: Distribution of gain indicators for the global network. The distributions shows statistics for routes which had at least 10 moves along them.

2.4.2 National Gains

Brain circulation is a major concern at country level and we can estimate the contribution to the growth of the national innovation systems from international mobility, but also domestic mobility. This allows us to compare knowledge output (change in total impact weighted output) across countries and identify which countries were the primary direct beneficiaries of international mobility.

The output produced within a country can be accounted for in the following way. Knowledge produced by authors staying in their city (S) moving domestically (D), *coming in from abroad* (I) and *leaving the country* (L). The total output for a given time period within a country before A^0 and after A^1 are given by $A^0 \equiv S^0 + D^0 + L^0$ and $A^1 \equiv S^1 + D^1 + I^1$ respectively. Note that in A^0 the output contains the production of those individuals who will leave the country L^0 in the second period and A^1 the production of those that will come in the second period I^1. Based on this breakdown we can define indicators identifying the growth due to the three types of scientists. Overall growth ν_A for the country, ν_S

City	Γ_u^{move}	Γ_u^{stay}	$\Delta\Gamma_u$	$g_{\star,u}$	$g_{u,\star}$	δ_u
Tokyo, Japan	-0.17	0.13	-0.31	0.04	0.16	-0.12
Saint Louis, MO, USA	-0.16	0.06	-0.22	0.20	-0.04	0.24
Paris, France	-0.08	0.11	-0.19	0.04	0.14	-0.10
Berlin, Germany	0.13	0.24	-0.10	0.12	0.20	-0.08
London, UK	0.04	0.12	-0.08	0.21	0.13	0.08
Washington, DC, USA	0.08	0.16	-0.07	0.13	0.08	0.05
New York, NY, USA	0.00	0.04	-0.04	0.20	-0.02	0.22
New Haven, CT, USA	-0.01	0.01	-0.02	0.28	-0.01	0.29
Chicago, IL, USA	0.08	0.09	-0.01	0.21	0.04	0.17
Stanford, CA, USA	0.01	-0.02	0.03	0.18	0.07	0.11
Seoul, Korea	0.62	0.58	0.04	0.23	0.51	-0.28
Boston, MA, USA	0.17	0.13	0.04	0.23	0.01	0.22
Cambridge, MA, USA	0.08	0.01	0.07	0.09	0.16	-0.06
Montreal, QC, Canada	0.19	0.11	0.08	0.11	0.08	0.03
Westwood, Los Angeles, CA, USA	0.16	0.07	0.09	0.23	0.11	0.12
Philadelphia, PA, USA	0.17	0.07	0.09	0.20	0.01	0.18
San Francisco, CA, USA	0.11	0.01	0.10	0.05	0.01	0.03
Toronto, ON, Canada	0.20	0.10	0.10	0.16	0.07	0.09
Cambridge, Cambridgesh., UK	0.20	0.10	0.11	0.17	0.03	0.15
Seattle, WA, USA	0.22	0.09	0.13	0.15	0.10	0.05
Baltimore, MD, USA	0.23	0.08	0.15	0.25	0.06	0.19
San Diego, CA, USA	0.08	-0.08	0.16	0.27	-0.13	0.40
Ann Arbor, MI, USA	0.29	0.12	0.16	0.23	0.07	0.16
Houston, TX, USA	0.25	0.07	0.18	0.24	-0.02	0.26
Los Angeles, CA, USA	0.25	0.07	0.18	0.09	0.13	-0.05
Durham, NC, USA	0.34	0.06	0.28	0.16	-0.01	0.17
Bethesda, MD, USA	0.39	0.07	0.33	0.23	-0.04	0.27
Pittsburgh, PA, USA	0.49	0.13	0.35	0.30	0.17	0.13
Beijing, China	1.32	0.87	0.45	0.36	0.75	-0.39
Atlanta, GA, USA	0.64	0.11	0.52	0.22	0.19	0.02

Table 4: The Γ_u values (growth due to turnover) for the 30 most central cities as listed in Section 2.3.1. The indicator of scientists gains from a move there u ($g_{\star,u}$), u $g_{u,\star}$, δ_u are also listed. The cities are ordered by Γ_u

growth due to stationary scientists, ν_D growth due to nationally mobile scientists and most relevant for the brain circulation discussion ν_I, the gain due to international turnover. Additionally, to have a indication of the generational turnover we also report the mean age of incoming ($\overline{\text{age}}_I$) and leaving scientists ($\overline{\text{age}}_L$). These indicators are defined in more detail in Table 5. The results for the largest countries in the dataset for the interval before ($\theta = 0 = [1999, 2004)$) and ($\theta = 1 = [2004, 2009)$) are reported in Table 6.

	Definition	**Description**
S^θ	$\sum_{i \in \mathcal{S}_{u,u}} P_i^\theta$	Output of stationary scientist in domestic city d
D^θ	$\sum_{i \in \mathcal{S}_{u,d}} P_i^\theta$	Output of scientist moving from domestic city u to domestic city d
I^θ	$\sum_{i \in \mathcal{S}_{f,d}} P_i^\theta$	Output of scientist coming from foreign city f to a domestic city d.
L^θ	$\sum_{i \in \mathcal{S}_{d,f}} P_i^\theta$	Output of scientist leaving the country for a foreign city f.
A^0	$S^0 + D^0 + L^0$	Total Output in the country before the move year
A^1	$S^1 + D^1 + I^1$	Total Output in the country after the move year
ν_A	$(A^1 - A^0)/A^0$	National output growth of output
ν_S	$(S^1 - S^0)/S^0$	Output growth of from stationary scientists
ν_D	$(D^1 - D^0)/D^0$	Output growth from domestically mobile scientists
ν_I	$(I^1 - L^0)/L^0$	Output growth from international exchange
$\overline{\text{age}}_I$		Average age (years from first publication) for incoming
$\overline{\text{age}}_L$		Average age of leaving scientists
$\Delta\overline{\text{age}}$	$\overline{\text{age}}_L - \overline{\text{age}}_I$	Age difference between Leaving and Incoming scientists

Table 5: Country level Brain Circulation indicators

From Table 6 we see that the USA, with international turnover, has increased its scientific output by 14% overall (ν_A). Among all the three types of scientists, the growth due to new arrivals (I) is largest, 61%. This also compared to growth due to stationary and domestically mobile scientists (9% and 17%, respectively). The stationary ($S = 0.71$) and domestically mobile scientists ($D = 0.23$) represents the largest portion of the population, however international exchange has had a net benefit on the output growth. However, not all countries have a higher than average growth from international mobility (i.e. $\nu_I > \nu_A$). This suggests that not all countries have the same direct gain from international exchange. Moreover, looking at the age differential between incoming and leaving scientists we see that the average scientists moving to the US (6.7) are younger than the ones they replace (7.9). This means that the US has been able to rejuvenate their scientific labor force, while simultaneously increasing their scientific output.

Clear beneficiaries of international exchange beyond the US, are Australia, Canada, Spain and Switzerland with $\nu_I > \nu_D > \nu_S$ and with a substantial contribution (i.e. more than 10% of output share). Argentina for example, has experienced only 5% output growth, the second lowest in the list and has lost 48% of output due to international exchange. Japan is also striking, the scientist leaving are young (6.7) compared to the scientists moving to Japan (8.6). This is accompanied by a negative loss from international exchange -21%. All other countries in this comparison either loose out or the effect is ambiguous. What is clear, is that international exchange as measured by direct scientific output does not benefit everyone in the same way.

	ν_A	ν_S	ν_D	ν_I	$\overline{\text{age}}_I$	$\overline{\text{age}}_L$	$\Delta\overline{\text{age}}$
Argentina	5%	23% (0.71)	79% (0.02)	-48% (0.27)	8.98	8.13	0.85
Australia	32%	28% (0.80)	24% (0.07)	56% (0.13)	7.34	6.93	0.40
Austria	18%	22% (0.77)	-7% (0.04)	9% (0.18)	7.35	7.34	0.01
Belgium	23%	28% (0.79)	15% (0.07)	0% (0.14)	8.23	8.21	0.02
Brazil	46%	53% (0.78)	57% (0.08)	1% (0.14)	7.66	6.88	0.77
Canada	20%	16% (0.73)	19% (0.11)	36% (0.16)	7.14	7.08	0.06
China	41%	117% (0.62)	212% (0.15)	58% (0.23)	4.52	3.36	1.16
Denmark	18%	18% (0.81)	18% (0.06)	17% (0.13)	8.00	7.93	0.07
Finland	3%	8% (0.77)	2% (0.10)	-21% (0.13)	9.06	8.92	0.14
France	10%	14% (0.76)	12% (0.09)	-10% (0.16)	8.14	7.47	0.66
Germany	16%	18% (0.66)	16% (0.19)	8% (0.15)	7.29	7.31	-0.02
India	42%	65% (0.66)	73% (0.10)	-33% (0.24)	7.91	5.64	2.27
Israel	14%	22% (0.74)	27% (0.09)	-23% (0.17)	9.67	7.12	2.54
Italy	32%	32% (0.82)	31% (0.10)	36% (0.08)	8.49	8.00	0.49
Japan	9%	12% (0.66)	13% (0.24)	-21% (0.10)	8.57	6.71	1.85
Korea	71%	66% (0.63)	87% (0.19)	76% (0.18)	5.55	4.70	0.85
Netherlands	26%	28% (0.74)	18% (0.13)	19% (0.12)	7.57	7.86	-0.29
Russia	13%	27% (0.73)	41% (0.01)	-27% (0.26)	8.03	7.37	0.66
Spain	35%	35% (0.81)	21% (0.07)	44% (0.11)	7.68	6.45	1.23
Sweden	10%	18% (0.74)	13% (0.08)	-25% (0.18)	8.47	8.07	0.40
Switzerland	14%	8% (0.67)	2% (0.08)	33% (0.24)	7.53	7.94	-0.42
Taiwan	37%	37% (0.74)	51% (0.15)	17% (0.11)	6.90	7.94	-1.04
UK	16%	14% (0.74)	25% (0.13)	17% (0.13)	7.16	7.27	-0.12
USA	14%	9% (0.71)	17% (0.23)	61% (0.06)	6.69	7.89	-1.20

Table 6: National scientific output growth figures for selected countries (at least 3,000 stationary scientists in 1999–2004). See Table 5 for definitions. In parentheses, the proportion of the total output in the first period (A^0) by category.

2.5 Preference for Global cities: Regression Analysis

The topological, geographic and impact gain analysis in the previous sections suggest that there is a spatial component to the mobility patterns of scientist, that certain cities are more central within this network and that not all moves offer the same gain for a mobile scientist. We test the hypothesis that more central cities (k-core, page rank or degree), not only attract a lot of scientist but attract more productive scientist. If it is indeed true that more productive scientists move preferentially to more central cities, we expect productivity to be positively correlated with the centrality of the destination. That is, after we control for various factors and account for selection bias in our data (i.e. not all scientists move), we should find that scientific output before the move is indicative of a move to a more central city.

We estimate a Heckman two stage regression to account for the fact that the majority of scientists do not move and as such we would not observe a change in the centrality of their relocation choice. The focal variable of this analysis are "Productivity^0" and "Centrality^1". The variable "Productivity^0" measures how prolific a scientist was before she moved. Specifically, this is the log of r^0, which is described in detail in Section 2.4.1. All other controls used in the regression are listed in Table 7.

Note that not all countries and fields are present in sufficient number or are relevant for the analysis. For this reason, we drop an author from the dataset if one of the following applies: (1) the author is a member of a country which has less than 500 scientists or (2) the author publishes predominantly in fields for which there are less than 500 papers in the period. These are mostly fields which are not considered life-sciences but are in MEDLINE (e.g. Economics).

As the measure of "Centrality^0" and "Centrality^1" we use PageRank since it is proportional to the stable distribution of a random walker on the observed mobility network. The PageRank of a city can be interpreted as the null model where the relocation choice is simply done at

Variable	Description
Centrality^0	The centrality of the source city (i.e. PageRank)
Centrality^1	The centrality of the destination city
Moved	1 if the author moves to a different city, 0 otherwise.
Productivity^0	The log of the annual productivity rate r^0(see Section 2.4.1)
$\Pr$(Move other Fields)	The proportion of authors moving away from the source city which do not publish in the same field as the focal researcher.
$\mathbb{I}$(Year)	The move year (i.e. 2000, 2002, 2004)
$\mathbb{I}$(Age Group)	Age is measured as the difference in years from first publication to the move year. The age-groups are split such that the cohorts are of comparable size.
Intermove	The years between the last observation in the first period and the first in the second.
$\mathbb{I}$(Country)^0	The country in which the author was working in period 0
$\log$(km dist)	The log of the distance from source to target city in kilometers
$\mathbb{I}$(Field)^0	The SCIMAGO thematic area the author publishes most in period 0

Table 7: Regression Variables

random without regards for productivity, distance or other features we assume are important, but following the empirically observed flows between cities. As a robustness check we also estimate the model for k-core and degree centrality, which yield qualitatively similar results (see Appendix Table 18 and Table 19). The dataset is constructed by combining three mobility network (2000, 2002 and 2004) all with a buffer of 5 years.

To estimate the Heckman model and correct for self-selection of scientists into the population of mobile scientists, we use as an exclusion restriction the probability to leave the city for all scientist not belonging to the focal field ($\Pr(\text{Move other Fields})$). For example, for authors predominantly publishing in "Embryology" the probability to move is computed as the fraction of scientists leaving the city in the same period, but who do not publish on "Embryology". The rational to use this variable as an exclusion restriction is that if we observe a lot of mobility originating from a city, it stands to reason that it increases the propensity of the focal author to move as well. By excluding the focal field, we reduce the likelihood that the focal author is influenced by competition or imitation of peers working in the same field.

$$\begin{aligned}\text{Moved}_i =&\gamma_0 + \gamma_1 \text{Centrality}_i^0 + \gamma_2 \text{Productivity}_i^0 + \\ &\gamma_3 \Pr(\text{Move other Fields})_i + \gamma_4 \text{Intermove}_i + \\ &\boldsymbol{\gamma_a} \mathbb{I}(\text{Age Group})_i + \boldsymbol{\gamma_y} \mathbb{I}(\text{Year})_i + \boldsymbol{\gamma_f} \mathbb{I}(\text{Field})_i^0 + \boldsymbol{\gamma_c} \mathbb{I}(\text{Country})_i^0 + \\ &\boldsymbol{\gamma_{pa}} \mathbb{I}(\text{Year})_i \times \text{Productivity}_i^0 + \\ &v_i\end{aligned} \tag{2.1}$$

$$\begin{aligned}\text{Centrality}_{1i} =&\beta_0 + \beta_1 \text{Centrality}_i^0 + \beta_2 \text{Productivity}_i^0 + \\ &\beta_3 \text{Intermove}_i + \\ &\boldsymbol{\beta_a} \mathbb{I}(\text{Age Group})_i + \boldsymbol{\beta_y} \mathbb{I}(\text{Year})_i + \boldsymbol{\beta_f} \mathbb{I}(\text{Field})_{0i} + \boldsymbol{\beta_c} \mathbb{I}(\text{Country})_i^0 + \\ &\boldsymbol{\beta_{pa}} \mathbb{I}(\text{Year})_i \times \text{Productivity}_i^0 + \\ &\log(\text{km dist}) + u_i\end{aligned} \tag{2.2}$$

	Pr(move)		PageRank destination	
PageRank source	-5.262	(-0.23)	0.0166	(0.76)
Productivity^0	0.0209***	(4.28)	0.000146***	(4.16)
Pr(Move other Fields)	1.564***	(6.10)		
log(km dist)			0.000108*	(2.26)
2002	0.0289***	(6.05)	0.0000352	(0.84)
2004	0.0331***	(4.42)	0.0000375	(0.70)
2002 $\times$ Productivity^0	0.00269	(0.86)	0.0000236*	(2.35)
2004 $\times$ Productivity^0	-0.00455	(-1.35)	0.0000100	(0.68)
inter-move	0.109***	(64.38)	-0.0000751***	(-4.25)
Constant	-1.754***	(-4.05)	0.00289***	(3.77)
Year Effects	Yes		Yes	
Origin country effects	Yes		Yes	
Age effects	Yes		Yes	
Field effects	Yes		Yes	
Observations	1,363,280		433,023	
tanh(ρ)	-0.12 (-4.84)	log(σ)	-5.73 (-35.80)	
Log pseudo-likelihood	1,000,412			

Table 8: Regression results for mobility and relocation choice. Results of the Heckman two stage regression for the PageRank. The standard errors have been clustered at source city for the first stage (Pr(move)) and on the destination city for the second stage (i.e. the centrality of the destination)

In the first stage (2.1) we estimate the probability that a given author decides to relocate And in the second stage (2.2) the PageRank of the destination is estimated conditional on observing a move.

2.5.1 Regression Results

The results of the regression are shown in Table 8. We find that in the second stage the propensity to move to a more central city is positively correlated with "Productivity^0". This confirms our hypothesis that controlling for various factors, more prolific scientists (before) tend to move to more central locations. However, this effect changes with age, with young scientists having a substantially higher propensity to move to a

more central city than more senior scientists (see Figure 9).

With regard to the first stage, note that $\Pr(\text{Move other Fields})$ is positive. This means that for any given location the probability to observe a move is positively correlated with the probability to move of other scientists, not working in the same field. So, in fact, we do find that the exclusion restriction has the desired sign. The probability to observe a move (see column $\Pr(\text{move})$) does not depend on the PageRank of the source city. However, we do find that Productivity^0 controlling for various factors has a significantly positive effect on the probability to move. This effect holds across centrality measures (see Appendix Tables 18 and 19).

From Section 2.3.2, we have observed an increased tendency to move abroad over the years. This is also confirmed by the increasing propensities to move (i.e. 2000, 2002, 2004). However, the influence of "Productivity^0" has remained constant across snapshots. Additionally, we find that having larger holes in the publication history (i.e. long Intermove) is strongly indicative of observing a move. This is to be expected since the longer we do not have a signal of presence the chances of finding a scientist again in the same location decrease. Additionally, a larger hole in the publication history means that the starting and destination locations are weak signals of actual presence and could have been spurious.

We can also observe a cyclical pattern in the probability to move by looking at the marginal effects by age-group (see Figure 9). In the years after the first publication the marginal probability to move increases but declines after about 5 years. At country level we find also differences in the propensity to move (see Figure 10). The US is identified as having the most mobile scientist population, followed closely by UK, Switzerland and Germany. Note that while internationally mobility from the US is low, as we have seen in Section 2.3.2, the overall mobility is high, which implies that most of the mobility is domestic, this is also confirmed by the analysis on country gains. We also find that the probability to move by field of research "$\mathbb{I}(\text{Field})$" varies greatly (see Figure 33), with "Physics and Astronomy" being considerably more mobile than fields such as Dermatology.

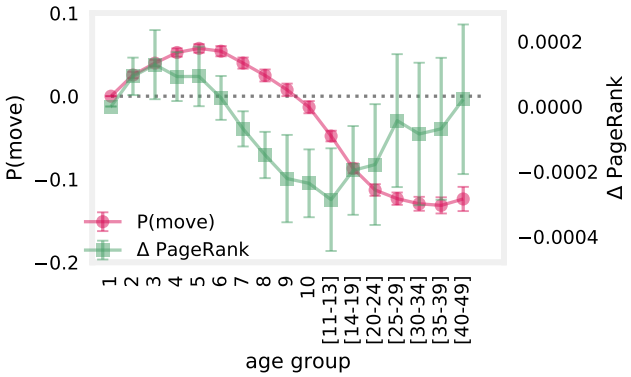


Figure 9: Probability and centrality gain by age This figure shows the marginal effects in probability to move compared to an author with age=1 (a year after the first publication in MEDLINE) and the marginal PageRank increase in destination due to age. An author with a 9 years career has the same probability to be observed moving as an author at the beginning of her career, however the PageRank of the chosen destination will be on average lower (error bars indicate 95% confidence interval).

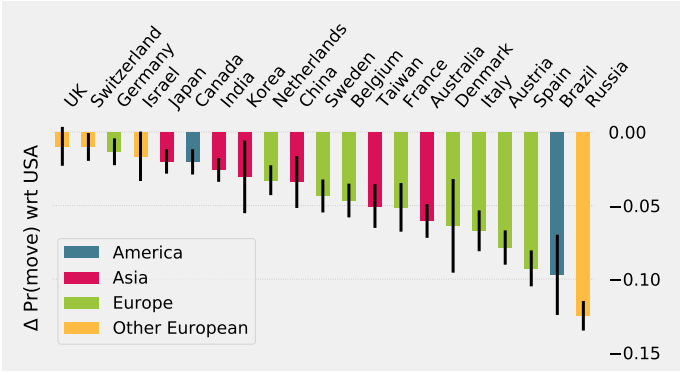


Figure 10: The figures shows the marginal effect on the probability to move compared to the base case USA and the 95% confidence interval (black bars). A negative value such as Taiwan (-5%) means that keeping everything else fixed, a scientist in Taiwan is 5% less likely to move than a colleague in the US. Only countries are shown here for which we have observed at least 3,000 scientists in the country in the period 2000–2004.

Star scientists, (i.e. extremely productive researchers) are a major topic in the mobility of researchers (Azoulay, Zivin, and Wang, 2010; Enrico Moretti et al., 2017; Trippl, 2013; Schiller and Diez, 2012). It could be the case that the patterns that we have estimated here are not representative of this elite group. We carry out the analysis we have done above also for the top 10% of scientists in terms of productivity before the move (Productivity^0). The results are shown in the Appendix in Table 20. We find that the results are identical to the main analysis, except for one key difference. Star scientists, the more productive they become the less likely they are to leave their city. In part this might be due to the fact that they are more advanced in their career and survived for longer in academia then the general life scientist population, thus are on average older. However, this might also suggest that star scientists are embedded in their local innovation system (i.e. they have many important local collaborators) and thus more reluctant to leave[5].

As an additional robustness check to alleviate concerns of imitation and the choice to relocate being influenced by collaborators, we subsample our dataset to 10% of all observations. By doing so we reduce the chance that we pick up groups of people, who are likely to influence each other. The PageRank regression for a subsample of 10% is available in the appendix Table 21. The results are, except for the confidence bounds, identical to the main analysis.

2.6 Discussion and Conclusions

Our results highlight several features of the mobility patterns of life scientists in the period 1999 to 2009. In general, we find that not all cities are equal and borders do matter, however shared language can reduce barriers. We find that not all countries and cities benefit equally from intercity and international exchange. It is apparent that for the analyzed period European cities are not well represented within the global life sciences research system. Moreover, gains in national scientific output, as

[5]This hypothesis could be test by controlling for the number of local and remote collaborators, a possible future extension.

highlighted by the output growth due to international turnover, do not provide a clear signal that international exchange is unequivocally beneficial to all participants. The results clearly point to the USA being a prime beneficiary, which according to the data can attract young and prolific scientists.

This study makes four contributions. First, it introduces a novel approach to extract mobility networks from bibliographic data and augments it with quality indicators. Second, it characterizes the international flows of life scientist highlighting the importance of national barriers. Third, it quantifies the gains from mobility to scientists, cities and countries. And finally, it offers evidence that young prolific scientist move to global cities early in their career.

Our study has several strengths. We are able to reconstruct intercity mobility networks for specific timer intervals, making it potentially useful for event studies, although here we have focused primarily on the 2004 cross section. The dataset has an extensive coverage of life scientists spanning multiple countries, career stages and productivity levels (i.e. not only star scientists). However, it is not without limitations. First, we do not have detailed personal information on the scientists such as gender, birth date or citizenship (only the origin of mobility, which may not coincide). This information is available in smaller but more focused datasets such as the ones used by (Franzoni, Scellato, and Stephan, 2012; Graf and Kalthaus, 2018). Second, we are restricted to 2009 by AUTHORITY, making the findings less current than we would like, however a more recent high quality MEDLINE author disambiguation could alleviate this problem. And third, this dataset covers primarily life scientists, omitting a large chunk of potentially relevant disciplines.

In this work, we have limited ourselves to a descriptive analysis of the mobility network, omitting causality claims. However, the richness of the dataset makes it potentially useful for use in determining causal relocation factors. The global nature and good temporal coverage means that several natural experiments can be identified, which can help to isolate the determinants of mobility. An example of this, is the estimation of the impact of stem cell legislation in the US on stem cell scientist mo-

bility (US states offer various degrees of support) which we will discuss in Chapter 3. Similarly, the effect of regional projects (e.g. opening a new research campus), aiming to improve scientific output or innovation, can be quantitatively analyzed. This dataset, in conjunction with natural language processing techniques and text mining, can also be used to follow the mobility and diffusion of new ideas and concepts in the life sciences. By estimating the relative importance of mobility and collaboration research policies optimizing diffusion could be devised. Moreover by exploiting the available information on collaborations and the presented data on mobility we can also estimate the propensity of scientists to remain in contact with their home-country and city. This approach allows us to replicate and build upon the analysis on the same topic by Agrawal, Cockburn, and McHale (2006).

In conclusion, this was has described a method to extract mobility networks from bibliographic data, used the resulting mobility networks to characterize the mobility patterns and output gains of life scientists at city, intercity and national level as well as provided evidence that prolific scientists gravitate preferentially towards global cities early in their career.

Chapter 3

The Impact of US State and Federal Legislation on the Geographic Preferences of Stem Cell Researchers

This chapter is a joint work with Prof. Massimo Riccaboni

Human Embryonic Stem Cell (hESC) Research is a controversial research topic, which has lead various countries to either heavily regulate or ban it outright. In the US, various laws have been passed at federal level, and several more at state level, restricting hESC research even further, adopting a more supportive approach or outright supplanting federal funding restrictions. The legislative events at federal and state level offer a unique opportunity to evaluate how the diverging interests at the two levels counteract or amplify each other. In fact, the differential impact of federal and state hESC law, is a prime example of the "sates as laboratory" view of state legislation by Supreme Court Justice Brandeis. Therefore our analysis, beyond the specific issue of hESC research laws can also be viewed as case study of how state laws have been used to counterbalance federal state laws.

An open question regarding the early years of this technology in the US is the impact of various federal and state legislative measures on the geographic preferences of hESC researchers.

With the extensive dataset on individual level scientist mobility obtained in Chapter 2, we estimate the impact of several policies on the relocation choice of US hESC researchers over the period 1998 to 2008. This period covers the early beginnings of the field up to the Bush years, which saw a federal moratorium on funding. 2001 to 2006 was according to Acosta and Golub (2016) a period in which the US Stem Cell laws were forged and influence to this day the research landscape. We look specifically at the US, because within its borders, several state and federal laws have been enacted to regulate this nascent field and the extant literature suggests that restrictive interventions have stymied research (Owen-Smith and McCormick, 2006) and caused hESC researchers to move abroad and supportive interventions to increase scientific output (Alberta et al., 2015). These state legislative changes are a source of exogenous shocks we can exploit to identify, if researchers voted with their feet. In addition, the US has a large resident scientist population, which allows us to identify several likely affected scientists as well as comparing their choices against a believable control group. Most importantly however, we have seen in Chapter 2, that the US is marked by a significant within mobility. This means that relocation is not uncommon and we can exploit this variability to study relocation choices specifically and geographic preferences more generally.

Specifically, with this work we want to answer an open question regarding the impact of state and federal level legislation on the mobility of affected researchers. Aaron D. Levine (2012) argues that hESC Researchers express a preference to relocate in US States, which are perceived as being less restrictive. The study, based on surveys, suggests that expressed preferences are a strong indication that the participants would act upon that preference. More specifically the authors find that a state's stance on hESC correlates with their ranking in their survey.

A similar policy shock as hESC support, is the USA PATRIOT Act[1],

[1] "Uniting and Strengthening America by Providing Appropriate Tools Required to In-

enacted by President George W. Bush in response to the 9/11 attacks. This measure required strict background checks and increased the security requirement for labs entrusted with the research of "select agents"[2] with the overall effect of reducing the number of labs involved in the research and the people given clearance interact with the substances. Dias et al. (2010) finds based on scientific output that research has suffered in this area. A collateral effect of this measure as hESC legislation beyond the immediate reduction of scientific output, is the induced mobility of the researchers. The overall reduction of research is given by the combined effect of (1) adversely affected researchers staying with their institution and reduce their work on the subject or stopping altogether, or (2) move away. In this work, we want to explore also the contribution of mobility.

With this work, we want to assess two general questions. Does the level of support correlate with more hESC research and its attractiveness for this field? And second, how did major federal and state legislation shape the mobility patterns of these researchers.

This chapter is organized as follows. In Section 3.1 an overview of legislation in general and the US in particular is offered. In Section 3.2 the datasets are introduced and the process by which we identify affected scientists. Then in Section 3.3 we explore how supportive legislation has affected the geographic distribution in general. We analyze two important legislative shocks (i.e. Bush Ban, Proposition 71) in Section 3.4. And finally, we discuss the implications of these findings for Stem Research in the US in Section 3.5.

3.1 Stem Cell Legislation in the US

In 1981 scientists were first able to derive embryonic stem cells from mice embryos (Department of Health and Human Services, 2015) marking the early beginning of stem cell research. The cloning of mammalian cells by

tercept and Obstruct Terrorism Act of 2001"

[2]Pathogens and toxins listed by the US government that pose a severe threat to public health and safety (Dias et al., 2010)

Wilmut et al. (1997) in 1997 and the derivation of the first hESC lines by Thomson (1998) highlighted the possibilities and ethical implications of the technology. The source of the ethical conundrum lies in the process by which hESC lines are derived, which involves the destruction of human embryos. States and the federal government have since stepped in to regulated the field (Aaron D. Levine, Lacy, and Hearn, 2013).

Stem Cells are cells, which have not yet specialized to become more specialized types of cells (e.g. a muscle, red blood or brain cells) but are still able to do so and can renew and divide themselves (Department of Health and Human Services, 2015). We distinguish two types, of stem cells *embryonic* and *adult*. Both can become more specialized however *embryonic* cells are thought to be more "versatile" and are far easier to be cultured while adult stem cells are more difficult to obtain in large numbers, something that is necessary for replacement therapies (Department of Health and Human Services, 2015). Takahashi and Yamanaka (2006) discovered in 2006 a method to "reprogram" specialized adult cells to regain certain hESC like features, these induced pluripotent stem cells (IPSCs) are an important new avenue for research which reduces the need for hESC cells in a number of cases. However, for the period under investigation, i.e. 1998 to 2008, hESC was the most prevalent form of Stem Cell research, however IPSC has influenced Stem Cell research trajectories since then (Scott et al., 2011).

Globally, hESC research laws are very disparate with some commentators (Caulfield et al., 2009; Russo, 2005) describing it as a "patchwork of patchworks", mimicking at a global level, what we observe in the US. This patchwork is best illustrated by the Stem Cell World Map, compiled by StemGen (2017) and the similar classification by Russo (2005). On the map we note the varied approaches governing EU countries, with Italy and Germany adopting a restrictive approach and the UK and Belgium being more permissive in the kinds of experiments tolerated and the variety of Cell Lines made available for research. By and large as already noted above at global scale similar to the US we have very different approaches to this emotionally and ethically charged field of research.

The first federal response to the potential future development of hESC

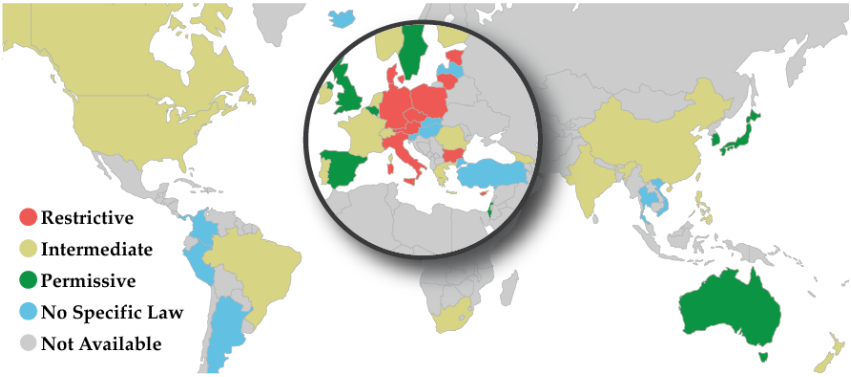


Figure 11: The state of global hESC Research policies as of 2014. The map is based on the information compiled by StemGen (2017), Russo (2005) and Aaron D Levine (2008).

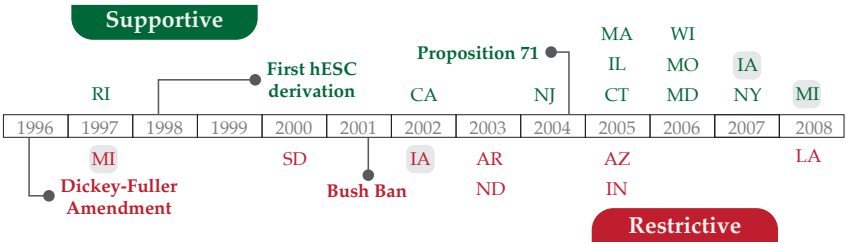


Figure 12: Time-line of main Stem Cell legislation events in the USA. The events above the time-line can be broadly classified as "supportive", the events below "restrictive". The states (abbreviations) listed above and below, indicate the year in which a given state has adopted a supportive or a restrictive stance. Note that IA and MI change from restrictive to supportive.

was the Dickey-Fuller Amendment in 1996. The Amendment bans any federal funding for " (1) the creation of a human embryo or embryos for research purposes; or (2) research in which a human embryo or embryos are destroyed, discarded ..."(Kearl, 2010). This Amendment has been attached to every appropriations bill for the Departments of Health and Human Services, Labor, and Education since 1996. This makes federal funding unavailable, but does not outright ban hESC research as such, as evidenced by Thomson's breakthrough in 1998.

Several states responded to the federal vacuum with legislation of their own. These laws address in part the legality of research on embryos, the consent requirement to conduct research on fetus/embryos or restrictions to purchase and sell human tissue[3]. US states, after the Dickey-Fuller amendment, adopted legislation, which has been classified by Aaron D. Levine, Lacy, and Hearn (2013) as either restrictive or supportive. For example 1998 Rhode Island was the first State to propose important state legislation in favor of hESC and in the same year Michigan adopted legislation hindering hESC research. However some states (e.g. Pennsylvania) had laws on their books even before the advent of the technology in 1998, on the subject of abortion, which indirectly affect the practicality of hESC research (National Conference of State Legislatures, 2016). Which countries are considered supportive or restrictive and can be seen in the time line in Figure 12 and the stance as of 2008 can be seen on the map in Figure 13.

However the lines between supportive and restrictive states are blurry as the nuances of civil and criminal liabilities listed by the National Conference of State Legislatures (2016) shows. Additionally, even if a state has not been formally classified as either supportive or restrictive by the National Conference of State Legislatures (2016) or Aaron D. Levine, Lacy, and Hearn (2013), state level Amendments and legislative proposals have been passed and repeatedly discussed as outlined in Table 22. This ongoing debate in state capitols suggests, that adopted legislation regarding funding and admissible procedures is not a sudden shock or

[3]The National Conference of State Legislatures (2016) has an extensive breakdown of the various legal aspects considered by the various states.

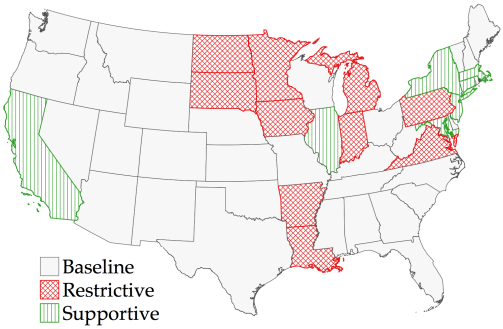


Figure 13: The map of the US highlights the stance of the individual states on hESC. Note: Territories which are not shown (e.g. Alaska) are all Baseline.

would come as a surprise to the resident scientist population, but is rather the culmination of legislative proposals and public opinion, which is also supported by the results of Aaron D. Levine, Lacy, and Hearn (2013). What this means is that while legislation may have been passed in a specific year, the stance of the states would have been known to the relevant population well in advance. This implies two things for the analysis, (1) changing from supportive to restrictive or vice versa is not a shock to the resident population and (2) the level of support, while not enacted by a specific law, characterizes the state further into the past.

Beyond these state level regulations, mostly affecting the practicality of hESC, there have been two important state and federal legislation affecting the crucial aspect of funding. Bush banned in the August of 2001 any federal funding for hESC, exempting research into cell lines derived prior to the announcement. The ban does not only affect future funding, but any research on hESC done in the future may not be conducted on federally funded premises or equipment (Adelson and Weinberg, 2010). As a direct answer to this, California adopted the ballot initiative "Proposition 71" devoting 3 Billion Dollars to hESC and become a precedent that several states have copied since then.

3.2 Data

For the analysis of relocation choices, we need primarily data on the mobility of Life scientists. Fort this reason we make extensive use of the data and methods developed in Chapter 2 to extract yearly mobility networks. Additionally, to control for productivity (i.e. impact weighted publication rate) we use SCIMago journal impact factors. All information on the stance of US state on the issue and the timing of their legislative changes is compiled from Aaron D. Levine, Lacy, and Hearn (2013), Aaron D. Levine (2012), and National Conference of State Legislatures (2016).

3.2.1 Mobility Data

We use the mobility network extracted for the analysis on Brain Drain in Chapter 2 to track if and when researchers chose to relocate.

For this analysis, we extract the mobility network in accordance with the definition given in Section 2.2, but using a buffer size of 2 instead of 5 years (i.e. period before and after the move). We use this shorter interval length for several reasons. By using 2 years instead of 5 years we trade some accuracy in the signal of presence for more certainty on the year of the move. The reduced certainty of presence is due to the fact that using 2 years worth of publications to determine the starting position and 2 years to determine the destination position uses fewer articles then a buffer of 5 years. However, by restricting this interval to this narrower interval we gain more certainty on the year the actual move took place, since the move could have been at most two years off, i.e. if only two publications at the extreme ends of the buffers (before and after intervals) are observed. The certainty on the event timing is important when evaluating the response to the policy shocks we investigate in this work. With the choice of 2 year we lose also a significant number of observations, since we can only use authors who had 2 publications at most 4 years apart. This is not a problem for individuals publishing with a high frequency, but authors with a lower publication rate are less likely to be included. On one hand, we do lose a lot of observations and give up

some certainty on the source location. On the other hand, we gain more certainty on the timing of the event (i.e. leave state, stay) and can extend our analysis to 2008, given that the last complete year of observations is 2009.

3.2.2 Identification of hESC Researchers

We identify Stem Cell researchers through the titles and abstracts of papers they have published in MEDLINE. We distinguish four types of Researchers. These four classes of researchers have an increasing likelihood of being subject to hESC regulation in their work. (1) **Other Life Scientists (NoSC)**, represent all scientists available in MEDLINE for the relevant period but have never mentioned "Stem Cells" in their work and have never worked with anyone mentioning "Stem Cell" in their work. This group should be largely unaffected by Stem Cell legislation. (2) **Stem Cell (SC)** researchers, are scientists which work on stem Cells in general but not on "embryos" or "human embryos". This group might be impacted directly or indirectly by state legislation. The same reasoning goes for (3) **Embryonic Stem Cell (ESC)** researchers which might experiment on "embryos" but not necessarily of the "human" kind. And finally the most likely researchers to be affected are (4) **human Embryonic Stem Cell (hESC)** researchers which explicitly mention human embryos in their research. Specifically, to identify researchers working on SC, ESC and hESC we proceed as follows. We collect all publications in MEDLINE which are marked as `Journal Article` and filter out papers which contain the following strings either in the title or the abstract[4].

SC "*stem cell**"

ESC "*stem cell**" and "*empry**"

hESC "*stem cell**" and "*human empry**"

[4]The full text of the articles is only available for a subset of MEDLINE, and only abstract and Title are readily available.

Here "*" represents a wild-card matching any subsequent letter if it isn't interrupted by whitespace. We restrict our search to Documents of type `Journal Article` to minimize the chance that we pick up commentary on stem cell and not actual hESC research. An alternative to the pure text based search is to use the MeSH (Medical Subject Headings), a key-word classification system maintained by the National Library of Medicine. However "Fetal Stem Cells" has only been introduced as its own concept in 2006 as a refinement of "Stem Cells". Additionally a large part of research articles in MEDLINE available as XML files are not marked up with MeSH qualifiers. To obtain a sufficiently large sample of scientist despite the missing MeSH term and incomplete coverage we have employed a text search strategy.

To take into account, that hESC researchers might not necessarily start their career carrying out hESC research we adopt the following convention. A scientist who has published a paper on hESC in year t is marked as being a hESC researcher 3 year before $(t - 3)$. A conceptual illustration of this definition is available in Figure 14. We adopt this convention

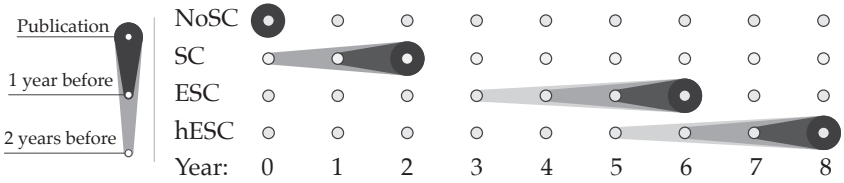


Figure 14: Identification of SC researchers. The illustration shows an author becoming active in year 0 by publishing a NoSC paper. In the subsequent years she publishes on SC, ESC and hESC. Since a publication is a delayed signal of activity, her type is determined several years prior. Using an offset of 3 years, she is an SC scientist in 0, ESC in 3 and hESC in 5.

for three reasons. First, the publication represents a delayed signal of the actual research, i.e. if a paper appears in 2005 it stands to reason that work on it has been done before that year. Second, we do only observe papers available in MEDLINE, which does not offer universal coverage. This means that we could have missed a prior publication and would classify the researcher as hESC later then we should have. And third, it

takes time for a young researcher to start publishing on hESC and thus a publication is a delayed signal of the fact that his career in the previous years was in this field.

3.2.3 State Stance

We distinguish between *Supportive* and *Not Supportive* (i.e. did not outright declare support yet or are restrictive). First we assume that explicitly signaling support is more important for attracting researchers then omitting or outright declare such research not welcome. In the Supplementary materials we carry out parts of the analysis also distinguishing between baseline and restrictive.

To appreciate the evolution over the years of the distribution of these scientists across the various policy regimes (including "restrictive") we show in Figure 15 the breakdown of the four types. The 4 types of researchers differ primarily in their number, with each group being substantially smaller than the previous. The low number of stem cell researchers becomes an issue if we consider that as we have seen in Chapter 2 that most scientists do not move in any given period.

3.3 Geographic Preferences of hESC Researchers

Where scientists choose to relocate is a multi-faceted issue starting well before observed individual mobility. A first factor, which might reasonably, influence the mobility of hESC researchers is *where* they are currently residing and working. If this distribution of hESC scientists is biased toward supportive states, it would imply that the scientist of interest who could potentially move from restrictive to supportive is skewed. Which leads us to formulate our first hypothesis.

$H_{\textbf{prev.}}$ hESC Researchers are more prevalent in supportive states.

If hESC scientists residing in supportive states, are also less likely to leave their state then their counterparts in non-supportive states then this would bias the propensity to move to supportive states. This is the basis for our second hypothesis.

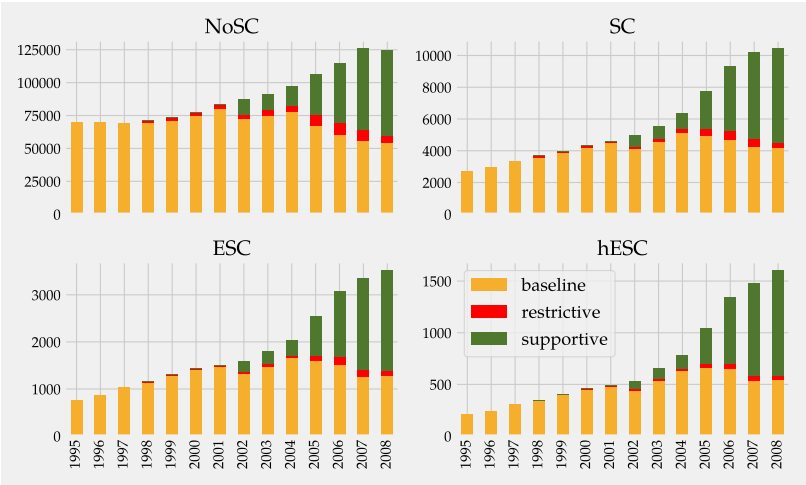


Figure 15: Distribution over the years of researchers as according to the stance of the state they work in. The stance of the state changes over the years.

$H_{\textbf{stay}}$ hESC Researchers are less likely to leave a supportive state.

Finally we look into the relocation choice itself keeping in mind that $H_{\text{prev.}}$ and $H_{\text{prev.}}$ can confound this estimate, and formulate the following hypothesis.

$H_{\textbf{supp.}}$ hESC Researchers *if* they move, are more likely to move to a supportive state.

In the regression analysis we will be using the following variables and controls.

Supportive Equal to 1 if the state is the given year is supportive and 0 otherwise (i.e. not supportive)

StemType Dummy variables identifying the 4 classes of Stem Cell researchers, the NoSC is the always the base case

Productivity The log of annual impact weighted publication rate, same definition as described in detail in Section 2.4.1

Age Group Age is measured as the difference in years from first publication to the move year. The age-groups are split such that the cohorts are of comparable size.

State The state the author was in before the year the move took place

Year The year in which the move took place

Inter-move The time between the last observation before the move, this variables quantifies how large the publication hole (i.e. no pubs available) is.

3.3.1 High Prevalence of hESC in Supportive states.

In testing hypothesis $H_{\text{prev.}}$ we look at the distribution of scientists across supportive and non-supportive states. Not all states are of comparable size in 2008 as suggested by Figure 16. Supportive states, as declared in 2008, make up 50% of the total population of Life Scientists, however this has not been the case in previous years, where few states enacted outright legislation in support as illustrated by Figure 15.

As first test for $H_{\text{prev.}}$ we compare the distributions of the populations across supportive and non-supportive states aggregating the position the scientists were in from 2000 to 2008. We restrict the interval to this sub period since prior to 2002 the number of scientists working in declared supportive states is negligible.

By comparing the relative prevalence in the geographic distribution of hESC researchers across supportive and not supportive states against the NoSC population, we can identify an imbalance. This analysis helps us to validate two things, (1) the validity of the classification itself and (2) geographic imbalance. If the classification of supportive states were not informative there should not be a discernible difference between the two groups. Specifically, a deviation from the distribution of the NoSC population in favor of supportive states would falsify the assumption that the two population are distributed proportionally. In fact, we do find evidence in support of $H_{\text{prev.}}$. As shown in Table 9, hESC scientists are more prevalent in supportive states. All three levels of exposure (i.e. SC,

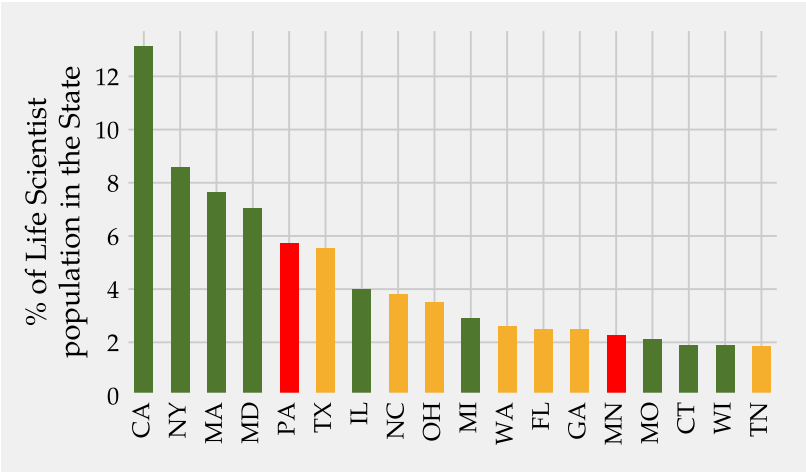


Figure 16: Percentage of the Life Scientists Population working in 2008 in the given state. The shown states cumulatively account for 80% of this population.

ESC and hESC) are between 15% and 33% more likely to be encountered in a supportive state then their NoSC counterparts.

To make sure that what we observe here is in fact also true accounting for state and year effects we estimate a panel fixed effects model covering the period 1995 to 2008. We include a longer period then before since by using a Fixed Effect model we can exploit the change from not supportive to supportive for a state.

$$\log(\text{ESC Pop.})_{st} = \beta_0 + \beta_1 \log(\text{Total Pop.})_{st} + \beta_2 \text{Supportive}_{st} + \alpha_s + u_{st}$$

We estimate the log of the hESC population as a function of the stance of the state in that year and the size of the entire Life Scientists population in that state. Note that the base-case for the state policy stance in the regression is "not supportive". The results of the panel regression in Table 10 suggest in accordance with Table 9 that there is a preference for hESC scientists to work in supportive states. A large part of the variance is explained by the scientist population of the state. This is validated

	Supportive		Not Supportive	
	Prop (%)	Δ %	Prop (%)	Δ %
NoSC	32.4		67.6	
SC	37.2	+15.0	62.8	-7.2
ESC	39.9	+23.2	60.1	-11.1
hESC	43.1	+33.1	56.9	-15.8

Table 9: Distribution for the four classes of researchers in the period 2000 to 2008 by the state's supportive stance as they are distributed. *Prop* indicates what the proportion of researchers are found in that class. For example authors without any connection to Stem Cells are have a concentration of 32.4% of in supportive states. The Δ% indicates by how much the Stem scientist populations deviate from the NoSC population, i.e. hESC / NoSC -1

in part by what we see in Figure 16, that the largest states are indeed supportive. In addition to the Fixed Effect specification, we test for robustness also random effect models and as well as regressing on the raw population count (see Table 23, 24 and 25 in the Appendix). The results are all in line with the Fixed Effect specification.

The high relative prevalence of hESC scientists in supportive states and the Panel Regression suggest quite strongly that scientists with high exposure to stem cell legislation tend to work in supportive states and offers strong support for $H_{\text{prev.}}$.

3.3.2 Propensity to Move

In the following section we address H_{stay}. If H_{stay} is true then we should find that hESC scientist are less likely to leave their state if it is supportive. For this purpose we estimate the probability that a scientists working in a given state leaves conditional on the stance of the state in that year. Specifically, we estimate the following repeated cross section model.

Table 10: Fixed Effect regression on the log of the population stocks

	(1) log(Pop. SC)	(2) log(Pop. ESC)	(3) log(Pop. hESC)
log(Pop Tot.)	1.157***	0.913***	0.753***
	(32.10)	(22.03)	(15.91)
Supportive	0.288***	0.357***	0.585***
	(5.04)	(5.00)	(6.14)
Constant	-4.048***	-3.467***	-3.238***
	(-16.13)	(-12.15)	(-10.06)
Observations	593	593	593
R^2_{within}	0.656	0.449	0.333
$R^2_{between}$	0.865	0.786	0.686
$R^2_{overall}$	0.861	0.781	0.658

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$$\begin{aligned} \text{Leave}_i = \beta_0 &+ \beta_1 \text{Supportive}_i + \beta_2 \text{StemType}_i \\ &+ \beta_3 \text{Productivity}_i^0 \\ &+ \beta_s \text{Supportive}_i \times \text{StemType}_i \\ &+ \text{Age Group}_i + \text{Year} + \text{State} + u_i \end{aligned}$$

We look at two types of move here, first "Leave City" which indicates whether the scientists moved to a different city in that year and "Leave State" which is 1 if the scientist left the state and 0 if he did not. According to H_{stay} we should expect that hESC scientists who are already in a supportive state are less likely to leave it.

The results are listed in Table 11. The most important result in this table with regards to H_{stay} are the two rows "hESC" and "supportive $\times$ hESC", with the first indicating the likelihood that a hESC researcher in general leaves his state and the second the probability of hESC researcher in a supportive state to leave his state. The results suggest that the probability to leave the state are higher for hESC researchers. However this

tendency is reversed for hESC scientists working in a supportive state, as "supportive $\times$ hESC" suggests. This is the case even after controlling for State effects. Similarly we find that hESC scientists are less likely to leave the city if it is located in a supportive state.

These results suggest that H_{stay} is true and hESC scientists are more reluctant to leave supportive states than their counterparts in not-supportive states.

3.3.3 Mobility Preference

So far we have noted that there are two forces affecting the relocation choice of hESC researchers which could confound our estimates to move from not-supportive to supportive. In fact, we have found that in general hESC researchers are already predominantly residing and working in supportive states and have a lower propensity to move then their counterparts in not supportive states.

Nevertheless, we can estimate for the researchers which do indeed change city in each year what the type of the target is most likely to be. More precisely we estimate the following Logit model.

$$\begin{aligned}\text{TargetType}_i =& \text{SourceType}_i + \text{StemType}_i \\ &+ \text{SourceType}_i \times \text{StemType}_i \\ &+ \text{Age Group}_i + \text{Year} + \text{State} + u_i\end{aligned}$$

The results are listed in Table 12.

We do find that researchers in general are more likely to move to supportive states and avoid Not Supportive states, and that ESC researchers are less likely to move to "Not supportive states" then the general Life Scientist Population, however we do not find a discernible difference for hESC researcher to move preferentially to supportive states above and beyond what we would expect from the general mobility patterns. This means, that we find mixed signals with regards to $H_{\text{supp.}}$. We do know that $H_{\text{prev.}}$ and H_{stay} are likely the case and they might negatively affect this estimate. The strongest claim we can make with regards to $H_{\text{supp.}}$ is

	(1) Leave State	(2) Leave State	(3) Leave City	(4) Leave City
main				
supportive	-0.0179	-0.0193	0.234***	-0.0243
SC	-0.0193	-0.0123	-0.0625***	-0.0613***
ESC	0.135***	0.136***	0.0553**	0.0485*
hESC	0.0971***	0.104***	0.0623*	0.0469
supportive × SC	0.0636	0.0674*	0.0651	0.0731*
supportive × ESC	0.0255	0.0342	0.0139	0.0333
supportive × hESC	-0.241***	-0.231***	-0.170**	-0.162**
Productivity	-0.0564***	-0.0549***	-0.0806***	-0.0867***
intermove	0.348***	0.347***	0.354***	0.355***
Constant	-2.033***	-2.287***	-1.750***	-2.100***
Age Effects	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes
State Effects	No	Yes	No	Yes
Observations	1132030	1132030	1132030	1132030
Loglikelyhood	-561804.7	-560587.5	-625739.5	-621958.5
Pseudo R^2	0.0245	0.0266	0.0252	0.0311

The errors are clustered at the Source State Year level (582 clusters)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Logit Regressions on the Probability to leave the State (1 and 2) and leaving the city (3 and 4).

	(1) Supportive	(2) Not Supportive	(3) Abroad
SC	0.138***	-0.110***	0.0642**
ESC	0.0700	-0.188***	0.187***
hESC	0.0988	-0.0230	-0.0177
supportive	1.047***	-0.819***	-0.0163
supportive × SC	-0.211***	0.0642	0.0406
supportive × ESC	-0.190	0.0647	0.0583
supportive × hESC	0.0339	-0.111	-0.0429
Productivity	0.0330***	-0.00950	-0.0114
Constant	-5.980***	1.769***	-1.606***
Age Effects	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes
State Effects	Yes	Yes	Yes
Observations	287913	287918	287918
Log Likelyhood	-109877.3	-176447.3	-158996.0
Pseudo R^2	0.232	0.107	0.00763

The errors are clustered at the Source State Year level (572 clusters)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Logit Regression with dependent variable the Destination Type

that there might be a tendency for hESC researchers to move to supportive states, but the effect is masked by the fact that most hESC researchers are already in supportive states and that these do not move as much, once there.

3.4 Impact of important legislative changes

We have noted in the previous section that there was no significant mobility induced overall by a state's stance on the probability to move to a supportive states. Which is most likely the result of more scientists already residing in supportive states and those scientists having a lower propensity to move. We have assumed that the propensity to move is constant across time and is only influenced by the stance of a state in a given year controlling for year and state fixed effects.

However as we have noted in the historical overview, state and federal legislation has not stood still and there have been two major events in the period of interest which could have reasonably affected the attractiveness for hESC research of the US in general and California in particular. These two events are, (1) the heavy restrictions on federal funding for hESC research imposed by Bush in August 2001, henceforth the "Bush Ban" and (2) Proposition 71 in California. The Bush Ban prevents federal funds from being used for hESC research on non-approved stem cell lines and the derivation of any new cell lines (see introduction for more details). Proposition 71, a Californian funding initiative passed in November of 2004, "authorizes a total of nearly $3 billion in tax-free, general obligation state bonds to support stem cell research at California hospitals, medical schools, universities and other research institutions over 10 years"(Baker, Deal, and Principal, 2004, p. 2). Importantly, this funding may be used to derive new hESC lines (Nature Cell Biology, 2010), in direct opposition of the Bush Ban.

With regards to these two legislative changes we formulate the following three hypotheses and test them in turn.

$H_{\textbf{abr.}}$ hESC Researchers moved with a higher probability abroad following the Bush Ban.

$H_{\textbf{P71}}$ hESC Researchers move in response to the approval of Proposition 71 with a higher probability to California.

$H_{\textbf{stayCA}}$ hESC researchers residing in California are less likely to leave the state following Proposition 71.

To assess if there has been any change in mobility patterns following these interventions, we will estimate the probability to move, not only overall as done in the previous sections, but year by year. If there is indeed a strong signal that mobility abroad or to/from California for hESC researchers changes after the two interventions they would show up as significant increases in mobility intensities above and beyond the general trend observed in the less exposed groups, i.e. NoSC, SC and ESC.

3.4.1 Move abroad

First we investigate $H_{\text{abr.}}$, the claim that hESC researchers left the US following the announcement of the Bush Ban in the August of 2001. We have noted in the previous section that there is no clear preferences for hESC scientists to move abroad (there is for ESC), but there could have been a peak over the years which we have averaged out. To test the hypothesis that there was an up-tick in moves abroad, we estimate the following model. We consider only researchers which did indeed move to another city, in fact only considering anyone who has actively expressed a choice.

$$\begin{aligned}\text{Move Abroad}_{it} = \beta_0 &+ \beta_1 \text{Supportive}_{it} \\ &+ \beta_2 \text{Year}_{it} \times \text{StemType}_{it} \\ &+ \beta_3 \text{Productivity} + \beta_4 \text{Intermove} \\ &+ \text{Age Group}_{it} + \text{Year} \\ &+ \text{State} + u_{it}\end{aligned}$$

The result of this regression are available in the Appendix in Table 26. We note, that controlling for state effects, we do not find a significant

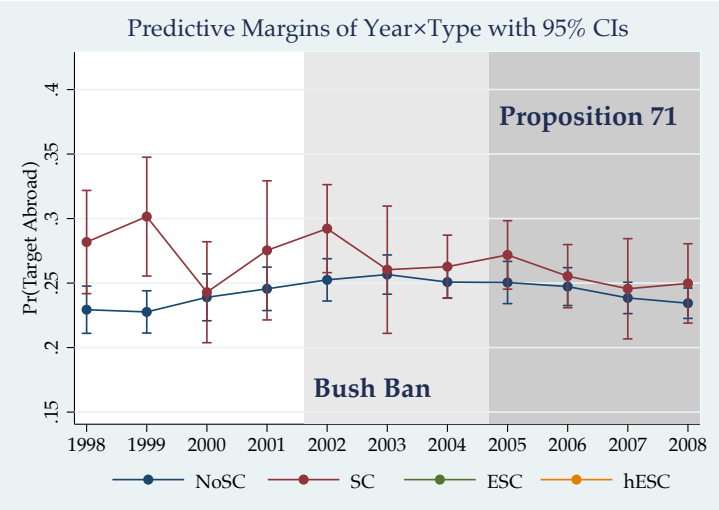


Figure 17: Marginal Probability to move abroad, for researchers which move in that year (i.e. $\Pr(\text{Move Abroad}|\text{Leaving city}, X)$). The errors are clustered at State×Year level. The difference in the error bar sizes is due to the different cohort sizes.

difference between supportive and restrictive states in their tendency to move abroad. However we do observe some differences in yearly marginal probabilities to move abroad for the 4 researcher types. In Fig-

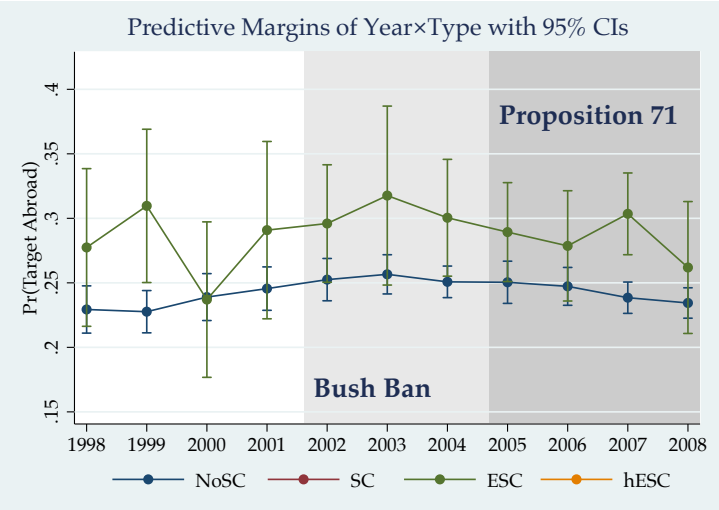


Figure 18: Marginal Probability to move abroad, for researchers which move in that year (i.e. $\Pr(\text{Move Abroad}|\text{Leaving city}, X)$).

ure 17 we show the probability to move abroad for unaffected group NoSC and the largely unaffected SC group. Note that with the exception of 1999 — well before both the Bush Ban and Proposition 71 — the two groups have the same propensity to move abroad, suggesting that SC has not been materially affected by either the Bush Ban or Proposition 71. This points to the fact the SC group, while researching Stem Cells in general the missing focus on human embryos means that they are largely unaffected. Similarly the more at-risk group ESC, shown in Figure 18, is more likely to leave the country then NoSC, however this has been true for most of the years and is in line with the results in Table 12. Importantly the trend does not seem to be materially affected by the two legislative measures. These results points again to the fact stem cell research as a whole is not materially affected. We do however find that hESC, researchers specifically working with human embryos seem

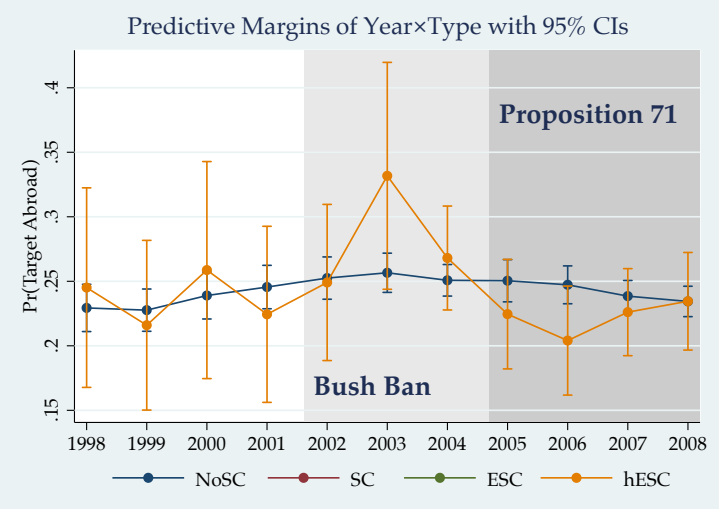


Figure 19: Marginal Probability to move abroad, for researchers which move in that year.

to respond to the two treatments, as shown in Figure 19. The trend from 1998 to 2002 is statistically not distinguishable from NoSC, suggesting that mobility abroad before the Bush Ban was in line with general Life Scientists. However in 2003 there is a sudden up-tick in mobility abroad, which is quickly reversed and tends below NoSC after 2005. The difference is not significant at the 95% confidence level, as the overlapping error bars suggest, the sudden shift in trend could however indicate the influence of the Bush Ban or Proposition 71. A similar conclusion is supported by comparing ESC and hESC, which is a comparable group in numbers and research focus.

With regards to $H_{\text{abr.}}$, we do not find conclusive evidence that hESC mobility has materially increased following the funding ban. We do however find that unlike the other at-risk groups (SC and ESC) the mobility abroad seems to be affected by the Ban. The sudden drop off after 2005 could be simply a reversal to the mean or it could have been influenced by Proposition 71 In the next section we will explore this possibility.

3.4.2 California and Proposition 71

The funding made available by Proposition 71 in November of 2004 is to date the largest state hESC funding scheme in the US. After California several other states have adopted similar funding schemes (e.g. New Jersey, Massachusetts), although nowhere near the size of the 3 Billion provided by Proposition 71 (Acosta and Golub, 2016). Both the generous funding and the signal to other states to start their own funding schemes are arguably a very important legislative shock. We argue therefore that this event has affected increased mobility to California in particular and prevented moves abroad in general.

Immediately after the approval of Proposition 71 litigation regarding its legality prevented the disbursement of funds for research until 2006. As Acosta and Golub (2016) recounts, this time was however not lost as the administrative apparatus to evaluate and monitor future investments was being built. This delay has then arguably two effects on the hESC population. On the one hand, it made the financing less certain, therefore some researchers may have decided to move anyway. On the other hand, the ongoing work on the administrative framework was a signal to the hESC scientists that California was serious about this initiative.

In line with the analysis in Section 3.4.1, where we estimated the propensity to leave the US we estimate the propensity to move from a US state to California.

We estimate the probability to move to California, conditional on observing a move to another city. In other words the variable "Moved To California" is 1 if a given scientist moves to California in the given year, and 0 if the given scientist moves to a different country or city which is not in California from his current city. We also do not include scientists which are already based in California and move to different city. Specifically, we estimate the following model.

$$
\begin{aligned}
\text{Move To California}_{it} = \beta_0 &+ \beta_1 \text{Supportive}_{it} \\
&+ \beta_2 \text{Year}_{it} \times \text{StemType}_{it} \\
&+ \beta_3 \text{Productivity} + \beta_4 \text{Intermove} \\
&+ \beta_5 \text{Age Group}_{it} + \text{Year} \\
&+ \text{StateFixedEffects} + u_{it}
\end{aligned}
$$

In Table 27 the regression results are outlined. We compare the marginal probability to move to California over time for the various researcher types to understand if there has in fact been a sudden influx of scientists following either the Bush Ban or Proposition 71.

As in the previous section, we look at the marginal predicted probabilities to move to California for the at-risk groups in sequence. We can verify again in Figure 20 that the SC group is virtually identical to the NoSC in its propensity to move to California and neither the Bush Ban nor Proposition 71 seem to have had an effect.

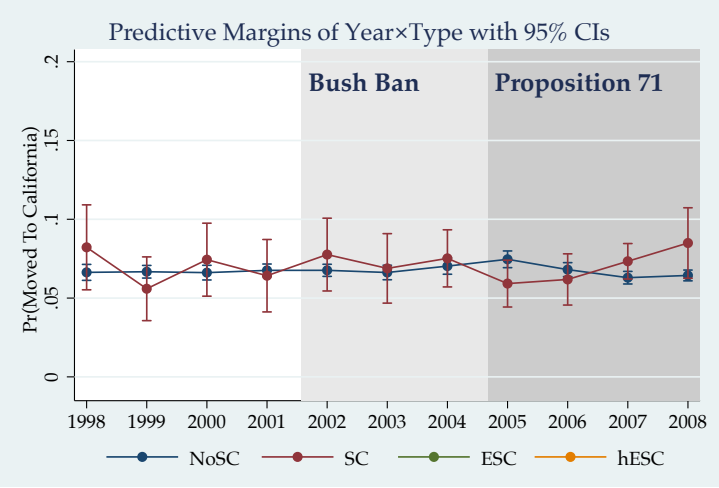


Figure 20: Marginal probability to move to California for both NoSC and SC.

The ESC group, the second most at-risk group, similarly does not deviate in its propensity to move to California from the NoSC group. However there seems to be a drop in mobility toward California in 2006, which is quickly reversed. What is clear is that the announcement of the Bush Ban did not materially affect ESC's propensity to move to California and Proposition 71 did not increase this propensity.

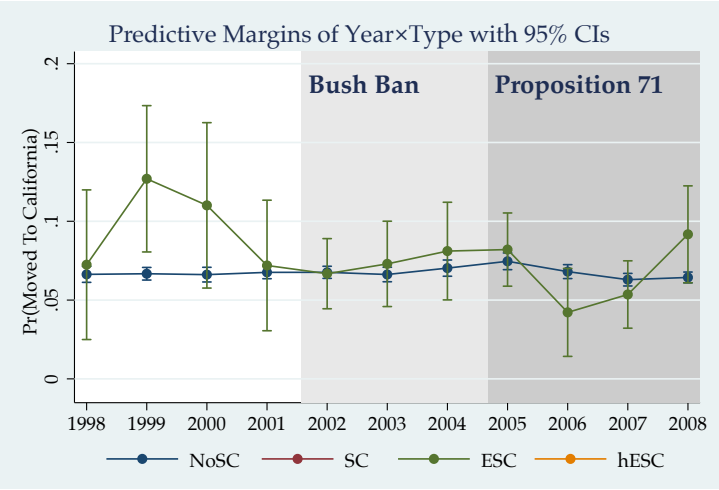


Figure 21: Marginal probability to move to California for both NoSC and ESC.

The impact of the two policies changes when considering the group most likely to be affected by both legislations, hESC. The marginal probability to move to California for hESC and NoSC are shown in Figure 22. Overall we can observe, disregarding confidence bounds for a moment, that the propensity of hESC researchers to move to California is in line with what we expect it to be in the three policy regimes, i.e. in the *pre* Bush Ban (1998 to 2001) to be in line with general mobility trends, in the years after but before the vote on Proposition 71 (2002 to 2004) to drop due to mobility abroad, and a reversal after the adoption of Proposition 71 (2005 to 2008). The effect size is not negligible, jumping from 5% just before Proposition 71 in 2003, to 12% in 2006. In the years that follow

this trend reverses to the NoSC mean of 7%. By considering the 95% confidence bounds, the jump from 2002 to 2006 is significant, assuming common trends with either NoSC or SC.

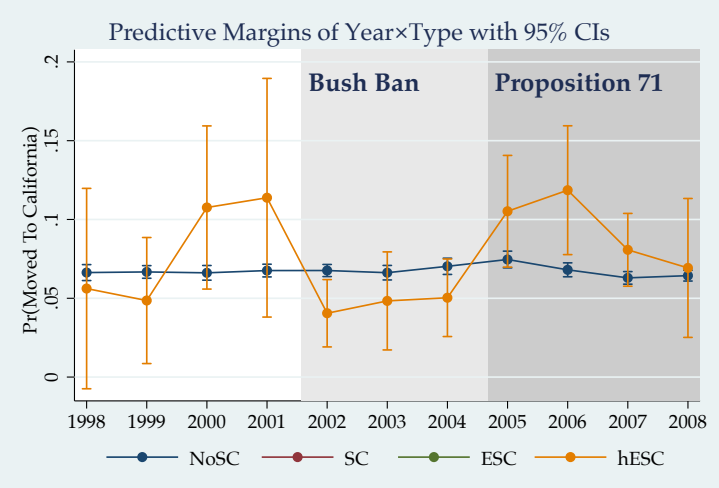


Figure 22: Marginal probability to move to California for both NoSC and hESC.

In addition to the divergence of the propensities to move to California we have also shown with $H_{\text{prev.}}$, that hESC scientists are concentrated in supportive states and with H_{stay}, that scientists in supportive states are less likely to move. These two factors arguably reduce the observed effect size, given that most scientists are already where they want to be, either because they moved there and don't want to relocate or because they started their career there.

By comparing Figure 19, the probability of hESC researchers to move abroad with Figure 22, the probability of hESC researchers to move to California, we note that the response to the two policies is mirrored. Where we observe an increase in mobility abroad, from 2002 to 2004 we observe a drop in mobility to California, and analogously after the vote on Proposition 71, we observe a drop in mobility abroad and an up-tick in mobility towards California. This "mirror effect" suggest the possible

existence of a substitution effect, whereby mobility to California is substituted for mobility abroad post Ban announcement and post Proposition 71 vote a shift from abroad to California.

To further strengthen our claim that Proposition 71 has had the overall effect to reduce mobility abroad and increase retention of hESC scientists in the US in general and in California in particular we look at the marginal predicted probability to leave California over the same period. These probabilities for the least affected groups are shown in Figure 23, where we note again the SC has not been affected. In fact, the propensity to leave California if anything has increased above and beyond the baseline NoSC group.

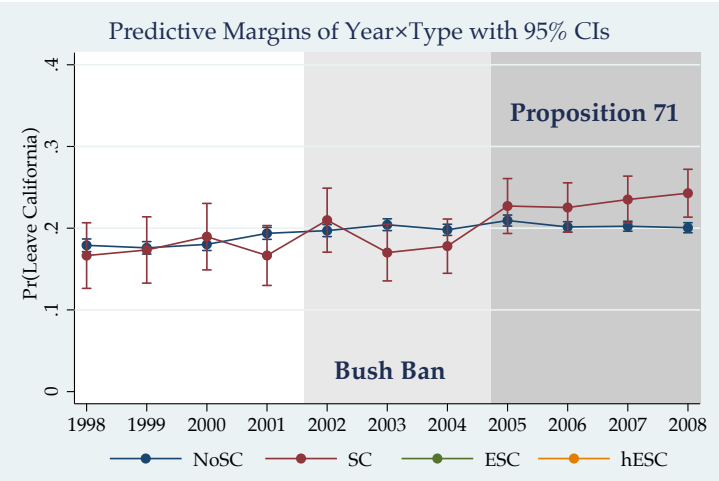


Figure 23: Marginal Predicted probability to leave California comparing NoSC with SC

However, as expected and in line with the results on the propensity to move abroad and to California, we find that the propensity for hESC researchers to leave California mirrors the propensity to leave the country as shown in Figure 19. This results suggests that the adoption of Proposition 71, has not only increased mobility towards California, but has dissuaded scientists from leaving the state.

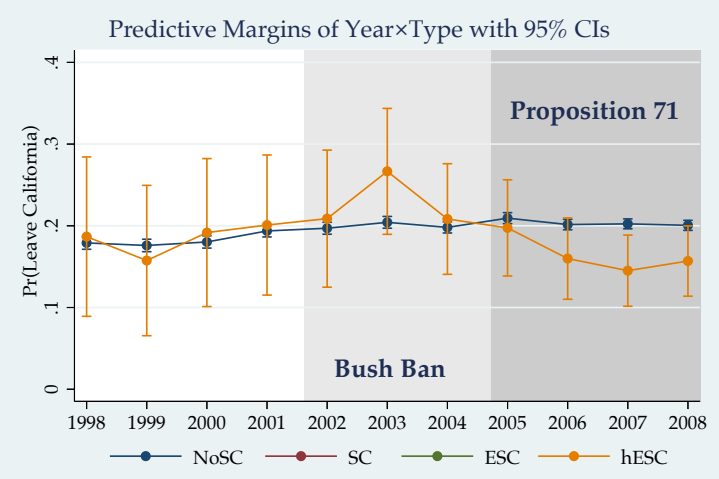


Figure 24: Marginal Predicted probability to leave California for hESC.

3.5 Discussion and Conclusion

From our analysis the following stylized facts regarding the mobility patterns of hESC researchers in the period 1998 to 2008 in the US emerges:

1. hESC scientists are more likely to work in a supportive state (see Section 3.3.1).

2. hESC scientists working in a supportive state are less likely to move then their counterparts in non-supportive states (see Section 3.3.2).

3. However, given the first two results, we do not find a significant preference for hESC researchers to move to supportive states (see Section 3.3.3).

4. Mobility abroad did not increase for the at-risk groups, significantly (see Section 3.4.1). The hESC group did experience a single up-tick in abroad moves, which was quickly reversed.

5. Following the vote on Proposition 71, the drop in moves towards California is reversed (see Section 3.4.2).

6. Following Proposition 71, hESC scientists leave California with a lower probability (see Section 3.4.2).

The overall theme that emerges from the analysis is that the varied approaches to stem cell research have had an impact on the geographic distribution of hESC researchers across the US. Supportive states do in fact employ more hESC scientist than non-supportive states, and more importantly, once working in a supportive state these scientists do not leave.

Specifically, with regard to the two most important pieces of legislation, i.e. restrictions on funding put in place by Bush, and the 3 Billion Dollar funding scheme in California, we find that the supportive state level legislation has counteracted the federal restrictive stance. There is some indication that hESC scientists were headed abroad following the Ban in 2001. This trend was however reverse nearly immediately, and in fact we do find that hESC researchers are less likely to leave the country or California, then their less exposed counterparts after 2004.

We find that state legislation and funding has had the overall effect an anticipated exodus of hESC researchers following the Bush Ban. We argue that this is because viable alternatives to move to within the US were created. A prime example of which is California with Proposition 71. According to our analysis, this funding scheme has attracted scientists to California and reduced the propensity of Californian scientists to leave the state. In fact, the nearly simultaneous drop in mobility abroad and increase in moves towards California suggests that this initiative, and several the state funding schemes that followed, have had the desired effect to avert an exodus of researchers.

More generally the supportive state response to the restrictive stance of the federal government, suggests that in this case, the US national research system is resilient to restrictions. HESC research as a whole has been affected by restrictive state and federal legislation as the geographic redistribution of hESC researchers suggest. However the ability of individual states to step in and experiment with new laws has by and large negated the effect of federal restrictions and allowed hESC research in the US to flourish. In part the observed dynamics of state legislation

counteracting federal rules is an example of partisan politics playing out at the two levels. It is probably not a coincidence that all state funding initiatives, enacted after Proposition 71, have been passed by Democratic majorities and with the lone exception of New Jersey, had a Republican governor (Acosta and Golub, 2016).

Overall this study points to the fact that hESC research, being a controversial topic in the US has lead to a spatial segregation, whereby hESC researchers are concentrated in supportive states. More importantly however, restrictive federal and state laws have been negated by permissive legislation which has most likely thwarted a possible exodus of hESC researchers. The two tier approach to hESC legislation is therefore an example of how the US national research system reacted to restrictions. According to some commentators (Salter and Harvey, 2008; Gottweis and Prainsack, 2006; Klitzman and Sauer, 2009; Aaron D Levine, 2008) the bearish stance on hESC by the federal government could have compromised the standing of US hESC research for years to come. However it looks like the the national research system was resilient to this threat as indicated by our analysis and the continued leadership of the US in Stem Cell research. Acosta and Golub (2016, p. 21) sum up the Stem Cell legislative battle in this period very well with:

> "There is an aphorism in laboratory science that not all experiments are successful but all experiments teach. In this case, the state variations on stem cell policy may be able to teach us much about the most desirable components of a future national policy."

Chapter 4

The Impact of Acquisitions on Inventor Turnover in the Pharmaceutical Sector

This chapter is based on the collaboration with my co-authors Jeroen Allard van Lidth de Jeude, Federica Parisi and Prof. Massimo Riccaboni

A firm's post Mergers and Acquisition's (M&A) performance depends on its ability to exploit the assets and capabilities of the target company. Crucially for Research & Development (R&D) intensive industries, this does not only entail tangible but also intangible assets. It follows therefor that intellectual property both in codified form (e.g. Patents) and tacit knowledge (i.e. key employee know-how) must be managed with care so as not to jeopardize the success of the maneuver. In fact, Gottweis and Prainsack (2006) find that employee turnover is a major reason for M&A failure, with reportedly up to three quarters of these deals failing (King et al., 2004). Specifically we look at the Pharmaceutical industry which saw a rush of M&As in the Nineties[1]. This sector is an exceptionally

[1]Mergers and Acquisition's (M&A) deals in 1999 involving a US companies were worth well in excess of 500 Billion USD (Danzon, Epstein, and Nicholson, 2007), highlighting the economic importance of the practice.

R&D intensive sector, with Pharmaceutical Researchers and Manufacturers of America (2012) claiming that as much as 17% of sales are devoted to R&D, compared to the US industry average of 4% (Danzon, Epstein, and Nicholson, 2007). The importance of Patents and Intellectual Property (IP) is therefore a clear priority. Inventors, the employees behind these innovations, represent by extension future production of patents and as such their retention is also important. This is especially relevant if we assume that these employees have relevant tacit knowledge and contribute to difficult to replicate "socially complex knowledge" (Barney, 1991), i.e. firm capabilities which emerge from rich social interactions. As Coff (1997) puts it, you don't want the source of your competitive advantage to simply "walk out the door".

An inventor working for an acquired company can respond in several ways to the acquisition. The most desirable outcome in our framework is that he (1) stays with the company and continues to produce patents, however this is not the only alternative he has. The alternatives we focus on mostly in this work is (2) that he continues his patenting activity, but he does so for a third party unrelated to the M&A or (3) retires. However there are various intermediate degrees with which an affected inventor could respond, he could (4) reduce his patenting activity and transition to a different role where he no longer applies for patents, thus for our patent based approach he would disappear, but still stay with the company.

The extant literature addresses turnover and related performance issues as they related to top management (e.g. Hambrick and Cannella (1993), Haveman (1995), and Haleblian et al. (2009)), turnover in general (Carriqury, 2017) and to a lesser degree specifically the impact of M&As on R&D personnel (Ernst and Vitt, 2000). We want contribute to this discussion by providing further evidence that M&As have a detrimental effect on retention of inventors.

Specifically, in this work we want to answer the question, whether an acquisition has a negative effect on inventor retention. In this work we look specifically at acquisitions in the Life Sciences (mainly biotechnology and pharmaceuticals) for two reasons. First, it represents an R&D intensive sector and hence relevant for our analysis on retention of employ-

ees and intellectual property and second, in the Nineties, the pharmaceutical sector has experienced an increased rate of M&A events, ranging from mergers between large pharmaceuticals companies to acquisitions of Biotechnology start-ups. Moreover, excluding the most recent deals gives us the possibility to look at the M&A effects on the long run. Finally, this choice allows us to avoid relying on the latter part of the patent applications, whose coverage might incomplete due delays in data availability.

Note that we do *only* consider outright acquisitions and not any weaker form of M&A such as partial or complete mergers. We do this to sidestep several difficulties we might encounter when considering also Mergers. First, by only considering Acquisitions there is a clear indication of which entity will have control and that no new legal entity might emerge from the transaction[2]. Second, the relationship of acquired and acquiring allows us to identify the party in the deal, which is arguably most likely to be affected by changes, i.e. the acquired firm.

The remainder of this work is organized as follows. In Section 4.1 we discuss the extant literature on employee turnover and M&As. In Section 4.2 we introduce the various datasets. The empirical strategy is presented in Section 4.3 as well as the various steps involved in our matching procedure. In Section 4.4 the regression results are shown. Finally, we discuss the results and their importance, how they fit in with the extant literature and possible improvements in Section 4.5.

4.1 Acquisitions and turnover

There is a rich literature on the motivations of M&A and their impact on both stock returns, employee turnover and the post-merger integration. Here we offer an overview of the literature on employee turnover and the various caveats regarding both the motivation for M&As as well as the type of companies, which are most likely to be involved.

[2]In the actual analysis we make sure that after the acquisition the name of the acquired company does not change and if it does we try to identify it through variations of the acquired and acquiring company's names.

Common reasons proposed for why M&As take place combine elements of vertical and horizontal integration, economies of scale and scope, and transfer of specific assets or capabilities (Danzon, Epstein, and Nicholson, 2007; Higgins and Rodriguez, 2006; Ravenscraft and Long, 2000). Higgins and Rodriguez (2006), analyzing pharmaceutical firms in the Nineties, finds that companies with expiring patents and declining internal productivity are more likely to engage in M&As. The rational behind this proposed mechanism is that expiring patents free up production and R&D capacity quickly and fixed and sunk cost assets (i.e. employees and factories) can be used to enhance the impact of an acquired company which has a viable compound but lacks productive and marketing capabilities.

M&A can lead to positive stock performance as Higgins and Rodriguez (2006) finds. M&As that were identified as being conducted with the expressed purpose of R&D consolidation, lead to abnormal returns to both the acquired and the acquiring company. However, with the caveat that the acquired firms were more likely to experienced financial difficulties (Danzon, Epstein, and Nicholson, 2007). With respect to the financial standing of companies engaging in M&As In particular Ravenscraft and Long (2000) finds that acquired firms tend to experience negative stock returns up to 18 months prior to takeover.

In this work we take a "knowledge base view of the firm"[3] and argue that among the various motives, in addition to the acquisition of codified intellectual property (i.e. Patents) the retention of tacit knowledge held by employees is an important factor for R&D intensive industries. Consequently, retention of employees involved in R&D efforts as witnessed by their Patent applications, is a desirable outcome.

Hambrick and Cannella (1993) finds evidence to suggest that the departure of important employees was an important predictor of poor post-merger performance. The authoritative work by Coff (1997) on the management of human assets and their tacit knowledge, reiterates this senti-

[3]In accordance with the definition of "knowledge base view of the firm" given by Grant (1996) a firm is considered to considered as vehicle to organize tacit and complex social knowledge held by individuals to create products and services.

ment and argues that retention of key employees is difficult to get right in acquisitions.

With respect to turnover several studies focusing on M&As in the Nineties have found that top executives and management are more likely to leave the firm following an acquisition. Haveman (1995) evaluating the retention rate in M&As involving financial companies finds that top executives are likely to leave the firm following the event. Ranft and Lord (2000) finds that according to a survey involving 89 firms which were part of M&As top executives were the most likely to leave and 22.7% R&D personnel left on average. However, he also finds in accordance with our hypothesis that R&D personnel is considered to be the most important class of employees to be retained. Specifically, Ernst and Vitt (2000), focusing on acquisitions in general in the US and Germany, finds that indeed inventors tend to leave their company.

As noted by Carriqury (2017) the disruptive nature to the firms by a major managerial intervention is cause to believe that it might have an overall negative effect on the turnover rate. Changing routines, managerial hierarchies and the added stress of uncertainty could all factor into the decision of employees to leave a company. Hobman et al. (2004) for example argues that this increased pressure can cause turnover. Similarly Holtom et al. (2005) argues that shocks such as M&A "trigger" a reevaluation of career and life goals of the affected employee. For example an employee wanting to leave all along takes the M&A as a sign to act, or similarly difference in the "culture" or managerial could trigger a reevaluation. What is clear also from a meta analysis of turnover in general by Griffeth, Hom, and Gaertner (2000) is that the motivations for turnover are many and varied, but in the case of strong shocks it is argued that higher than average turnover is to be expected. Additionally, as we have seen in Chapter 2 and Chapter 3 spatial mobility is not taken lightly, given tendency of scientists to stay in their city. If a company is acquired by an external firm which might require the inventors to relocate, we would also expect several inventors to look for employment with a company located closer to home. Unfortunately in our data we do not have information on whether the M&A could have required a reloca-

tion, something that might be addressed by information on the location of the research centers of the two companies. What we can control for is however how "exclusive" an inventor is, specifically the proportion of patents filed for his company. We should expect, that if an inventor perceives the M&A to be a shock or adverse to his career prospects, he would look outside opportunities, and having collaborated with several outside companies his ability would be known to those companies as well as having already shown his ability to switch employer.

We look therefore in this work at the specific question: "Are inventors working for acquired companies more likely to leave for a third company?".

It is this stream of research that we want to contribute to and provide further evidence that acquisitions are negative shocks and important employees respond by leaving the company and take their know-how with them.

4.2 Data

For the analysis we require two main sources of information, (1) when M&A events took place and which companies were involved, and (2) Patents Data containing both details on Assignees (i.e. Firm applying for Patents) and their inventors (i.e. Employees of the company).

Specifically, for the identification of M&A events and the relevant companies we rely on the following datasets, "Thomson Reuters Recap"[4] and "Evaluate"[5]. Moreover, we use the disambiguated and georeferences Patent Dataset by Morrison, Riccaboni, and Pammolli (2017) offering a comprehensive set of patents filed with the USPTO, EPO and PCT.

The "patent data" contains 9,290,268 Patent Applications filed between 1978 and 2010 as well as disambiguated inventors and Assignees. Additionally, we have data on the geo-location of the inventors at the

[4]Now the dataset is sold and maintained by Clervite. https://clarivate.com/products/cortellis/cortellis-deals-intelligence/

[5]The data has been obtained from http://www.evaluate.com/

time of application as well as the IPC classification of the patent. The latter we use to estimate technological distances among inventors and firms. A high-level overview of the data available and the relationships between assignees, patents and inventors are illustrated in Figure 25. The coverage of the dataset is however not complete past 2005, as the sudden drop in patent applications past 2005 suggests (shown in shown in Figure 26). For this reason we do not rely on any data in the dataset after 2003. This restricts our analysis to smaller time period. It is however necessary to omit this part of the data, so as not to falsely classifying an inventor as leaving only because we do miss observations.

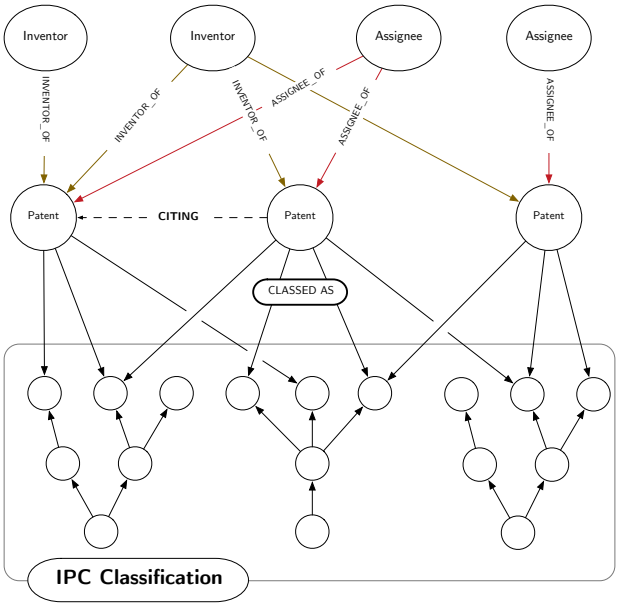


Figure 25: The Patent database offers information on several types of relationships between Patents, Assignees, inventors and Paten Classifications. With the dataset we are able to identify which inventors worked for which Assignees (i.e. Companies) and under what IPC these patents have been classified.

"Thomson Reuters RECAP" contains data on major deals in the Phar-

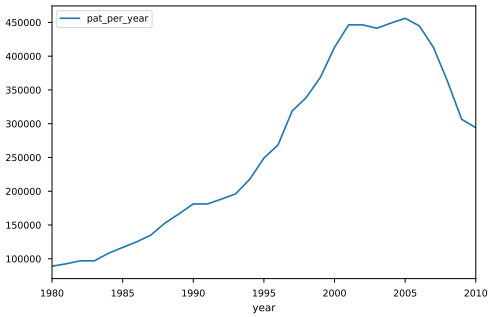


Figure 26: Patent Applications filed per year available in the dataset.

maceutical sector from 1981 to 2012 and covers 46,135 deals (of various type). RECAP contains detailed information on the parties, dates and types of deals in the Pharmaceutical sector. Specifically, for our purposes we are interested in the 3,794 outright acquisition events[6]. Similarly, in the Evaluate Dataset we have 6,604 Acquisition events from as early as 1980 up to 2018. Since we need several years after the event to assess retention we will only use deals up to 1998. Hence we chose only deals falling within the 1990 and 1998 interval. After checking manually the deals to be genuine and proper acquisitions (i.e. not only announcement or falling through) we are left with 135 acquisitions to work with. This implies that we have 135 acquired firms. We use the remaining data on Mergers and Acquisitions, to identify and exclude companies and inventors who have been acquired or worked for an acquired company. In this way we exclude from our set of potential controls anyone who could have been influenced by an acquisition. The number of acquisition events per year found in the RECAP and Evaluate are shown in Figure 27.

[6] An "Acquisition" is defined in RECAP as: One company acquires legal control (greater then 50% of voting shares) of the other company, including both assets and liabilities. Acquisitions may be paid with cash or through an exchange of shares from the buying company to the selling company".

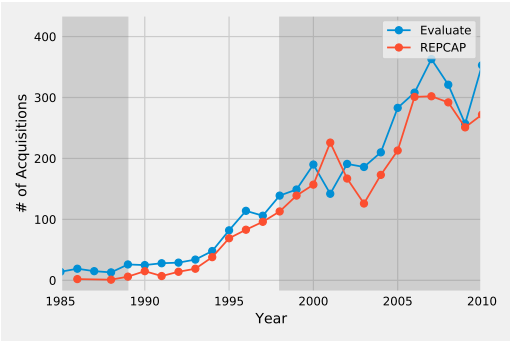


Figure 27: Number of all "Acquisitions" available in RECAP and Evaluate. The period of interests, 1990 to 1998, is highlighted.

4.3 Methodology

We set out to estimate the effect of an Acquisition on the probability that an employee at an acquired firm stays with the firm or leaves it for a third firm. For this purpose we set up a differences in difference style regression with matching at both firm and inventor level.

Specifically, at a high-level we carry out the following analysis. We match treated (i.e. acquired) and control firms (i.e. never subject of M&A) on patenting rate (i.e. patents per year), age (i.e. years since first patent application) and technological distance (i.e. cosine similarity on IPCs). Given this firm level match we proceed to match inventors working for the control firms with the inventors working for the treated firms, thus obtaining pairs of inventors which work for similar firms and are themselves similar on observables. With the matched pairs we then proceed to estimate their probability (1) to continue patenting and (2) and if so which firm is the assignee. A difference in these probabilities emerging after the acquisition event implies that the treated group has responded to the shock. In this section we will describe in detail how the matching is done and how the effect is estimated.

The identification of who *is a valid treated employee* is ambiguous for several reasons. First, new employees enter constantly and closing in on

the actual acquisition event, the reason for working might be driven by the event itself. For example, new employees are hired in expectation of the takeover or employees are not hired due to uncertainty. Similarly, not all inventors named on a Patent can be unambiguously identified as working for *the treated company*. We call this type of inventors "freelancers" as no clear employment relation can be inferred. We find in our data several instances of inventors working within a given period intermittently filing patents for unrelated Assignees, with no clear signal of affiliation. For example, an inventor is named on 4 patents assigned to 4 different companies, thus filing at most 25% of his patents for the focal company.

We address the first concern by setting up a framework to clearly define which inventors we are interested in and which we exclude from the treated group. A more detailed discussion on this setup follows below. With regards to the "freelancers" problem we have adopted the following convention. Inventors filing exclusively for the company of interest are clearly employees. If in the identification phase we do find that an inventor has filed for multiple companies we consider him to be an employee of a firm if and only if he has filed more than $1/3$ of his patents in the relevant period for focal company. If however he files less than $1/3$ for the focal company, we drop him from the pool of treated employees and exclude him from the pool of potential controls. This last step, guarantees that we not pollute the control set with "partially treated" individuals.

Here we introduced now the various phases in the framework. The setup defines, (1) when inventors are identified as treated or controls, and (2) their decisions on staying with the company before and (3) after the event. The event year is denoted by t_E and the various phases are defined as offsets from this base year. Moreover, we identify treatment at both firm and inventor level with a superscript, where 1 is treated and 0 not treated, and with a subscript we denote their identity. Accordingly, a treated inventor k is denoted as I_k^1 and an untreated firm h is denoted as F_h^0.

The framework, i.e. identification and the evaluation periods, is com-

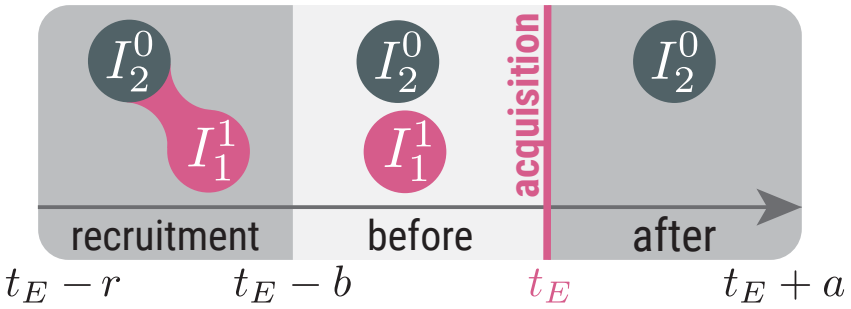


Figure 28: In the "recruitment" phase the matching covariates for the inventors are computed. And the matching of the inventors is done (e.g. treated inventor I_1^1 is matched to I_2^0). In both "before" and "after" periods being active and staying with the company are measured. In this example I_1^1 is active in "before" but we do not see him again in "after", I_2^0 on the other hand is active in both periods.

prised of three distinct phases (see Figure 28). In the first phase, the *"recruitment phase"* in the window $[t_E - r, t_E - b)$, valid treated and control firms and inventors are identified. In the second phase, the *"before phase"*, we observe the various outcome of interest *before* the event, and covers the period $[t_E - b, t_E)$. Finally, in the third phase $[t_E, t_E + a)$., the *"after phase"*, we evaluate the same outcome again, knowing that the event took place.

As noted in the introduction, studies by both Ravenscraft and Long (2000) and Higgins and Rodriguez (2006), show respectively that firms under stress tend to be acquired and that firms nearing a patent expiration of important compounds are more likely to engage in M&As. It is for this reason that we think that the crucial "recruitment" phase must be pushed back before the actual event sufficiently to reduce the probability that these firms do not operate in what could be called a "going concern". Ideally, we would have access to financial details to match companies. This would allow us to match on financial distress, yielding a match of two firms that have a similar propensity to be acquired. To alleviate this concern as best we can with the available information, we recruit firms and inventors at least 3 to 4 years prior to th event. How-

ever, by requiring that companies and inventors be observable for at least 4 years after they are first recruited to observe their decision to leave the company we reduce the potential pool of inventors considerably as well as eliminating small and young bio-tech startups from our analysis.

In the "recruitment" phase several things happen[7].

1. Companies which are acquired outright in the relevant period are marked as treated and added to the pool of treated firms.

2. These firms are then matched through their names in REPAC and Evaluate with the disambiguated Assignees in the Patents dataset. By doing so we obtain matched Patents for these Firms and by extension the inventors working on these patents (i.e. potential employees).

3. Treated and control firms are matched in a one to many fashion. This means that we match a treated firm with possibly multiple controls if they are "sufficiently" similar. If however, no adequate match can be found the firm is dropped.

4. Treated inventors working for the acquired firm h (F_h^1) are matched one to one with employees working for one of the matched control companies of (F_i^0). The result is a list of treated/control inventor pairs which are the focus of the analysis in the subsequent stages.

In the "before" phase (i.e. $[t_E - b, t_E)$) two outcomes for each inventor are observed. First, do they continue filing patents in this period? If they do we record them as "active". We should note that cessation of patent production after the event, could simply be caused by inventors being placed in administrative or managerial positions, but in fact do not leave the company. Those inventors, both treated and controls, who are active in the "before" period are then classified as either filing patents for the focal company and thus "staying on" or exclusively for a third party and thus "leaving". Since the acquisition did not yet take place, ideally the propensity to leave the company should not be affected, however

[7] the datasets will be described in more detail in Section 4.2

as suggested by Ravenscraft and Long (2000) acquisition targets tend to show signs of distress quite some time earlier.

Finally, we look at the "after" period (i.e. $[t_E, t_E + a)$), the time interval after the acquisition took place (including the acquisition year). If there would be no effect of acquisition on the propensity to leave we should find that treated and controls have the same propensity to leave the company, if on the other hand we find that treated companies experience a higher level of turnover we have an indication that the acquisition, if not necessarily caused the exodus, at least hastened it. Since the acquisition took place at the beginning of this period, the legal identify of the acquiring firm is no longer guaranteed to be the same. In other words the acquired company as a consequence of the takeover has either changed name or has ceased to exist and become a division of the acquiring company. To account for this possibility in the after period an inventor is considered to be "staying on" if he files at least one patent for the acquiring or acquired company or if the company name he files for now has a high string similarity to either of the two. If there is a match which is sufficiently close[8], we check manually if the new assignee name is correct.

4.3.1 Matching Firms

To be able to claim any causality of an acquisition on defection rates we need to construct a valid set of control Firms and inventors to compare the defection rate against.

In this work we adopt a two stage matching procedure, whereby as noted above we first match treated con control firms and in the second stage we match treated and control inventors. Conceptually, these two steps are illustrated in Figure 29.

In the first matching step, the *firm matching* step, we try to find companies which are similar to the treated companies we have identified. Since Pharmaceutical companies are R&D intensive and have a high propen-

[8]To determine in a first parss the string similarity we compute the Liechtenstein distance between the acquiring firm and the potential alias and the acquired firm and the potential alias.

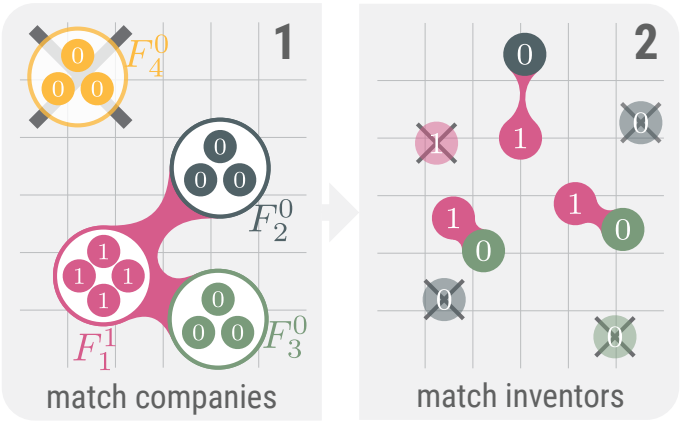


Figure 29: The matching of treated to untreated inventors is carried out in two stages. First, treated companies (e.g. F_1^1) are matched up with untreated companies (e.g. F_2^0 and F_3^0, one to many) on their covariates. In the second stage, the treated inventors (labeled 1) are matched to control inventors (labeled 0) in a one to one fashion. Any unmatched inventors are discarded as is any company for which we do not find a good match.

sity to patent their work, we will make extensive use of patent data to find good matches. In a first stage we identify all treated firms F_i^1 and their patents in the recruitment phase. For each treated firm we obtain its "IPC" technological profile (τ) at main-group level (e.g. "C08G063"). We obtain a vector of the form $\tau_h = [n_{ipc1}, n_{ipc2}, n_{ipc3}, \ldots]$, where each entry corresponds to the number of patents published in that IPC class by the company. Given their technological profile, we obtain a set of candidate control firms who have published a patent in those fields within the period. In other words, if a treated company F_h^1 filed a patent in IPC main-group `C08G063` in the recruitment phase, we identify all companies, which also filed a patent under `C08G063` in the same period. To make sure that we do not pollute the control set with treated companies we exclude any company that was or will be subject to any M&A event type listed in RECAP or Evaluate. We will denote this set of potential control firms for firm F_h^1 with $\mathcal{P}_h^0$. For each potential control company in $\mathcal{P}_h^0$ we compute "technological similarity", "age similarity", "patenting

rate similarity" and combine these measures into a single "similarity" through a weighted average. For the main analysis we have weighted technological similarity at 0.5 and the other two similarities with 0.25. Specifically, we compute the technological distance between a treated company F_h^1 and its potential controls $\mathcal{P}_h^0$ by constructing a tech profile for all companies, as defined above. We define the technological similarity (s_τ) between two profiles τ_h and τ_k using the cosine similarity measure as shown in (4.1).

$$s_{hk}^{\tau} = \frac{\tau_h \cdot \tau_k}{||\tau_h||_2||\tau_k||_2} \tag{4.1}$$

This measure is equal to 1 if the two vectors are identical and 0 if there is no common element (i.e. no shared IPC). The two similarity edges closer two one the more elements the two vector share and the closer the number of patents filed in these IPCs are.

We then compute the number of patents applied for in the same period (proxy for the size of the company) and its age (years since the firs Patent has been filed) and obtain a similarity defined by (4.2).

$$s_{hk}^{\text{age}} = \frac{|\text{age}_h - \text{age}_k|}{\text{age}_h + \text{age}_k} \tag{4.2}$$

$$s_{hk}^{\text{patents}} = \frac{|\text{patents}_h - \text{patents}_k|}{\text{patents}_h + \text{patents}_k} \tag{4.3}$$

These similarities are again 0 if the two firms have the same value and edge closer to 1 the wider the gab between the two is. The final similarity score s_{hk} between F_h^1 and F_k^0 is then given by (4.4).

$$s_{hk} = w_\tau s_{hk}^{\tau} + w_{\text{age}} s_{hk}^{\text{age}} + w_{\text{patents}} s_{hk}^{\text{patents}} \qquad w_\tau + w_{\text{age}} + w_{\text{patents}} = 1 \tag{4.4}$$

If a potential control firm in $\mathcal{P}_h^0$ is 90% similar to the treated firm, it is included in the control set $\mathcal{C}_h^0$ of firm F_h^1 and all its employees become potential controls for the treated employees of F_h^1.

4.3.2 Matching inventors

In a similar fashion to the firm matching we match treated and control inventors. For a treated inventor I_j^1 working for treated firm F_h^1 we obtain

all inventors working for matched control companies $\mathcal{C}_h^0$ and compute again several similarity metrics. We match again on the IPC technological profile (τ_j) for the treated inventor I_j^1 and its controls by using cosine similarity (see Equation (4.1)). Similarly, we use tenure (i.e. years since first patent with company) and patenting activity in the recruitment period to determine similarity (see Equation (4.2)). These similarities are again aggregated through the same weighted average and the closest match is chosen as the matched inventor, provided that it is at least 90% similar.

At the end of the two matching steps, i.e. firm matching and inventor matching, we are left with a set of treated-control inventor pairs, which we use in the analysis.

4.4 Estimation and results

Now that we have matched treated and control inventors we want to estimate how they behaved in the "before" and "after" periods. Specifically, we want to assess if the two groups differ in their propensity remain active (i.e. not stop to apply for patents) and second if they do continue their patenting activity into the after period if they do so for either the acquired company or the acquiring company (i.e. stayed on) or for a third party (i.e. left).

For the actual estimation we still have 3 hyper parameters to choose, namely r, b and a. These three parameters jointly define the length of the recruitment period (r), the length of the before period (b) and the length of the after period (a). For the main analysis we fix ($r = 7, b = 4, a = 4$), however we have carried out the same analysis yielding the same defection rate for ($r = 6, b = 3$) and varying lengths of a. The choice to conduct the recruitment well in advance (i.e. 7 to 4 years before) is conservative. It is conservative, since we lose a considerable amount of inventors, due to the fact that for any additional year after recruitment the number of active scientists can only decrease. Moreover, recruiting well in advance allows us to mitigate the fact that on average acquired firms experience financial distress nearing their takeover. We will also use the hyper pa-

	treated	
age	0.0112	(0.45)
tenure	-0.00200	(-0.06)
patents	0.0135	(0.79)
year=1993	0.0481	(0.18)
year=1994	0.0700	(0.32)
year=1995	0.0472	(0.14)
year=1996	0.214	(0.45)
year=1997	-0.0105	(-0.05)
year=1998	0.0170	(0.08)
Constant	-0.179	(-0.76)
Observations	1214	

t statistics in parentheses

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

Table 13: Propensity to be treated on covariates. The Logistic regressions shows that the propensity to be treated is not correlated with any of the covariates, the inventors have been matched on.

rameter a, the length of the window to observe defections as a means to estimate the effect of the acquisition after the event. So for example, by comparing the estimate for $a = 3$ with $a = 6$ we can infer if defection rates have increased in the 3 years that followed or remained unchanged, implying no effect in the last 3 years. In fact, we do find that

Before we move to the actual estimation, we show that the matched inventors on observables (i.e. age, tenure, patents before) are not indicative of treatment status (i.e. work for an acquired company) through a Logit regression. In other words we show that there is no bias in the those variables across groups. The results, for the particular parameter choice of ($r = 7, b = 4, a = 4$) shown in Table 13, suggest that neither of the covariates is predictive of treatment. This provides some measure of confidence that the matching procedure has worked. For the main chosen windows in the main analysis of ($r = 7, b = 4, a = 4$) we have 1,214 inventors in our sample with an equal number of treated and untreated (i.e. 607). Of these 370 are active in the after period.

Variable	Definition
active	If the inventor applies for at least one patent in the given period he is marked as active and incative otherwise (1=active, 0=inactive)
stay on	If in the period the inventor files any patent for the company he stayed and the value is 1 and 0 otherwise
deal year	The year in which the deal took place
acquired	The value is equal to 1 if the inventor is working for firm that has been acquired
after	Identifies to which period this observations belongs (i.e. before the deal, or after the deal)
age	Number of years the inventor has been active by the time the deal took place — an inventor applying for patent in 1990 and his company is acquired in 1995 is 5 years old
tenure	Number of years the inventor has been with the company at deal year
patents	The number of patents applied for in the recruitment phase by the inventor
exclusivity	Proportion of Patents filed for the company in the recruitment phase by the inventor

Table 14: Variables used in the Estimation

The dependent variables of interest for the analysis are *active* and *stay on*. We observe each inventor twice in the dataset, once in the before period and once in the after period. This means that the time subscript t is equal to 0 before and 1 after. In addition to these two variables we use several additional covariates to control for age and patenting activity as listed in Table 14

We estimate a Heckman 2 stage selection model with a Difference in Difference specification. This, in addition to the pairwise matching of treated and controls, should help us to alleviate several concerns. We use the Heckman selection model to control for the censoring, i.e. inventors not being active. If an inventor is not active, we are not able to observe whether she stayed on or defected to a third company. The matching on observables helps us to make sure that we compare two similar groups

of inventors, working for similar companies. The Difference in Difference approach, assuming that the common trends assumption is not violated, allows to estimate the effect of acquisitions on defection rates on the treated population.

The Heckman regression requires in the selection equation an exclusion restriction, i.e. a parameter which influences the probability to be observed/censored but does not in turn affect the outcome of the main regression. With the selection equation, i.e. the first stage, the propensity to be observed is estimated and this propensity is corrected for in the second stage. In other words, the exclusion restriction should affect an inventors continued activity, but not his permanence with the company. We propose as an exclusion restriction the number of patents filed in the recruitment period. The continued production of patents is an indication of momentum, in the sense that having patented more it is more likely to continue this activity. Highly productive individuals might also be more valuable for the company but their proven expertise mean also that they have more outside options and could defect if they so choose. We therefor argue that this variable, only used in the first stage, allows us to identify which inventors are most likely to be active.

We control in the second stage for the "exclusivity" of the inventor to his employer, by controlling for proportion of patents he has assigned to the focal company over all patents he has applied for. This means that an employee with exclusivity equal to 1 has filed patents exclusively for the focal company, and 0 otherwise. Note, that due to the "freelance rule" discussed in Section 4.3, all exclusivity values are at least $1/3$.

We estimate in the first stage (probit) the propensity to be active using the specification shown in Equation (4.5). In the second stage, an OLS or linear probability, we estimate the interaction of acquired $\times$ after — the DiD parameter and main focus of this analysis.

$$\begin{aligned}\text{active}_{it} = \beta_0 &+ \beta_1 \text{acquired}_{it} + \beta_2 \text{after}_{it} \\ &+ \beta_3 \text{acquired}_{it} \times \text{after}_{it} + \beta_4 \text{age}_i \\ &+ \beta_5 \text{age}_i^2 + \beta_6 \text{tenure}_i + \beta_7 \text{tenure}_i^2 \\ &+ \beta_8 \text{exclusivity}_i + \beta_9 \text{deal year}_i + u_{it}\end{aligned} \tag{4.5}$$

$$\begin{aligned}\text{stayed on}_{it} = \beta_0 &+ \beta_1 \text{acquired}_{it} + \beta_2 \text{after}_{it} \\ &+ \beta_3 \text{acquired}_{it} \times \text{after}_{it} + \beta_4 \text{age}_{it} \\ &+ \beta_5 \text{age}_i^2 + \\ &+ \beta_6 \text{patents}_i + \beta_7 \text{deal year}_i + u_{it}\end{aligned} \tag{4.6}$$

The results of this regression are shown in Table 15. We note first and foremost that the interaction term acquired $\times$ after, the DiD effect, for the scientist, who stay on, is significantly negative. In other words we do find that according to the OLS estimate, inventors working for a treated company are 27% less likely to keep on working then the control group in the 4 years that follow the acquisition event. We also find that acquired is equal across groups for both the probability to be active and the probability to leave the company. This means that the matching procedure has been able match treated and controls such that there is no difference in outcomes before the event. This is an important result, since it implies that in the difference in difference estimation we do not need to rely on the first difference (i.e. difference before the event).

As we would expect we find across the board, that irrespective of treatment the two groups are less active in both periods and retention decreases as suggested by the after parameters. The exclusion restriction, i.e. patents parameter, in line with our hypothesis shows that being active is strongly positively correlated with continued activity.

As a robustness check we vary the length of the after period (i.e. a) and use 3, 4, 5 and 6 years. Within the first 3 years we have a DiD effect of

Stage 2	stayed on	
stayed		
acquired	-0.00719	(-0.28)
after	-0.0939**	(-2.70)
acquired × after	-0.269***	(-6.26)
age	-0.0285	(-1.25)
age^2	0.00137	(1.39)
tenure	-0.0124	(-0.48)
tenure^2	0.000570	(0.49)
exclusivity	0.841***	(15.94)
Constant	0.237**	(2.70)
Stage 1	active	
acquired	-0.00725	(-0.11)
after	-0.906***	(-10.33)
acquired × after	-0.331**	(-2.79)
age	0.189***	(3.83)
age^2	-0.00793***	(-3.32)
patents	0.353***	(15.84)
Constant	-1.119***	(-4.45)
Deal Year Effects	Yes	
athrho	-0.409***	(-4.73)
lnsigma	-1.040***	(-50.63)
Observations	2,428	
Censored	1,377	

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Regression Results of the 2 stage Heckman regression. The first stage estimates the propensity to be active and the second stage, the main focus, is on OLS on the probability to leave the company. A linear probability model.

-25%, in the first 5 years -31% and at 6 years 32%[9]. This results suggests that the bulk of the effect takes place around the event. However, since we know that potential take-over candidates might experience some sort of financial or productivity distress and patent applications are a delayed signal of activity. Therefor we cannot rule out that some of the defections are not caused by the Acquisition but are due to distress.

4.5 Discussion and Conclusions

From the analysis it emerges that defection rate is significantly higher following an acquisition. In fact, we find that the defection rate before the event is equal to the control group, but in the 3 to 6 years that follow this defection rate goes from 26% to 31% above and beyond the control group. Surprisingly, the descriptive analysis of inventor defection by Ernst and Vitt (2000) finds that one third of inventors leave their respective company, a value in line with our finding. Considering that R&D employees are considered by the firms themselves especially important (Ranft and Lord, 2000), our results suggesting a high turnover can have detrimental effects on the post-merger performance and integration. We do find also that the degree to which an inventor works exclusively for the focal company is strongly indicative of continued work there. Conversely this means that inventors, who have worked for external companies before and in line with our initial hypothesis have more "outside options" tend to defect at a higher rate. This findings suggests that if an acquisition target has inventor working exclusively there, their retention is easier. This result should however be considered in light of the fact the young researchers and possibly less productive researchers are very likely to be "exclusive" to the company, however their value to the acquiring firm might not be equal to more seasoned inventors.

These findings are of interest for managers and analysts evaluating the success of M&A deals. Specifically those deals that have a strong focus on intellectual property and retention of tacit knowledge. Further

[9]The regressions are all equal to the main regression using an after window of 4. All other parameters have similar magnitude and significance.

analysis and more information on the particular deal type (e.g. interested only in equipment, vertical integration) could help to identify in which cases R&D retention is an objective and more importantly if the post-merger integration in these deals was successful from this perspective.

With this study we contribute to the literature on post-merger integration, offering additional evidence that acquisitions are accompanied by a higher than expected defection rate. Our analysis has several strengths. We employ a matching, difference in difference approach in addition to the Heckman selection model to control for various confounding factors. The disambiguated patent and assignees dataset allows us to track employment relationships across many companies focusing on our employees of interest (R&D).

However, we did not used any financial information regarding the acquired or acquiring firms. This is in large part due to the difficulties to obtain annual reports on these firms, given that they were active in the Nineties. And similarly we do not have personal information on the inventors (i.e. physical age, gender, education) which could alleviate several concerns in the matching stage. Our proposed methodology could be applied to a more recent vintage of M&A events by using a more up to date patent database, possibly reaching into 2018 and thus further strengthen the conclusion. However, the finding is robust to various checks (i.e. varying windows sizes) and the magnitude of turnover is in line with Ernst and Vitt (2000), Ranft and Lord (2000) and Carriqury (2017).

In summary, with this work we have shown that inventors, crucially important, for the success of firms operating in the R&D intensive pharmaceutical sector, respond to acquisitions by leaving. This effect is reduced for inventors who did not patent for an other company. However, the defection rate is high at 30%, a number which is confirmed by several other studies using different data sources and methods. What is clear, is that inventors might in fact vote with their feet and leave to the detriment of the acquiring firm.

Appendix A

Mobility Supplementary Materials

	Year	Affiliation City	PubMedID
1	2003	Stony Brook, NY, USA	12703729
2	2003	Stony Brook, NY, USA	12595470
3	2005	Kansas City, KS, USA	15936007
4	2005	Stony Brook, NY, USA	15791955
5	2005	Stony Brook, NY, USA	15944300
6	2005	Milwaukee, WI, USA	16299285
7	2007	Milwaukee, WI, USA	17311921
8	2007	Milwaukee, WI, USA	17490406
9	2008	Boston, MA, USA	18566416
10	2008	Stony Brook, NY, USA	18591234

Table 16: Example of career path of a specific author (Zhang Y.). For each record we have the year of publication, the city of the affiliation and the relative PubMed ID identifying the paper

Asymmetric mobility D-core

To further understand the diversity in the exchange between cities we look at the D-core decomposition (Giatsidis, Thilikos, and Vazirgiannis,

2013).

The D-core analysis (directed generalization of the k-core) allows us to analyze simultaneously the centrality and "coreness" of a city while taking the asymmetric nature of global mobility into account (i.e. the cities feeding scientists to a given location are not the same they receive scientists from). This algorithms instead of a list of k-Core, yields a matrix of (in-degree; out-degree)-cores, see Figure 30.

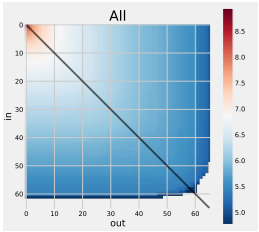
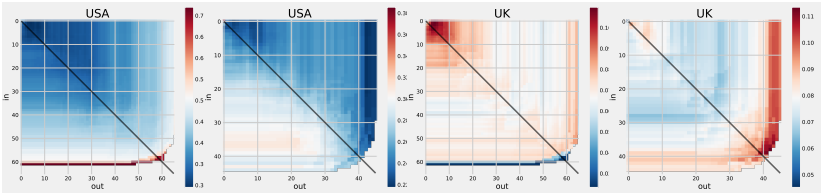


Figure 30: Every square at coordinate (k; l) corresponds to the (in; out)-core of the mobility network in 2004. The color indicates the number of cities in the D-core (in $\log_{10}$).

In the All matrix (see Figure 30) the squares have been colored according to the size of the core and the rest according to the proportion of cities therein contained belonging to that country (see. Figure 31). We observe in the "All" matrix, that the D-core decomposition is skewed toward having high out and low in-degrees, as can be seen by the slightly brighter blue on the right frontier than on the bottom portion (high in, low out). Comparing this result with Figure 31 for the USA, we see that US cities are represented more in the high in, low out portion of the plot, the opposite of what we would expect if all cities were distributed according to "All". This means that most US cities which are in the most central cores have more cities feeding into them than they are feeding. In other words central US cities on average source from a wide variety of cities but their scientists move to a less diverse set of cities.

Since the size of each country influences the number of cities they contain it could be the case that what we observe represents mainly the national configuration of cities. To explore this idea further we can look

at how the D-core looks like if we remove all national connections. In other words we leave for each city only its international connections such that the D-core they are part of is only induced by being part of an international network (see right side of Figure 31). We observe that the USA is still marked by a strong presence in cores with stronger IN degree than OUT. This suggests that US cities sources from a wide variety of cities, but are the origin of moves to a more restricted set of cities.



(a) (b) (c) (d)

Figure 31: D-core profiles for USA (a-b) and UK (c-d). The coordinate $(k;l)$ corresponds to the $(k;l)$-core of the mobility network. The matrices show the proportion cities belonging to the US contained within a given $(k;l)$-core. The color scale has been adjusted such that the average proportion across all D-cores is white. Complete network (a, c); Only international moves (b, d). Note: the transparent cells on the border belong to empty $(k;l)$-cores.

National Border Effects

Table 17: Breakdown of communities by country. For each modularity class the number of cities belonging to a country are listed along with their proportion of cities in the class. For example community 15 is composed of 25% Swedish, 22% Finnish, 20% Norwegian and 16% Danish cities as well as 17% smaller cities. Note: only countries with at least 5% member cities are listed.

Community (x)	Country (c)	% of cities of c in x	No. Cities
1	USA	92%	392
2	USA	91%	161
3	Spain	27%	65

Community (x)	Country (c)	% of cities of c in x	No. Cities
	Mexico	22%	53
	Argentina	11%	27
	Chile	6%	14
	USA	6%	15
4	USA	94%	355
5	USA	98%	197
6	USA	85%	197
7	France	57%	225
	Belgium	10%	40
8	Germany	69%	444
	Switzerland	14%	88
	Austria	7%	46
9	Russia	45%	53
	Taiwan	21%	25
	USA	20%	23
	Ukraine	9%	10
10	USA	95%	151
11	USA	91%	192
12	Australia	41%	81
	USA	12%	24
	Thailand	11%	22
	New Zealand	10%	20
13	Czech Republic	36%	39
	Croatia	16%	17
	Slovenia	12%	13
	Serbia	11%	12
	Slovakia	9%	10
	USA	6%	6
14	Netherlands	60%	110
	Belgium	14%	26
15	USA	51%	41
16	Sweden	25%	86
	Finland	22%	75

Community (x)	Country (c)	% of cities of c in x	No. Cities
	Norway	20%	69
	Denmark	16%	55
17	Korea	88%	43
	USA	10%	5
18	USA	89%	71
19	USA	98%	122
20	Japan	81%	91
21	Canada	54%	127
	Iran	11%	25
	Saudi Arabia	9%	22
	USA	8%	19
	Egypt	7%	17
22	Greece	80%	35
	Jordan	9%	4
	Cyprus	7%	3
23	Brazil	58%	55
	Portugal	21%	20
24	Israel	70%	39
	USA	21%	12
25	China	70%	171
	Hong Kong	7%	17
	Malaysia	7%	18
26	Poland	92%	57
27	UK	66%	315
28	Turkey	91%	53
29	Italy	75%	121
	USA	6%	10
30	USA	83%	84
31	Nigeria	47%	34
	South Africa	39%	28
32	India	73%	122

Community (x)	Country (c)	% of cities of c in x	No. Cities
	USA	8%	14

Regression Analysis

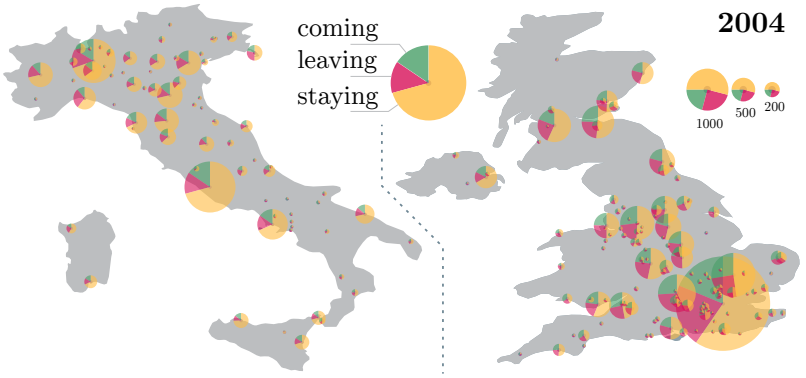


Figure 32: These two maps show the scientists population for the UK and Italy, which have at least 30 scientists stationary there in the period 1999 to 2008 (move year 2004). The pie chart indicates the proportion of scientist which are incoming, leaving and or staying. The size of the pie-char is proportional to the sum of all three types.

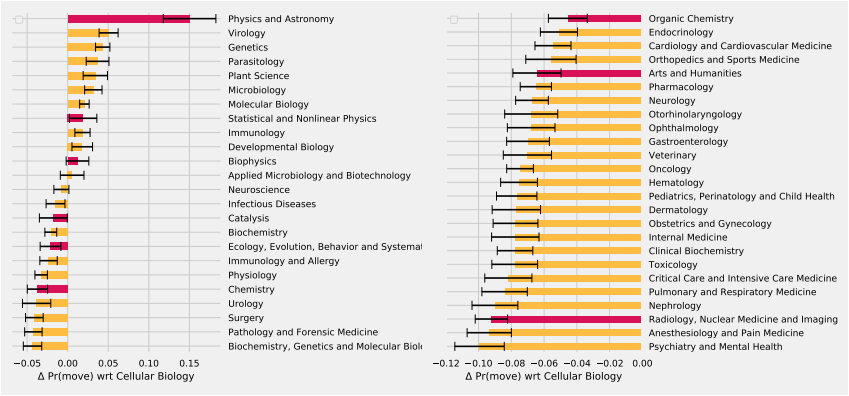


Figure 33: Marginal probability to move compared to Cellular Biology (i.e. the largest field). The 95% confidence interval is illustrated as black bars.

	$\Pr(\text{move})$		k-core destination	
k-core source	-0.000362	(-0.45)	0.0128	(0.77)
productivity	0.0218***	(4.68)	4.313***	(6.57)
inter-move	0.109***	(64.92)	-2.583***	(-6.77)
2002	0.0335**	(3.12)	10.62***	(10.07)
2004	0.0452	(1.72)	29.04***	(16.70)
2002 $\times$ productivity	0.00283	(0.88)	0.943**	(3.26)
2004 $\times$ productivity	-0.00414	(-1.12)	1.543***	(3.59)
$\log$(km distance)			5.301***	(5.43)
$\Pr$(move other fields)	1.510***	(6.12)		
Constant	-1.711***	(-3.89)	118.5***	(6.79)
Origin country effects	Yes		Yes	
Age effects	Yes		Yes	
Field effects	Yes		Yes	
Observations	1,363,280		433,023	
$\tanh(\rho)$	-0.14 (-4.60)	$\log(\sigma)$	4.46 (39.80)	
Log pseudo-likelihood	-3,193,349			

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 18: Heckman two stage regression for the centrality measure "k-core". The standard errors have been clustered at source city for the first stage ($\Pr$(move)) and on the destination city for the second stage (i.e. the centrality of the destination)

	Pr(move)		norm. degree	
norm. degree source	-0.574	(-0.38)	0.0174	(0.92)
productivity	0.0217***	(4.35)	0.00249***	(6.53)
inter-move	0.109***	(64.91)	-0.00126***	(-6.05)
2002	0.0294***	(6.15)	0.000775*	(2.17)
2004	0.0352***	(4.12)	0.00310***	(5.81)
2002 × productivity	0.00261	(0.83)	0.000194	(1.21)
2004 × productivity	-0.00470	(-1.39)	0.0000164	(0.07)
log(km distance)			0.00241***	(4.30)
Pr(move other fields)	1.534***	(6.33)		
Constant	-1.733***	(-3.98)	0.0498***	(5.12)
Origin country effects	Yes		Yes	
Age effects	Yes		Yes	
Field effects	Yes		Yes	
Observations	1,363,280		433,023	
tanh(ρ)	-0.14 (-4.64)	log(σ)	-3.117 (-31.04)	
Log pseudo-likelihood	-75,548.17			

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 19: Regression results for Mobility and relocation choice. This are the results of the Heckman two stage regression for the centrality measure "normalized degree". The standard errors have been clustered at source city for the first stage (Pr(move other fields)) and on the destination city for the second stage (i.e. the centrality of the destination)

	Pr(move)		PageRank destination	
PageRank source	-5.019	(-0.20)	0.0123	(0.71)
productivity	-0.0599**	(-3.21)	0.000356***	(3.95)
inter-move	0.145***	(37.02)	-0.000100***	(-4.63)
2002	0.0527**	(2.67)	-0.000101	(-0.99)
2004	0.0747**	(2.91)	-0.0000714	(-0.68)
2002 × productivity	-0.0137	(-0.92)	0.000148*	(2.57)
2004 × productivity	-0.0384*	(-2.01)	0.000187**	(3.12)
log(km distance)			-0.0000381	(-0.74)
Pr(move other fields)	1.692***	(3.29)		
Constant	-1.218***	(-5.90)	0.00448***	(4.58)
Origin country effects	Yes		Yes	
Age effects	Yes		Yes	
Field effects	Yes		Yes	
Observations	150,384		39,230	

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 20: Regression results for Mobility and relocation choice. Heckman two stage regression for the centrality measure "PageRank" considering only the subset of authors which are above the 90th percentile. The standard errors have been clustered at source city for the first stage (Pr(move)) and on the destination city for the second stage (i.e. the centrality of the destination)

	Pr(move)		PageRank destination	
PageRank source	-4.840	(-0.21)	0.0122	(0.94)
productivity	0.0175*	(2.37)	0.000136***	(3.68)
inter-move	0.111***	(39.33)	-0.0000650***	(-4.31)
2002	0.0179	(1.51)	0.000146*	(2.17)
2004	0.0747**	(2.91)	-0.0000714	(-0.68)
2002 × productivity	0.00278	(0.36)	0.0000827**	(2.63)
2004 × productivity	0.00555	(0.74)	0.0000105	(0.37)
log(km distance)			0.000115*	(2.16)
Pr(move other fields)	1.547***	(5.74)		
Constant	-1.099***	(-8.81)	0.00337***	(6.21)
Origin country effects	Yes		Yes	
Age effects	Yes		Yes	
Field effects	Yes		Yes	
Observations	135,931		40,945	

t statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 21: Regression results for Mobility and relocation choice. Heckman two stage regression for the centrality measure "PageRank" with a random 10% subsample.

Appendix B

Stem Cell Research

B.1 State Legislation

Table 22: Legislation listed as Key Votes on the Vote Smart Website, a non partisan voter information Website. This is an excerpt of State Laws and Amendendments, which are listed under the "Stem Cell Research" theme or mention "stem cells" explicitely in their text and have been approved or ratified

State	Year	Bill Title
AL	2014	Authorizes Health Care Providers to Abstain fr...
AZ	2010	Treatment of Human Embryos
AZ	2017	Requires Physicians to Attempt to Revive Viabl...
DE	2007	Delaware Regenerative Medicine Act
GA	2007	Umbilical Cord Blood and Stem Cell Banks
GA	2009	In Vitro Embryo Creation Limitations
IA	2007	Stem Cell Research and Penalties
ID	2006	Controlled Substances Relating to Pregnant Wom...
ID	2010	Authorizing Health Care Professionals to Refus...
IL	2007	Stem Cell Research and Human Cloning
IN	2007	Operating Budget
IN	2008	Pre-Abortion Notification Requirement
KS	2011	Amending Statutes Regulating Abortion

Continued on next page

Table 22: Legislation listed as Key Votes on the Vote Smart Website, a non partisan voter information Website. This is an excerpt of State Laws and Amendendments, which are listed under the "Stem Cell Research" theme or mention "stem cells" explicitely in their text and have been approved or ratified

State	Year	Bill Title
KS	2013	Establishes the Midwest Stem Cell Therapy Center
KY	2008	Abortion Regulations
KY	2014	Requires an Ultrasound Prior to an Abortion
LA	2008	Banning Public Funds for Human Cloning
LA	2009	Health Care Protection
MA	2008	Life-Science Bill
MD	2006	Maryland Stem Cell Research Act of 2006
MD	2006	Stem Cell Funding Amendment
MI	2012	School Aid Budget
MN	2007	Public Funding for Stem Cell Research
MN	2008	Surrogate Mother Regulations
MN	2008	Dean Amendment: Stem cell research that does n...
MN	2008	Public Funding for Stem Cell Research
MN	2008	Dean Amendment: Stem cell research that does n...
MN	2016	Appropriates Funds for Education
MO	2012	Provides Exemptions on Participating in Certai...
MO	2014	Authorizes Health Care Providers to Opt Out of...
ND	2009	Amending Abortion Requirements
ND	2011	Definition of Human Being
ND	2013	Defines a "Human Being"
NH	2006	Stem Cell Research Resolution
NH	2012	Prohibits Violent Offenses Against Fetuses
NH	2015	Amends Definition of "Another" for Certain Cri...
NJ	2006	Cigarette Tax Securitization Proceeds Fund; St...
NJ	2012	Amends Surrogacy Laws
NM	2008	Authorizing Stem Cell Research
NM	2009	Allowing Embryonic Stem Cell Research
OH	2008	Abortion and Ultrasound Bill
OK	2008	Ultrasound Requirement for Abortions
OK	2009	Ban on Certain Types of Stem Cell Research
OK	2010	Requiring Women Seeking Abortions to Answer Qu...

Continued on next page

Table 22: Legislation listed as Key Votes on the Vote Smart Website, a non partisan voter information Website. This is an excerpt of State Laws and Amendendments, which are listed under the "Stem Cell Research" theme or mention "stem cells" explicitely in their text and have been approved or ratified

State	Year	Bill Title
OK	2010	Allowing Health Services Employees to Refuse P...
OK	2014	Prohibits Nontherapeutic Research on Human Emb...
OK	2014	Prohibits Certain Embryonic Stem Cell Research
OK	2014	Amends Standards for Abortion Facilities
SC	2007	Pre-Abortion Ultrasound
SC	2009	24-Hour Waiting Period for Abortions
SC	2011	Freedom of Conscience Act
TN	2012	Expands Definition of Offense Against Fetuses
TX	2011	Requires an Ultrasound Prior to an Abortion
TX	2011	Abortion Procedures Bill
VA	2006	Stem Cell Funding Prohibition Amendment
VA	2007	Stem Cell Research Amendment
WA	2010	Legalizing Paid Surrogacy
WI	2008	Corporate Research and Development Tax Credit

	(1) Pop SC	(2) Pop ESC	(3) Pop hESC
log(Pop Tot.)	76.35***	24.05***	9.375***
	(9.36)	(8.16)	(6.84)
Supportive	207.6***	77.50***	43.57***
	(14.19)	(13.93)	(13.43)
Constant	-389.9***	-122.3***	-48.72***
	(-6.97)	(-6.08)	(-5.26)
Observations	593	593	593
aic	.	.	.
r2_w	0.379	0.349	0.309
r2_b	0.434	0.417	0.407
r2_o	0.449	0.427	0.392

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 23: Random Effect Panel regression on the stock of Scientists with raw population counts as dependent variable

B.1.1 Stock of Scientists

In addition to the Random effects Panel regression on the elasticity of the population (i.e. log of hESC population and log of NoSC population) we carry out also Fixed Effect versions of the regression and keeping the raw population counts

B.2 Move Abroad and California

	(1) log(Pop. SC)	(2) log(Pop. ESC)	(3) log(Pop. hESC)
log(Pop Tot.)	1.157***	0.913***	0.753***
	(32.10)	(22.03)	(15.91)
Supportive	0.288***	0.357***	0.585***
	(5.04)	(5.00)	(6.14)
Constant	-4.048***	-3.467***	-3.238***
	(-16.13)	(-12.15)	(-10.06)
Observations	593	593	593
R^2_{within}	0.656	0.449	0.333
$R^2_{between}$	0.865	0.786	0.686
$R^2_{overall}$	0.861	0.781	0.658

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 24: Random Effect Panel regression on the stock of Scientists with log population counts as dependent variable

	(1) log(Pop. SC)	(2) log(Pop. ESC)	(3) log(Pop. hESC)
log(Pop Tot.)	1.417***	1.154***	1.171***
	(29.60)	(18.74)	(14.24)
Supportive	0.211***	0.280***	0.454***
	(3.79)	(3.91)	(4.74)
Constant	-5.804***	-5.087***	-6.010***
	(-18.61)	(-12.68)	(-11.21)
Observations	593	593	593
aic	28.97	327.3	671.1
r2_w	0.659	0.453	0.348
r2_b	0.864	0.785	0.681
r2_o	0.861	0.779	0.653

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 25: Fixed Effect Panel regression on the stock of Scientists with raw population counts as dependent variable

	Moved Abroad
SC × 1999	0.360***
SC × 2000	0.00395
SC × 2001	0.130
SC × 2002	0.171**
SC × 2003	-0.00537
SC × 2004	0.0411
SC × 2005	0.0961*
SC × 2006	0.0277
SC × 2007	0.0233
SC × 2008	0.0647
ESC × 1999	0.393**
ESC × 2000	-0.0333
ESC × 2001	0.223
ESC × 2002	0.198
ESC × 2003	0.268
ESC × 2004	0.217*
ESC × 2005	0.180
ESC × 2006	0.139
ESC × 2007	0.316***
ESC × 2008	0.135
hESC × 1999	-0.117
hESC × 2000	0.0746
hESC × 2001	-0.155
hESC × 2002	-0.0540
hESC × 2003	0.349*
hESC × 2004	0.0539
hESC × 2005	-0.173
hESC × 2006	-0.255**
hESC × 2007	-0.0561
hESC × 2008	0.000477
supportive	-0.0138
Productivity	-0.0115
Constant	-1.611***
Year Effects	Yes
Age Effects	Yes
State Effects	Yes
Observations	287918
ll	-158981.5
r2_p	0.00772

errors clustered at State-Year level (576 clusters)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 26: Regression Result for the estimation of the yearly propensity to move abroad. The individuals

	Moved Abroad
moved_to_california	
SC × 1999	-0.189
SC × 2000	0.127
SC × 2001	-0.0554
SC × 2002	0.149
SC × 2003	0.0416
SC × 2004	0.0733
SC × 2005	-0.249
SC × 2006	-0.103
SC × 2007	0.165*
SC × 2008	0.301**
ESC × 1999	0.715***
ESC × 2000	0.561**
ESC × 2001	0.0677
ESC × 2002	-0.0139
ESC × 2003	0.105
ESC × 2004	0.155
ESC × 2005	0.104
ESC × 2006	-0.508
ESC × 2007	-0.172
ESC × 2008	0.386**
hESC × 1999	-0.338
hESC × 2000	0.535*
hESC × 2001	0.575
hESC × 2002	-0.543*
hESC × 2003	-0.335
hESC × 2004	-0.357
hESC × 2005	0.380*
hESC × 2006	0.614***
hESC × 2007	0.270*
hESC × 2008	0.0788
supportive	0.0127
Productivity	0.0812***
Constant	-3.746***
Age Effects	Yes
State Effects	Yes
Year Effects	Yes
Observations	242,845
ll	-59539.5
r2_p	0.00944

errors clustered at State-Year level (576 clusters)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 27: Regression Result for the estimation of the yearly propensity to move to California only for researchers which were not in California alaready. The individuals

Bibliography

Acosta, Nefi D. and Sidney H. Golub (Sept. 2016). "The New Federalism: State Policies Regarding Embryonic Stem Cell Research". In: *The Journal of Law, Medicine & Ethics* 44.3, pp. 419–436. ISSN: 1073-1105. DOI: 10.1177/1073110516667939. arXiv: 15334406. URL: http://journals.sagepub.com/doi/10.1177/1073110516667939.

Adelson, Joel W. and Joanna K. Weinberg (2010). "The California stem cell initiative: Persuasion, politics, and public science". In: *American Journal of Public Health* 100.3, pp. 446–451. ISSN: 00900036. DOI: 10.2105/AJPH.2009.168120.

Agrawal, Ajay, Iain Cockburn, and John McHale (2006). "Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships". In: *Journal of Economic Geography* 6.5, pp. 571–591. ISSN: 14682702. DOI: 10.1093/jeg/lbl016. arXiv: arXiv:1011.1669v3.

Agrawal, Ajay, Devesh Kapur, et al. (2011). "Brain drain or brain bank? The impact of skilled emigration on poor-country innovation". In: *Journal of Urban Economics* 69.1, pp. 43–55. ISSN: 00941190. DOI: 10.1016/j.jue.2010.06.003.

Alberta, Hillary B. et al. (2015). "Assessing state stem cell programs in the United States: How has state funding affected publication trends?" In: *Cell Stem Cell* 16.2, pp. 115–118. ISSN: 18759777. DOI: 10.1016/j.stem.2015.01.007.

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang (May 2010). "Superstar Extinction *". In: *Quarterly Journal of Economics* 125.2, pp. 549–589. ISSN: 0033-5533. DOI: 10.1162/qjec.2010.125.2.549. URL: https://academic.oup.com/qje/article-lookup/doi/10.1162/qjec.2010.125.2.549.

Baker, Laurence, Bruce Deal, and Managing Principal (2004). *Economic Impact Analysis Proposition 71 California Stem Cell Research and Cures*

Initiative. Tech. rep. URL: http://www.analysisgroup.com/uploadedfiles/content/news%7B%5C_%7Dand%7B%5C_%7Devents/news/proposition%7B%5C_%7D71%7B%5C_%7Dreport.pdf.

Barney, Jay (Mar. 1991). "Firm Resources and Sustained Competitive Advantage". In: *Journal of Management* 17.1, pp. 99–120. ISSN: 0149-2063. DOI: 10.1177/014920639101700108. arXiv: 0803973233. URL: http://journals.sagepub.com/doi/10.1177/014920639101700108.

Blondel, Vincent D et al. (Oct. 2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2008/10/P10008.

Bohannon, John and Kirk Doran (May 2017). "Introducing ORCID". In: *Science* 356.6339, pp. 691–692. ISSN: 0036-8075. DOI: 10.1126/science.356.6339.691. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.356.6339.691.

Cardwell, Donald Stephen Lowell (1972). *Turning points in western technology; a study of technology, science and history*. New York, NY Science History Pub.

Carriqury, Juan Martin (2017). "Impact of M&A on Employee Turnover". In: pp. 1–26.

Catini, Roberto et al. (2015). "Identifying geographic clusters: A network analytic approach". In: *Research Policy* 44.9, pp. 1749–1762. ISSN: 00487333. DOI: 10.1016/j.respol.2015.01.011.

Caulfield, Timothy et al. (June 2009). "The Stem Cell Research Environment: A Patchwork of Patchworks". In: *Stem Cell Reviews and Reports* 5.2, pp. 82–88. ISSN: 1550-8943. DOI: 10.1007/s12015-009-9071-3. URL: https://link.springer.com/content/pdf/10.1007%7B%5C%%7D2Fs12015-009-9071-3.pdf%20http://link.springer.com/10.1007/s12015-009-9071-3.

Chambers, E et al. (1998). "The war for talent". In: *The McKinsey Quarterly* 3, pp. 44–57. ISSN: 10367128. DOI: 10.1080/03071840308446873.

Chessa, Alessandro et al. (Feb. 2013). "Is Europe Evolving Toward an Integrated Research Area?" In: *Science* 339.6120, pp. 650–651. ISSN: 0036-8075. DOI: 10.1126/science.1227970. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1227970.

Coff, Russell W. (Apr. 1997). "Human Assets and Management Dilemmas: Coping With Hazards On the Road to Resource-Based Theory". In: *Academy of Management Review* 22.2, pp. 374–402. ISSN: 0363-7425.

DOI: 10.5465/amr.1997.9707154063. URL: http://journals.aom.org/doi/10.5465/amr.1997.9707154063.

Culotta, Elizabeth (May 2017). "People on the move: The science of migrations". In: *Science*. ISSN: 0036-8075. DOI: 10.1126/science.aan6884. URL: http://www.sciencemag.org/news/2017/05/people-move-science-migrations.

Danzon, Patricia M., Andrew Epstein, and Sean Nicholson (June 2007). "Mergers and acquisitions in the pharmaceutical and biotech industries". In: *Managerial and Decision Economics* 28.4-5, pp. 307–328. ISSN: 01436570. DOI: 10.1002/mde.1343. URL: http://doi.wiley.com/10.1002/mde.1343.

Department of Health and Human Services (June 2015). *Stem Cell Basics*. URL: https://stemcells.nih.gov/info/basics.htm (visited on 10/15/2018).

Deville, Pierre et al. (2014). "Career on the move: Geography, stratification, and scientific impact". In: *Scientific reports* 4, p. 4770.

Dias, M. B. et al. (2010). "Effects of the USA PATRIOT Act and the 2002 Bioterrorism Preparedness Act on select agent research in the United States". In: *Proceedings of the National Academy of Sciences* 107.21, pp. 9556–9561. ISSN: 0027-8424. DOI: 10.1073/pnas.0915002107. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0915002107.

Enrico Moretti, By et al. (2017). "The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists †". In: *American Economic Review* 107.7, pp. 1858–1903. ISSN: 0002-8282. DOI: 10.1257/aer.20150508. URL: https://doi.org/10.1257/aer.20150508.

Ernst, Holger and Jan Vitt (2000). "The influence of corporate acquisitions on the behaviour of key inventors". In: *R and D Management* 30.2, pp. 105–120. ISSN: 0033-6807. DOI: 10.1111/1467-9310.00162. URL: http://doi.wiley.com/10.1111/1467-9310.00162.

Fortunato, Santo and M. Barthelemy (Jan. 2007). "Resolution limit in community detection". In: *Proceedings of the National Academy of Sciences* 104.1, pp. 36–41. ISSN: 0027-8424. DOI: 10.1073/pnas.0605965104. URL: www.pnas.orgcgidoi%20http://www.pnas.org/cgi/doi/10.1073/pnas.0605965104.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan (2012). "Foreign-born scientists: mobility patterns for 16 countries". In: *Nature Biotechnology* 30.12, p. 1250.

— (2014). "The mover's advantage: The superior performance of migrant scientists". In: *Economics Letters* 122.1, pp. 89–93.

— (2018). "Context Factors and the Performance of Mobile Individuals in Research Teams". In: *Journal of Management Studies* 55.1, pp. 27–59. ISSN: 14676486. DOI: 10.1111/joms.12279.

Geuna, Aldo (2015). *Global mobility of research scientists: The economics of who goes where and why*. Elsevier, Academic Press.

Giatsidis, Christos, Dimitrios M. Thilikos, and Michalis Vazirgiannis (May 2013). "D-cores: measuring collaboration of directed graphs based on degeneracy". In: *Knowledge and Information Systems* 35.2, pp. 311–343. ISSN: 0219-1377. DOI: 10.1007/s10115-012-0539-0.

Girvan, M and M. E. J. Newman (June 2002). "Community structure in social and biological networks." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.12, pp. 7821–6. ISSN: 0027-8424. DOI: 10.1073/pnas.122653799.

Gottweis, Herbert and Barbara Prainsack (Nov. 2006). "Emotion in political discourse: contrasting approaches to stem cell governance in the USA, UK, Israel and Germany". In: *Regenerative Medicine* 1.6, pp. 823–829. ISSN: 1746-0751. DOI: 10.2217/17460751.1.6.823. URL: https://www.futuremedicine.com/doi/10.2217/17460751.1.6.823.

Graf, Holger and Martin Kalthaus (2018). "International research networks: Determinants of country embeddedness". In: *Research Policy* 47.7, pp. 1198–1214. ISSN: 00487333. DOI: 10.1016/j.respol.2018.04.001.

Grant, Robert M (Dec. 1996). "Toward a knowledge-based theory of the firm". In: *Strategic Management Journal* 17.S2, pp. 109–122. ISSN: 01432095. DOI: 10.1002/smj.4250171110. arXiv: /ehis.ebscohost.com/ [http:]. URL: http://doi.wiley.com/10.1002/smj.4250171110.

Griffeth, Rodger W, Peter W Hom, and Stefan Gaertner (June 2000). "A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millennium". In: *Journal of Management* 26.3, pp. 463–488. ISSN: 0149-2063. DOI: 10.1177/014920630002600305. URL: https://journals.sagepub.com/doi/pdf/10.1177/014920630002600305.

Haleblian, Jerayr et al. (2009). "Taking stock of what we know about mergers and acquisitions: A review and research agenda". In: *Journal of Management* 35.3, pp. 469–502. ISSN: 01492063. DOI: 10.1177/0149206308330554.

Hambrick, Donald C and Albert A. Cannella (Aug. 1993). "Relative Standing: A Framework for Understanding Departures of Acquired Exec-

utives". In: *Academy of Management Journal* 36.4, pp. 733–762. ISSN: 0001-4273. DOI: 10.5465/256757. URL: http://journals.aom.org/doi/10.5465/256757.

Haveman, Heather A. (Dec. 1995). "The Demographic Metabolism of Organizations: Industry Dynamics, Turnover, and Tenure Distributions". In: *Administrative Science Quarterly* 40.4, p. 586. ISSN: 00018392. DOI: 10.2307/2393755. URL: https://www.jstor.org/stable/2393755?origin=crossref.

Higgins, Matthew J and Daniel Rodriguez (May 2006). "The outsourcing of R&D through acquisitions in the pharmaceutical industry". In: *Journal of Financial Economics* 80.2, pp. 351–383. ISSN: 0304405X. DOI: 10.1016/j.jfineco.2005.04.004. URL: http://linkinghub.elsevier.com/retrieve/pii/S0304405X05001807.

Hobman, Elizabeth et al. (2004). "Uncertainty During Organizational Change : Types , Consequences ". In: *Journal of Business and Psychology*, 18.4, pp. 507–532. ISSN: 08893268. DOI: 10.1023/B:JOBU.0000028449.99127.f7.

Hoekman, Jarno, Koen Frenken, and Robert J.W. Tijssen (June 2010). "Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe". In: *Research Policy* 39.5, pp. 662–673. ISSN: 0048-7333. DOI: 10.1016/J.RESPOL.2010.01.012.

Holtom, Brooks C. et al. (2005). "Shocks as causes of turnover: What they are and how organizations can manage them". In: *Human Resource Management* 44.3, pp. 337–352. ISSN: 00904848. DOI: 10.1002/hrm.20074.

Jackson, Matthew O and Brian W Rogers (May 2007). "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" In: *American Economic Review* 97.3, pp. 890–915. ISSN: 0002-8282. DOI: 10.1257/aer.97.3.890.

Kearl, Megan (2010). "Dickey-Wicker Amendment, 1996". In: *Embryo Project Encyclopedia*. URL: https://embryo.asu.edu/pages/dickey-wicker-amendment-1996.

King, David R. et al. (Feb. 2004). "Meta-analyses of post-acquisition performance: indications of unidentified moderators". In: *Strategic Management Journal* 25.2, pp. 187–200. ISSN: 0143-2095. DOI: 10.1002/smj.371. URL: http://doi.wiley.com/10.1002/smj.371.

Klitzman, Robert and Mark V. Sauer (2009). "Payment of egg donors in stem cell research in the USA". In: *Reproductive BioMedicine Online* 18.5, pp. 603–608. ISSN: 14726483. DOI: 10.1016/S1472-6483(10)

60002-8. URL: http://dx.doi.org/10.1016/S1472-6483(10)60002-8.

Levine, Aaron D (June 2008). "Identifying Under- and Overperforming Countries in Research Related to Human Embryonic Stem Cells". In: *Cell Stem Cell* 2.6, pp. 521–524. ISSN: 19345909. DOI: 10.1016/j.stem.2008.05.008. URL: http://www.%20http://linkinghub.elsevier.com/retrieve/pii/S1934590908002221.

Levine, Aaron D. (2012). "State stem cell policy and the geographic preferences of scientists in a contentious emerging field". In: *Science and Public Policy* 39.4, pp. 530–541. ISSN: 03023427. DOI: 10.1093/scipol/scs038.

Levine, Aaron D., T. Austin Lacy, and James C. Hearn (2013). "The origins of human embryonic stem cell research policies in the us states". In: *Science and Public Policy* 40.4, pp. 544–558. ISSN: 03023427. DOI: 10.1093/scipol/sct005.

Mokyr, Joel (2016). *A culture of growth: the origins of the modern economy*. Princeton University Press.

Morrison, Greg, Massimo Riccaboni, and Fabio Pammolli (May 2017). "Disambiguation of patent inventors and assignees using high-resolution geolocation data". In: *Scientific Data* 4, p. 170064. ISSN: 2052-4463. DOI: 10.1038/sdata.2017.64. URL: https://www.nature.com/articles/sdata201764.pdf%20http://www.nature.com/articles/sdata201764.

National Conference of State Legislatures (2016). *Embryonic and fetal research laws*. Accessed: 2018-09-16.

Nature Cell Biology (July 2010). "Human embryonic stem cell research in the US: time for change?" In: *Nature Cell Biology* 12.7, pp. 627–627. ISSN: 1465-7392. DOI: 10.1038/ncb0710-627. URL: https://www.nature.com/articles/ncb0710-627.pdf.

Newman, M. E. J. (Jan. 2001). "The structure of scientific collaboration networks". In: *Proceedings of the National Academy of Sciences* 98.2, pp. 404–409. ISSN: 0027-8424. DOI: 10.1073/pnas.98.2.404.

Newman, M. E. J. and M. Girvan (Feb. 2004). "Finding and evaluating community structure in networks". In: *Physical Review E* 69.2, p. 026113. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.69.026113.

OECD (2017). *OECD Science, Technology and Industry Scoreboard 2017*. OECD Publishing, Paris.

Owen-Smith, Jason and Jennifer McCormick (Apr. 2006). "An international gap in human ES cell research". In: *Nature Biotechnology* 24.4,

pp. 391–392. ISSN: 1087-0156. DOI: 10.1038/nbt0406-391. URL: http://www.nature.com/articles/nbt0406-391.

Ozden, Caglar and Hillel Rapoport (2018). "Cross-country Perspectives on Migration and Development: Introduction". In: *The Economic Journal*.

Page, Lawrence et al. (1998). "The PageRank Citation Ranking: Bringing Order to the Web". In: *World Wide Web Internet And Web Information Systems* 54.1999-66, pp. 1–17. ISSN: 1752-0509. DOI: 10.1.1.31.1768.

Pharmaceutical Researchers and Manufacturers of America (2012). *Pharmaceutical Researchers and Manufacturers of America Report 2012*. Tech. rep., pp. 1–73. URL: http://phrma-docs.phrma.org/sites/default/files/pdf/phrma%7B%5C_%7Dindustry%7B%5C_%7Dprofile.pdf.

Ranft, Annette L and Michael D Lord (Nov. 2000). "Acquiring new knowledge". In: *The Journal of High Technology Management Research* 11.2, pp. 295–319. ISSN: 10478310. DOI: 10.1016/S1047-8310(00)00034-1. URL: http://linkinghub.elsevier.com/retrieve/pii/S1047831000000341.

Ravenscraft, David J and William F Long (2000). "Paths to creating value in pharmaceutical mergers". In: *Mergers and productivity*. University of Chicago Press, pp. 287–326.

Russo, Eugene (2005). "Follow the money - The politics of embryonic stem cell research". In: *PLoS Biology* 3.7, pp. 1167–1171. ISSN: 15449173. DOI: 10.1371/journal.pbio.0030234.

Salter, Brian and Olivia Harvey (July 2008). "Stem cell innovation in the USA: the benefits of the minimal state". In: *Regenerative Medicine* 3.4, pp. 597–610. ISSN: 1746-0751. DOI: 10.2217/17460751.3.4.597. URL: https://www.futuremedicine.com/doi/10.2217/17460751.3.4.597.

Sassen, Saskia (2016). "The Global City: Strategic Site, New Frontier". In: *Managing Urban Futures*. Routledge, pp. 89–104.

Saxenian, AnnaLee (June 2005). "From Brain Drain to Brain Circulation: Transnational Communities and Regional Upgrading in India and China". In: *Studies in Comparative International Development* 40.2, pp. 35–61. ISSN: 0039-3606. DOI: 10.1007/BF02686293.

Schiller, Daniel and Javier Revilla Diez (2012). "The Impact of Academic Mobility on the Creation of Localized Intangible Assets". In: *Regional Studies*. ISSN: 0034-3404. DOI: 10.1080/00343404.2011.571241.

Scott, Christopher Thomas et al. (June 2011). "Democracy Derived? New Trajectories in Pluripotent Stem Cell Research". In: *Cell* 145.6, pp. 820–826. ISSN: 00928674. DOI: 10.1016/j.cell.2011.05.032. URL: https://ac.els-cdn.com/S0092867411005939/1-s2.0-S0092867411005939-main.pdf.

Serafinelli, Michel and Guido Tabellini (2017). "Creativity over Time and Space". In:

Solimano, Andrés (2008). *The international mobility of talent: Types, causes, and development impact*. Vol. 394. Oxford University Press Oxford.

StemGen (2017). *International Database on the Legal and Socio-Ethical Issues in Stem Cell Research*. Accessed: 2018-10-09 through web.archive.org.

Takahashi, Kazutoshi and Shinya Yamanaka (2006). "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors". In: *Cell* 126.4, pp. 663–676. ISSN: 00928674. DOI: 10.1016/j.cell.2006.07.024. arXiv: NIHMS150003.

Taylor, Peter J and Ben Derudder (2015). *World city network: a global urban analysis*. Routledge.

Thomson, J. A. (1998). "Embryonic Stem Cell Lines Derived from Human Blastocysts". In: *Science*. ISSN: 0036-8075. DOI: 10.1126/science.282.5391.1145.

Torvik, Vetle I. (Nov. 2015). "MapAffil: A Bibliographic Tool for Mapping Author Affiliation Strings to Cities and Their Geocodes Worldwide". In: *D-Lib Magazine* 21.11/12. ISSN: 1082-9873. DOI: 10.1045/november2015-torvik.

Torvik, Vetle I. and Neil R. Smalheiser (July 2009). "Author name disambiguation in MEDLINE". In: *ACM Transactions on Knowledge Discovery from Data* 3.3, pp. 1–29. ISSN: 15564681. DOI: 10.1145/1552303.1552304. URL: http://portal.acm.org/citation.cfm?doid=1552303.1552304.

Traag, Vincent (Nov. 2017). *vtraag/louvain-igraph: 0.6.1*. DOI: 10.5281/zenodo.1054103.

Trippl, Michaela (2013). "Scientific Mobility and Knowledge Transfer at the Interregional and Intraregional Level". In: *Regional Studies*. ISSN: 00343404. DOI: 10.1080/00343404.2010.549119.

Wagner, Caroline S. and Loet Leydesdorff (Dec. 2005). "Network structure, self-organization, and the growth of international collaboration in science". In: *Research Policy* 34.10, pp. 1608–1618. ISSN: 0048-7333. DOI: 10.1016/J.RESPOL.2005.08.002.

Wilmut, I. et al. (1997). "Viable offspring derived from fetal and adult mammalian cells". In: *Nature*. ISSN: 00280836. DOI: 10.1038/385810a0.
Zucker, Lynne G and Michael R Darby (2007). *Star scientists, innovation and regional and national immigration*. Tech. rep. National Bureau of Economic Research.