The Commercialization Pipeline

Christopher Esposito (UCLA)

Olav Sorenson (UCLA)

Abstract

Despite widespread belief that scientific research fuels innovation and economic growth, the mechanisms that accelerate the translation of scientific discoveries into commercial applications remain poorly understood, particularly those not immediately connected to patenting. We develop an empirical framework for studying science-to-technology commercialization as a multi-step process by tracing knowledge trajectories along the shortest citation paths connecting scientific articles to downstream patents. Using a matched differences-in-differences design, we examine the relationship between two factors—the science being pursued in a regional hub or in a corporate research setting—and the speed with which these trajectories move toward commercial applications. Hubs and corporate settings both accelerate progress toward commercialization. Our framework offers a scalable approach to studying commercialization dynamics far upstream of patents and provides evidence that organizational and geographic environments shape the pace at which science becomes economically useful.

** The authors acknowledge research support from the UCLA Price Center for Entrepreneurship and Innovation.*

1. Introduction

The private and public sectors spend billions of dollars on basic science. For example, in the United States, the National Science Foundation (NSF) funds roughly $10 billion in projects each year. The Department of Defense and the National Institutes of Health (NIH) combine for another $50 billion in annual spending. Private companies, from pharmaceutical manufacturers to quantum computing startups, also allocate tens of billions of dollars to scientific research (Arora et al., 2024). These investments pay off in the form of novel products and services that improve lives and that accelerate economic growth (Bush, 1945; W. Cohen et al., 2002; Jaffe, 1989; Mowery et al., 2001). If anything, science has been becoming increasingly important to innovation and economic growth (Fleming et al., 2019).

Although research strongly suggests that investments in science eventually yield economically valuable innovations, our ability to study these processes has been limited. Commercialization often requires a long pipeline from lab the product. Research to date has generally focused either on the correlations between scientific activity and invention, usually measured by patents or economic growth (Adams, 1990; Jaffe, 1989; Valero & Van Reenen, 2019), or on the final step in the process, the research that immediately proceeded a patent (Arora et al., 2024; Fleming & Sorenson, 2004; Marx & Fuegi, 2020; Siegel et al., 2003). But the transition from basic science to the commercialization of novel products and services can require years, if not decades, and often involves multiple steps to move from principles to applications (Ahmadpoor & Jones, 2017; Van de Ven et al., 2008; Werth, 1994). Existing approaches have had little insight into how anything up until that final step influences the pace of commercialization.

We propose an empirical framework for studying the factors that accelerate the advancement of science toward commercial applications along the full pipeline connecting basic science to applied science and on to commercialization. We first trace trajectories of scientific research as they progress from basic, fundamental discoveries to more applied scientific research and eventually to downstream commercial applications. We then use those trajectories to identify factors and mechanisms that appear to accelerate their progression.

As an example, we use this framework to explore two factors that have been thought to accelerate the progression of science toward application: corporate engagement in science and commercialization hubs. Companies pursue scientific research to gain earlier access to academic research (W. M. Cohen & Levinthal, 1990; Stern, 2004). But they then probably steer their own research toward more commercial applications (Arora et al., 2024; Rosenberg, 1990; Scharfmann et al., 2025). Geographic hubs, meanwhile, have been seen as important because they provide entire ecosystems to support the commercialization process, from early-stage funding to the mentoring of scientists (Clayton et al., 2018; Samila & Sorenson, 2010; Stuart & Ding, 2006).

To study how corporate engagement and embedding in hubs influence the translation of basic and applied science toward commercial applications, we trace knowledge trajectories connecting scientific articles to downstream patents through multi-step citation paths. Each citation step brings the trajectory closer to being cited by a patent. We trace these trajectories along the shortest citation path connecting a paper to a downstream patent, focusing our analysis on the most direct route from scientific research to commercial uses (Ahmadpoor & Jones, 2017).

We characterize each paper according to its commercialization potential (CP), using the measures developed by Masclans, Hasan, and Cohen (2025). CP assesses the commercial potential of a scientific paper by comparing the textual similarity of its abstract to the abstracts of papers cited by patents. As the advances reported in papers come closer to being commercializable, CP rises. We therefore assess changes in the pace at which knowledge trajectories move toward becoming products through increases in this CP measure.

Our analysis follows a differences-in-differences logic. We identify trajectories that transition to hubs and to corporate settings and match them to "placebo" trajectories that did not undergo such transitions. We conduct these matches based on trajectories' pre-transition CP levels and trends, as well as their publication year and broad scientific field. We find that both the treated trajectories experience a post-treatment acceleration in their increases in commercialization potential, relative to the placebo trajectories that remain in academic settings and outside of hubs.

We offer two primary contributions. First and foremost, we propose a method for constructing and analyzing how the structure of the commercialization pipeline influences the speed at which it translates science into technology. Second, we use the framework to examine how hubs and corporate science influence commercialization. Both appear to accelerate its movement toward application meaningfully.

We begin by introducing the datasets and describing the methods used to identify the shortest-citation-path knowledge trajectories connecting articles to downstream patents. We then detail the procedure used to match the transitioning trajectories to the counterfactual placebo trajectories.

We then present two sets of results: the effect of transitions to hubs and corporations on CP, and the effect of such transitions on patenting hazards. We conclude with a discussion of the findings and possible areas for future research.

2. Data

Our data come from OpenAlex, the largest and most comprehensive repository of publication metadata. It covers more than 240 million publications from around the world. For our purposes, OpenAlex importantly includes information on publication year, research fields, the geographic location of authors, and paper-to-paper citations.

Our key dependent variable, the commercialization potential of the science reported in articles, draws on a dataset described in Masclans, Hasan, and Cohen (2025). They created their CP measure by training a machine learning model to predict the probability that an article would be cited by a patent based on the text of its abstract. Because OpenAlex only contains abstracts for a subset of its repository, using this measure requires us to limit our analysis to the 31 million papers with abstracts. In general, these papers have been published in higher-impact and more international journals than the full population of OpenAlex articles.

To build knowledge trajectories, we trace citations made by patents to scientific articles back multiple steps into the scientific literature. Beginning with patents means that we only include science that has already become commercialized. That selection means that we probably miss trajectories with the longest paths as they end up being right censored in our data, not yet reaching

a patent. Our approach therefore allows us to analyze what influences the speed to commercialization for science that becomes commercialized.

We build the trajectories backwards. We first identify scientific articles directly cited by USPTO patents using the "Reliance on Science" dataset (Marx & Fuegi, 2020). We label these papers as "distance=0 papers" because they sit right on the boundary between commercialization and science. If a paper receives citations from multiple patents, we randomly select and retain one of these citation edges.

We then identify the papers at $distance = 1$ from the nearest patent. Distance refers to the length of the citation path from a focal paper to a paper cited by a patent, so these papers all receive citations from $distance = 0$ papers. We repeat this process until we arrive at the $distance = 10$ papers, papers ten steps away from the nearest patent along the patent-paper citation graph.

Papers could fall on more than one path, potentially even being at different distances from being cited by a patent along the different trajectories. Following Ahmadpoor and Jones (2017), we restrict the citation graph to the edges that fall along the shortest paths, the geodesics, connecting papers to patents. For example, if a paper is cited by both a distance=3 and a distance=6 paper, we would retain only the citation from the distance=3 paper. This restriction provides both conceptual and computational tractability. As the number of edges in a graph increases, the number of possible paths grows exponentially, to the point that computational analysis becomes infeasible.

This restriction to the shortest-path network also influences the interpretation of our results: Because we analyze how hub and corporate environments increase commercialization potential along the shortest path trajectories, our estimates tell us how exposure to these environments make the shortest paths from papers to patents even shorter.

We connect these data to two sources to determine exposure to hub and corporate environments. We use the Duke DISCERN 2.0 dataset (Arora et al., 2023) to assign authors to corporate affiliations. We then geocode corporate and university affiliations using the full text address strings, often recorded at the lab or plant level, to identify the locations in which the science had likely been developed. In total, our dataset includes roughly 9.9 million scientific papers that fall on these shortest paths.

3. Methods

We adopt a differences-in-differences logic for our analysis, comparing trajectories that transition into hub locations or corporate environments to a set of otherwise-similar trajectories that do not undergo such transitions. To minimize the extent to which selection into these transitions might influence our estimates, we match the transitioning trajectories to the counterfactual placebo trajectories on multiple dimensions.

We match trajectories at both the paper and the trajectory level. We begin by identifying papers in which a trajectory first transitions to a hub location or a corporate environment. We identify corporate environments as those where the last author reports a corporate affiliation. In most fields, the principal investigator (PI), who oversees the lab or group producing the research, customarily

appears last in the list of authors. We then define hub-produced papers as those for which the last author's affiliation has an address in one of the following U.S. cities: San Francisco, San Jose, Los Angeles, San Diego, New York City, Boston, Philadelphia (which includes Trenton/Princeton, NJ), Raleigh-Durham NC, Chicago, Houston TX, and Washington DC.

Let us label these transitioning patents as step $s = 0$ in trajectory *i. W*e use the shortest-path citation graph to identify the immediate and two-step antecedent papers to these transitions (denoted by papers at the coordinates $[i, s = -1]$ and $[i, s = -2]$), as well as the papers that follow these transitions (denoted by $[i, s = 1]$ and $[i, s = 2]$).

We then match each transitioning paper to a paper in a trajectory that does not transition to a hub or a corporation in $s = 0$, or in $s = -1$ or $s = -2$. From this set of possible matches, we first restrict the available candidates to those with $s = -2$ papers published in the same year and in the same broad field (N=26) as the $s = -2$ paper for the transitioning paper, as well as in the same CP value decile. For each remaining candidate, we calculate the change in CP from $s = -2$ and $s = -1$ and choose the one closest to the CP change observed from $s = -2$ and $s = -1$ for the transitioning paper. In other words, we use nearest-neighbor matching on CP changes combined with coarsened-exact matching on field, year, and CP levels. Together, these restrictions should ensure that the transitioning trajectories and the placebos had been advancing toward commercialization at nearly the same rates prior to the transitions to hubs and commercial settings.

Because the matching procedure requires information on trajectories with at least three observable steps ($s = -2$ to $s = 0$), our 9.9 million papers yield only 559,519 trajectories. This restriction tends to drop shorter trajectories (i.e. more applied areas of science).

Of the 559,519 trajectories with observable pre-period data, 39,498 transition to hubs and 9,706 transition to corporations in $s = 0$. Of those transitioning trajectories, we are able to match 39,414 (99.8%) of the hub-transitioning trajectories and 9,706 (100%) of the corporation-transitioning trajectories to placebo trajectories.

Matching generated two datasets, described in Table 1. The first consists of trajectories that transition to hubs in $s = 0$, matched to placebo trajectories that did not. The second comprises trajectories that transition to corporations in $s = 0$, matched to placebos that did not. Table 1 describes each dataset and the balance of the transitioning and placebo papers in each. The matching yielded almost identical distributions of change-in-CP from $s = -2$ to $s = -1$, with residual differences of only 0.01. The high precision of these matches stems in large part from the large number of candidate placebos available after the coarsened exact matching on field, publication year, and CP level. Trajectories in the hubs panel had an average of 655 candidates, while those in the corporations panel had an average of 597 potential matches.

< Table 1 about here>

Table 1 also reports the shortest-path citation distances to the nearest patent for transitioning and placebo trajectories. We would not expect balance on these dimensions. Most obviously, the

matching procedure did not include these measures as criteria. More importantly, citation distance to the nearest patent provides a forward-looking measure that may itself depend on the transition. For example, a trajectory that shifts into a corporate setting might realize a patent at an earlier step than its matched trajectory that did not (leading to a shorter path length to the nearest patent, even in steps -2 and -1).

Rather than providing a diagnostic on the success of the matching procedure, distance to the nearest patent points to the possible causal effects of transitioning to a hub or corporate setting in accelerating commercialization. With each step, the shortest path distance to a patent mechanically falls by one. Because trajectories move closer to patenting with each step, level differences across the transitioning and placebo groups suggests a causal effect of transitioning to hubs and corporations on path length. For example, transitions to hubs reduce the distance of otherwise-similar trajectories to the patent boundary by (4.94 - 4.63) = 0.31 citation steps.

The shortening of science-to-commercialization trajectories induced by the earlier patenting of transitioning trajectories also affects the composition of the sample. When a trajectory transitions to a patent, it ends. Naïve estimation that does not account for the changing sample could therefore understate the effect of transitioning on CP (if increases in CP correlate positively with patenting, as expected).

Table 2 reports the number of observations in transitioning and placebo trajectories relative to the matched transitioning papers. Both sets of trajectories have the same number of observations at steps -2, -1, and 0, as required by the matching procedure. Differences in the number of

observations emerge in steps 1 and 2, as some of the transitioning trajectories end in patents earlier than the placebos.

<Table 2 about here>

Finally, Table 3 reports the commercialization potential (CP) of paths by step. Prior to the transitions, transitioning and placebo trajectories have identical commercialization trajectories. Differences emerge in step 0, as trajectories that transition to hubs increase in CP by 0.05 more than their placebos and trajectories that transition to corporations by 0.1 more than their placebos. These differences attenuate in steps 1 and 2.

<Table 3 about here>

4. Effects of Treatments on the Commercialization Potential of Science

We next estimate the change in commercialization potential (CP) in a linear model:

$$CP_{i,s} = \sum_{s=-2}^{2} \beta_s^p StepsToTransition_i^p + \sum_{s=-2}^{2} \beta_s^t StepsToTransition_i^t + \alpha_i + \varepsilon_{i,s}$$

The model Includes trajectory fixed effects, α_i, and two separate event-time factor variables: the first, $StepsToTransition_i^p$, codes a vector of indicator variables for the step (s) relative to the transitioning paper for both transitioning and matched placebo trajectories; the second, $StepsToTransition_i^t$, encodes an equivalent vector for only transitioning trajectories. The

coefficients associated with the first vector captures the change in commercialization potential by step across all trajectories, relative to the reference step $s = -1$. The coefficients associated with the second, meanwhile, provides a flexible, non-parametric (in steps) estimate of the effects of transitioning to a hub or corporate setting on commercialization potential.

Table 4 Column 1 provides estimates for transitioning to a hub while Column 2 reports the estimates for transitioning to a corporate environment. As one would expect, commercialization potential rises with each step in both transitioning and placebo trajectories. Transitioning trajectories nevertheless experience even faster acceleration in this potential. Consider first the transition to hubs. While placebo trajectories increased in CP by 0.0547 between the reference step ($s = -1$) and the next step ($s = 0$), transitioning steps increased in CP by 0.0523 more. In other words, transitioning to a hub appeared to accelerate the movement toward commercialization by nearly 100%. This positive effect of transitioning to a hub persists for at least two additional steps after the transition.

<Table 4 about here>

The lagged coefficient in Table 4 Column 1, $\beta^{t}_{s=-2}$, also suggests that the transitioning sample contains minimal selection into hubs. The coefficient offers a test of the parallel pre-trend assumption, a necessary condition for being able to interpret the differences-in-differences estimate as being causal. Though the magnitude of the coefficient relative to the standard error suggests that transitioning trajectories have slightly lower CP leading up to the transition, the difference represents a mere 1.6% of the magnitude of the estimated *s*=0 transition effect.

Column 2 of Table 4 reports results for transitioning into corporations. The coefficients tell a similar story as the transitions to hubs: Although all trajectories increase in commercialization potential with each step, transitioning to a corporate setting accelerates that process. The magnitudes of the effect sizes for the corporate effect are roughly double those for hubs. The lagged coefficient for transitioning again suggests that selection into these transitions probably accounts for little of the estimated effect.

Figure 1 plots the coefficients associated with the $StepsToTransition_i^t$ variables against event time, with 95% confidence intervals. The figures show (1) the lack of substantive pre-trends, (2) the instantaneous impact of hubs and corporations on CP in $s = 0$, and (3) the persistence of the treatment effects in the subsequent steps, $s = 1$ and $s = 2$.

5. Effects of Transitions on Patenting Hazards

Although the commercialization potential measure gives us a continuous dependent variable that allows us to observe changes in the proximity to commercialization from one step to the next, we can also directly estimate the effect of transitioning to a hub or to a corporate environment on the observed rates of patenting for trajectories.

<Figure 1 about here>

Examining patenting rates directly can also help us to address a potential methodological issue with our CP regressions. When a trajectory culminates in a patent, its sequence ends and no longer contributes to our coefficient estimates. Trajectory attrition changes the composition of the data

set for $s = 1$ and $s = 2$. To the extent that transitions and patenting correlate positively with increases in CP, attrition could negatively bias our estimated effect of CP on later steps. In other words, it could explain the apparent attenuation of the transition effects in our prior analyses.

To estimate the effects of transitions to hubs and corporations on the rate of patenting, we estimate a discrete-time hazard model (as a pooled logit). The dependent variable is an indicator equal to one if a trajectory patents at a given step and zero otherwise. We restrict the analysis to $s = 1$ and $s = 2$ since our construction of the trajectories means that they could never result in a patent before that. Because the model conditions on survival up to each step, removing trajectories from the risk set once they patent, it accounts for the right censoring induced by patenting.

We again define a trajectory as transitioning to a hub or to a corporate setting based on whether these shifts occurred at the $s = 0$ paper. We include an indicator variable for each transition, an indicator variable for $s = 2$ (which one could see as a non-parametric estimate of the step "clock"), and the interaction between transitioning and $s = 2$. We cluster the standard errors at the matched-pair level.

Table 5 reports the results. Both types of transitions increase the patenting hazards post-transition. The negative interaction effect, however, suggests that this effect attenuates over time (and therefore also suggests that the attenuation in the CP models does not stem from attrition in the sample).

<Table 5 about here>

To facilitate interpretation of the hazard estimates, we convert the estimated step-specific patenting rates into cumulative patenting probabilities by step for transitioning and placebo (control) trajectories. Using the fitted model, we compute the probability that a trajectory patents by step one and by step two, accounting for the fact that patenting at earlier steps removes trajectories from the risk set at later steps. We plot these values in Figure 2.

The resulting cumulative patenting curves reveal that transitions to hubs and to corporate settings accelerate patenting. Treated trajectories exhibit higher cumulative patenting probabilities immediately after treatment. These differences then level out in the subsequent step.

<Figure 2 about here>

6. Changing Counterfactual Composition

Our differences-in-differences approach examines the effects of transitions, to a hub or to a corporate setting, at step $s = 0$. Our analysis therefore estimates the effect of a transition rather than the absolute effect of being in a hub or a corporation. The distinction matters for two reasons. First, our data construction eliminates all trajectories that remain in hubs or in corporations from the beginning of the trajectory until commercialization. Second and more subtlety, our placebo cases sometimes transition at a later point in time (i.e. $s = 1$ or $s = 2$). These transitions likely raise the CP and the patenting hazards of these trajectories, leading to measurement error in our "treatment" and a downward bias in our estimates of the effect of these transitions.

Figure 3 illustrates this issue. It plots the proportions in each environment, by step, for transitioning and placebo trajectories. A small percentage of trajectories begin at $s = -2$ in hubs and corporations, but by construction, most transitioning trajectories enter the relevant environment at step *s*=0, while the placebos do not. However, the figure reveals that some placebo trajectories transition into hubs or corporations in steps $s = 1$ and $s = 2$.

<Figure 3 about here>

This changing composition in the counterfactual set over time may partially explain the attenuation across steps of the estimated transition effects on CP and patenting hazards. As placebo trajectories enter hubs and corporations at later steps, their outcomes begin to converge toward those of the earlier transitioning trajectories, reducing the estimated gap in subsequent periods. Attenuation therefore may reflect the expanding exposure of the placebos to hubs and corporations rather than a decay in the underlying commercialization advantages of these settings.

6. Discussion

We offer an empirical framework for extending the study of commercialization back from the boundary between papers and patents deep into the scientific pipeline leading up to those innovations. We do so by tracing scientific trajectories through citations, finding the shortest paths connecting science to patents.

The primary contribution of our research is to introduce a rigorous, scalable, and plausibly causal framework to study the factors that accelerate the full science-technology commercialization

pipeline. Existing research focused on the subset of science that directly impacts technology (Arora et al., 2021; Bikard & Marx, 2020; Marx & Fuegi, 2020; Scharfmann et al., 2025), and thus does not address the many steps over the course of years that are taking by scientists to prepare scientific knowledge for commercialization. Our framework thus enables researchers to study the commercialization process upstream. Given the large volume of scientific effort that is expended to prepare knowledge for commercialization, this framework has the potential to substantially accelerate the commercialization process and improve the return on scientific investments.

As an example, we used our framework to estimate the importance of commercialization hubs and of corporate science on the speed of commercialization. Our results suggest that both have substantial effects, accelerating the speed with which ideas become commercial innovations.

Our framework can be readily deployed to study additional factors that may affect commercialization pipelines. For example, factors at the level of individuals, networks, institutions, and nations may all affect the speed of commercialization (Esposito, 2026; Koffi & Marx, 2026; Marx & Hsu, 2022). By combining shortest-path trajectories connecting science produced by such entities to downstream patents (Ahmadpoor & Jones, 2017), combining the papers in those trajectories with information on their commercialization potential (Masclans et al., 2025), and by matching the "treated" trajectories to placebo controls, our framework can be extended to study multitudes of factors that may affect the commercialization pipeline.

References

Adams, J. (1990). Fundamental Stocks of Knowledge and Productivity Growth. *Journal of Political Economy*, *98*(4).

Ahmadpoor, M., & Jones, B. (2017). The Dual Frontier: Patented Inventions and Prior Scientific Advance. *Science*, *357*(6351), 583–587.

Arora, A., Belenzon, S., Cioaca, L., Sheer, L., & Shvadron, D. (2023). DISCERN 2.0: Extending and Enhancing the DISCERN Dataset. *Working Paper*.

Arora, A., Belenzon, S., Kosenko, K., Suh, J., & Yafeh, Y. (2024). The Rise of Scientific Research in Corporate America. *Organization Science*, *36*(4), 1466–1488.

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

Bikard, M., & Marx, M. (2020). Bridging Academia and Industry: How Geographic Hubs Connect University Science and Corporate Technology. *Management Science*, *66*(8), 3295–3798.

Bush, V. (1945, July). As We May Think. *The Atlantic*.

Clayton, P., Feldman, M., & Lowe, N. (2018). Behind the scenes: Intermediary organizations that facilitate science commercialization through entrepreneurship. *Academy of Management Perspectives*, *32*(1), 104–124.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Cohen, W., Nelson, R., & Walsh, J. (2002). Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science*, *48*(1), 1–23.

Esposito, C. (2026). Global Science Sustains U.S. Innovation. *Working Paper*.

Fleming, L., Greene, H., Li, G., Marx, M., & Yao, D. (2019). Government-Funded Research Increasingly Fuels Innovation. *Scinece*, *364*(6446), 1139–1141.

Fleming, L., & Sorenson, O. (2004). Science as a map in technological search. *Strategic Management Journal*, *25*(8–9), 909–928. https://doi.org/10.1002/smj.384

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

Koffi, M., & Marx, M. (2026). Cossatts in the Attic: Is There a Gender Gap in the Commercialization of Science. *American Economic Journal: Applied Economics*, *Forthcoming*.

Marx, M., & Fuegi, A. (2020). Reliance on Science: Worldwide Front-Page Patent Citations to Scientific Articles. *2020*, *41*(9), 1572–1594.

Marx, M., & Hsu, D. (2022). Revisiting the Entrepreneurial Commercialization of Academic Science: Evidence from "Twin" Discoveries. *Management Science*, *68*(2), 1330–1352.

Masclans, R., Hasan, S., & Cohen, W. (2025). Measuring the commercial potential of science. *Strategic Management Journal*, *46*(9), 2199–2236.

Mowery, D., Nelson, R., Sampat, B., & Ziedonis, A. (2001). The growth of patenting and licensing by U.S. universities: An assessment of the effects of the Bayh–Dole act of 1980. *Research Policy*, *30*(1), 99–119.

Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy*, *19*(2), 165–164.

Samila, S., & Sorenson, O. (2010). Venture Capital as a Catalyst to Commercialization. *Research Policy*, *39*(10), 1348–1360.

Scharfmann, E., Marx, M., & Fleming, L. (2025). Pasteur's quadrant researchers bring novelty, impact to publishing, and patenting. *Science*, *390*(6776), 891–893.

Siegel, D., Waldman, D., Atwater, L., & Link, A. (2003). Commercial knowledge transfers from universities to firms: Improving the effectiveness of university–industry collaboration. *Journal of High Technology Management Review*, *14*(1), 111–133.

Stern, S. (2004). Do Scientists Pay to Be Scientists? *Management Science*, *50*(6), 709–853.

Stuart, T., & Ding, W. (2006). When Do Scientists Become Entrepreneurs? The Social Structural Antecedents of Commercial Activity in the Academic Life Sciences. *American Journal of Sociology*, *112*(1), 97–144.

Valero, A., & Van Reenen, J. (2019). The Economic Impact of Universities: Evidence from Across the Globe. *Economics of Education Review*, *68*, 53–67.

Van de Ven, A., Polley, D., Garud, R., & Venkataraman, S. (2008). *The Innovation Journey*. Oxford University Press.

Werth, B. (1994). *The Billion-Dollar Molecule*. Touchstone.

Tables

Table 1: Balance of Data Panels

	Treatment: Transition to Hubs		Treatment: Transition to Corporations	
	Treated	**Placebo**	**Treated**	**Placebo**
Number of trajectories	39,414	39,414	9,706	9,706
Mean publication year$_{s=-2}$	2007.81	2007.81	2007.46	2007.46
Modal publication field$_{s=-2}$	Engineering	Engineering	Engineering	Engineering
Mean change in CP between s=-2 and s=-1	0.06	0.06	0.08	0.07
Mean bin size used for matching	655.3	655.3	597.3	597.3
Residual change in CP after matching	0.01	0.01	0.01	0.01
Mean shortest path distance to patent$_{s=-2}$	4.63	4.94	3.93	4.51
Mean shortest path distance to patent$_{s=-1}$	3.64	3.95	2.93	3.51

Table 2: Observations by Step and Treatment Group

	Treatment: Transition to Hubs		Treatment: Transition to Corporations	
Step	**Treated**	**Placebo**	**Treated**	**Placebo**
-2	39,414	39,414	9,706	9,706
-1	39,414	39,414	9,706	9,706
0	39,414	39,414	9,706	9,706
1	28,901	31,031	5,670	6,962
2	20,261	22,655	3,466	4,732

Table 3: CP by Step and Treatment Group

	Treatment: Transition to Hubs		Treatment: Transition to Corporations	
Step	**Treated**	**Placebo**	**Treated**	**Placebo**
-2	0.39	0.39	0.48	0.48
-1	0.44	0.44	0.55	0.55
0	0.55	0.50	0.67	0.57
1	0.55	0.52	0.67	0.59
2	0.57	0.55	0.67	0.60

Table 4: Regression Estimates for Changes in Commercialization Potential

	(1) Effects of Transitions to Hubs	(2) Effects of Transitions to Corporations
$\beta^{p}_{s=-2}$	-0.0558*** (0.00136)	-0.0746 (0.00287)
$\beta^{p}_{s=0}$	0.0547 (0.00141)	0.0228*** (0.00285)
$\beta^{p}_{s=1}$	0.125*** (0.00177)	0.0810*** (0.00367)
$\beta^{p}_{s=2}$	0.175*** (0.00217)	0.118*** (0.00458)
$\beta^{t}_{s=-2}$	-0.000862*** (0.000141)	-0.000988*** (0.000326)
$\beta^{t}_{s=0}$	0.0523*** (0.00168)	0.0974*** (0.00329)
$\beta^{t}_{s=1}$	0.0347*** (0.00214)	0.0745*** (0.00456)
$\beta^{t}_{s=2}$	0.0200*** (0.00270)	0.0445*** (0.00597)
NOBS	33,332	79,066
Trajectories	78,828	19,412

Note: Trajectories are treated in event time = 0, the first step in the trajectory published in a hub or corporation. The control samples contains trajectories that do not transition to a hub or corporate environment in event time =0. All models contain trajectory fixed effects. Standard errors clustered at the matched pair.

Table 5: Effects of Transitions on Patenting Hazards

	(1) **Effect of Hubs**	(2) **Effect of Corporations**
Intercept	-1.59*** (0.0183)	-0.930*** (0.0225)
$\beta_{s=2}$	0.596*** (0.0219)	0.177*** (0.0342)
Treated	0.201*** (0.0234)	0.590*** (0.0271)
Treated * $\beta_{s=2}$	-0.0585** (0.0299)	-0.290*** (0.0472)
Observations	102,848	32,041

Figures

Figure 1: Event Plots

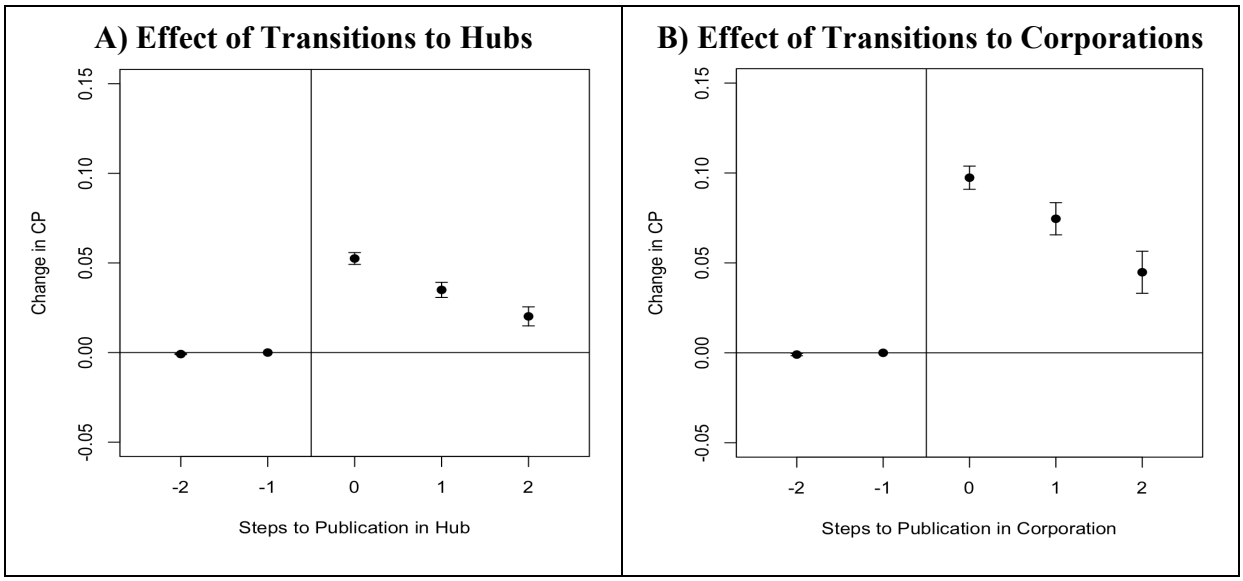


Figure 2: Cumulative Patenting Rates

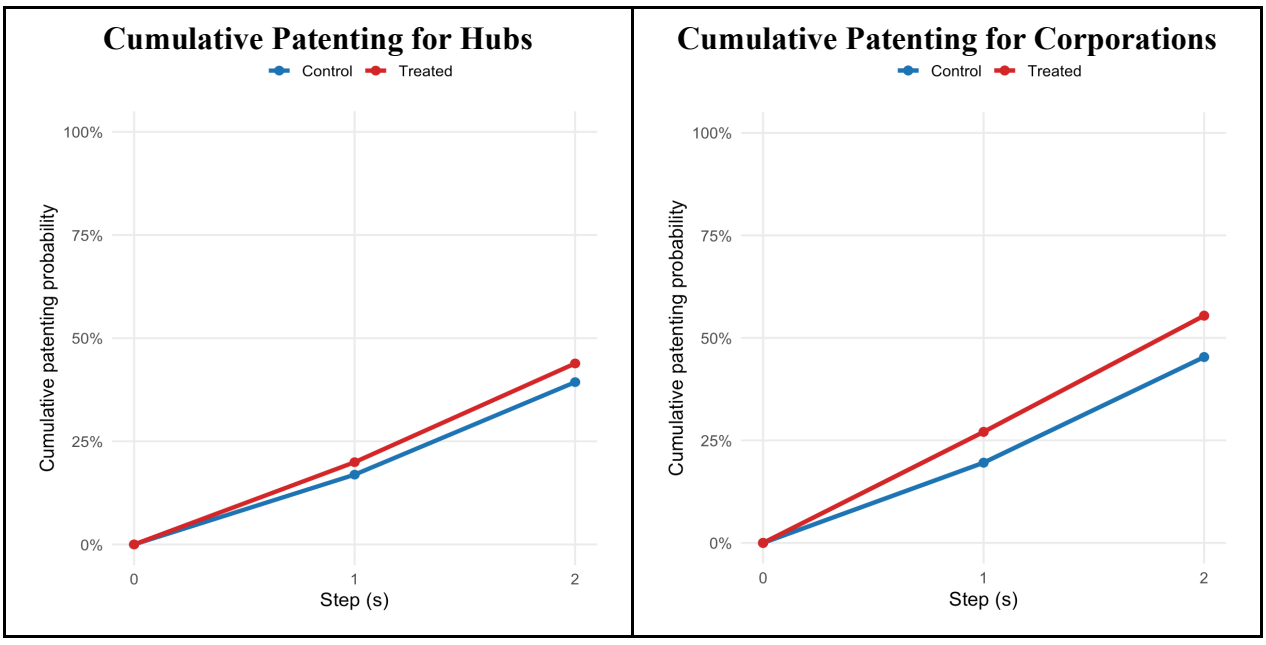
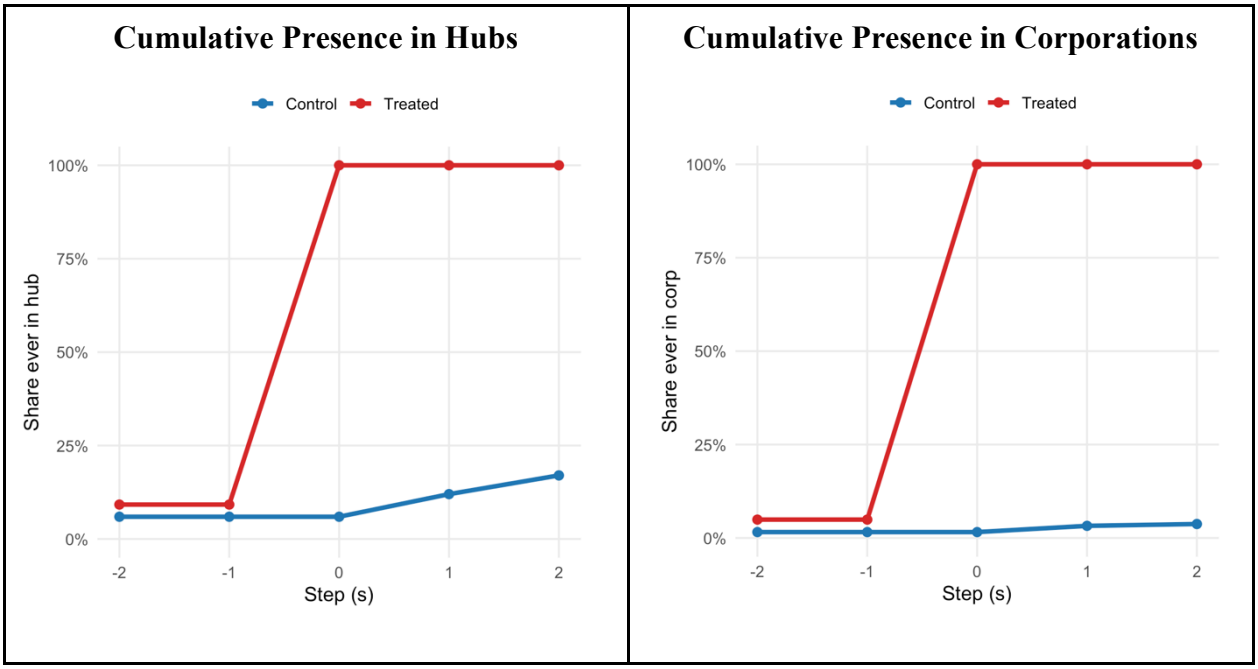


Figure 3: Presence in Hubs and Corporations for the Treated and Control Groups



Appendix: Representation of Scientific Fields in the Dataset

Table A: Trajectory Counts by Field

Field	Field name	Hubs		Corporations	
		Treated	Control	Treated	Control
11	Medicine	1,014	1,014	243	243
12	Biochemistry, Genetics & Molecular Biology	143	143	19	19
13	Biology	4,567	4,567	1,026	1,026
14	Pharmacology, Toxicology & Pharmaceutics	263	263	75	75
15	Neuroscience	133	133	44	44
16	Earth Sciences	995	995	323	323
17	Physics	1,920	1,920	1,155	1,155
18	Immunology & Microbiology	297	297	95	95
19	Agricultural & Biological Sciences	1,232	1,232	143	143
20	Health Professions	586	586	94	94
21	Dentistry	118	118	11	11
22	Materials Science	3,377	3,377	1,358	1,358
23	Chemistry	1,689	1,689	230	230
24	Psychology	797	797	207	207
25	Chemical Engineering	1,372	1,372	386	386
26	Economics, Econometrics & Finance	764	764	175	175
27	Engineering	12,229	12,229	2,870	2,870
28	Environmental Science	1,565	1,565	234	234
29	Energy	166	166	35	35
30	Decision Sciences	154	154	70	70
31	Computer Science	2,701	2,701	476	476
32	Mathematics	1,074	1,074	143	143
33	Social Sciences	1,211	1,211	130	130
34	Arts & Humanities	53	53	16	16
35	Business, Management & Accounting	139	139	33	33
36	Veterinary Science	855	855	115	115