

SUPERSTAR EXTINCTION

PIERRE AZOULAY
JOSHUA S. GRAFF ZIVIN
JIALAN WANG

We estimate the magnitude of spillovers generated by 112 academic “superstars” who died prematurely and unexpectedly, thus providing an exogenous source of variation in the structure of their collaborators’ coauthorship networks. Following the death of a superstar, we find that collaborators experience, on average, a lasting 5% to 8% decline in their quality-adjusted publication rates. By exploring interactions of the treatment effect with a variety of star, coauthor, and star/coauthor dyad characteristics, we seek to adjudicate between plausible mechanisms that might explain this finding. Taken together, our results suggest that spillovers are circumscribed in idea space, but less so in physical or social space. In particular, superstar extinction reveals the boundaries of the scientific field to which the star contributes—the “invisible college.”

“Greater is the merit of the person who facilitates the accomplishments of others than of the person who accomplishes himself.”

Rabbi Eliezer, *Babylonian Talmud, Tractate Baba Bathra 9a*

I. INTRODUCTION

Although the production of ideas occupies a central role in modern theories of economic growth (Romer 1990), the creative process remains a black box for economists (Weitzman [1998] and Jones [2009] are notable exceptions). How do innovators actually generate new ideas? Increasingly, discoveries result from the voluntary sharing of knowledge through collaboration, rather than individual efforts (Wuchty, Jones, and Uzzi 2007). The growth of scientific collaboration has important implications for the optimal allocation of public R&D funds, the apportionment of credit among scientists, the formation of scientific reputations, and ultimately

*Part of the work was performed while the first author was an Alfred P. Sloan Industry Studies Fellow. We thank the editor Larry Katz and the referees for their constructive comments, as well as various seminar audiences for their feedback, and we gratefully acknowledge the financial support of the National Science Foundation through its SciSIP program (Award SBE-0738142) and the Merck Foundation through the Columbia–Stanford Consortium on Medical Innovation. The project would not have been possible without Andrew Stellman’s extraordinary programming skills (<http://www.stellman-greene.com/>). The authors also express gratitude to the Association of American Medical Colleges for providing licensed access to the AAMC Faculty Roster, and acknowledge the stewardship of Dr. Hershel Alexander (AAMC Director of Medical School and Faculty Studies). The National Institutes of Health partially supports the AAMC Faculty Roster under Contract HHSN263200900009C. The usual disclaimer applies. pazoulay@mit.edu.

© 2010 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

The Quarterly Journal of Economics, May 2010

the design of research incentives that foster innovation and continued economic growth. Yet, we know surprisingly little about the role of collaboration among peers as a mechanism to spur the creation of new technological or scientific knowledge.

This paucity of evidence is due largely to the empirical challenges inherent in this line of inquiry. Individual-level data on the contributors to a particular innovation are generally unavailable. Furthermore, the formation of collaborative teams is the outcome of a purposeful matching process (Mairesse and Turner 2005; Fafchamps, Goyal, and van de Leij 2008), making it difficult to uncover causal effects. The design of our study tackles both of these challenges. To relax the data constraint, we focus on the academic life sciences, where a rich tradition of coauthorship provides an extensive paper trail of collaboration histories and research output. To overcome the endogeneity of the collaboration decision, we make use of the quasi-experimental variation in the structure of coauthorship networks induced by the premature and sudden death of active “superstar” scientists.¹

Specifically, we analyze changes in the research output of collaborators for 112 eminent life scientists who died suddenly and unexpectedly. We assess eminence based on the combination of seven criteria, and our procedure is flexible enough to capture established scientists with extraordinary career achievement, as well as promising young and mid-career scientists. Using the Association of American Medical Colleges (AAMC) Faculty Roster as a data source—a comprehensive, longitudinal, matched employee–employer database pertaining to 230,000 faculty members in all U.S. medical schools between 1975 and 2006—we construct a panel data set of 5,267 collaborator–star pairs, and we examine how coauthors’ scientific output (as measured by publications, citations, and National Institutes of Health (NIH) grants) changes when the superstar passes away.²

1. Other economists have used the death of prominent individuals as a source of exogenous variation in leadership, whether in the context of business firms (Bennedsen, Pérez-González, and Wolfenzon 2008) or even entire countries (Jones and Olken 2005). To our knowledge, however, we are the first to use this strategy to estimate the impact of scientific collaboration. Oettl (2008) builds on our approach by incorporating helpfulness as implied by acknowledgements to generate a list of eminent immunologists. Aizenman and Kletzer (2008) study the citation “afterlife” of 16 economists who died prematurely, shedding light on the survival of scientific reputation.

2. To be clear, our focus is on faculty peers rather than trainees, and thus our results should be viewed as capturing inter-laboratory spillovers rather than mentorship effects. For evidence on the latter, see Azoulay, Liu, and Stuart (2009).

The study's focus on the scientific elite can be justified on both substantive and pragmatic grounds. The distribution of publications, funding, and citations at the individual level is extremely skewed (Lotka 1926; de Solla Price 1963) and only a tiny minority of scientists contribute through their published research to the advancement of science (Cole and Cole 1972). Stars also leave behind a corpus of work and colleagues with a stake in the preservation of their legacy, making it possible to trace back their careers, from humble beginnings to wide recognition and acclaim.

Our results reveal a lasting 5% to 8% decrease in the quality-adjusted publication output of coauthors in response to the sudden and unexpected loss of a superstar. Though close and recent collaborators see their scientific output fall even more, these differential effects are small in magnitude and statistically insignificant. Therefore, the process of replacing missing skills within ongoing collaborative teams cannot, on its own, explain our core result.

The importance of learning through on-the-job social interactions can be traced back to the Talmudic era (as evidenced by the epigraph to this paper), as well as canonical writings by Alfred Marshall (1890) and Robert Lucas (1988).³ Should the effects of exposure to superstar talent be interpreted as laying bare the presence of knowledge spillovers? Because we identify 47 coauthors per superstar on average, we exploit rich variation in the characteristics of collaborative relationships to assess the relative importance of several mechanisms that could plausibly account for our main finding.

A jaundiced view of the academic reward system provides the backdrop for a broad class of stories. Their common thread is that collaborating with superstars deepens social connections that might make researchers more productive in ways that have little to do with scientific knowledge, for example, by connecting coauthors to funding resources, editorial goodwill, or potential coauthors. Yet we find no differential impact on coauthors of stars well-connected to the NIH funding apparatus, on coauthors of stars more central in the collaboration network, or on former trainees. These findings do not jibe with explanations stressing the gatekeeping role of eminent scientists.

3. A burgeoning empirical literature examines the influence of peer effects on shirking behavior in the workplace (Costa and Khan 2003; Bandiera, Barankay, and Rasul 2005; Mas and Moretti 2009). Because "exposure" does not involve the transmission of knowledge, these spillovers are conceptually distinct from those that concern us here.

Rather, the effects of superstar extinction appear to be driven by the loss of an irreplaceable source of ideas. We find that coauthors proximate to the star in intellectual space experience a sharper decline in output, relative to coauthors who work on less related topics. Furthermore, the collaborators of stars whose work was heavily cited at the time of their death also undergo steeper decreases, relative to collaborators of superstars of less renown. Together, these results paint a picture of an invisible college of coauthors bound together by interests in a fairly specific scientific area, which suffers a permanent and reverberating intellectual loss when it loses its star.

The rest of the paper proceeds as follows. In the next section, we describe the construction of the sample of matched superstars and collaborators, as well as our empirical strategy. Section III provides descriptive statistics at the coauthor and dyad level. We report the results in Section IV. Section V concludes.

II. SETTING, DATA, AND MATCHED SAMPLE CONSTRUCTION

The setting for our empirical work is the academic life sciences. This sector is an important one to study for several reasons. First, there are large public subsidies for biomedical research in the United States. With an annual budget of \$29.5 billion in 2008, support for the NIH dwarfs that of other national funding agencies in developed countries (Cech 2005). Deepening our understanding of knowledge production in this sector will allow us to better assess the return to these public investments.

Second, technological change has been enormously important in the growth of the health care economy, which accounts for roughly 15% of U.S. GDP. Much biomedical innovation is science-based (Henderson, Orsenigo, and Pisano 1999), and interactions between academic researchers and their counterparts in industry appear to be an important determinant of research productivity in the pharmaceutical industry (Cockburn and Henderson 1998; Zucker, Darby, and Brewer 1998).

Third, academic scientists are generally paid through soft money contracts. Salaries depend on the amount of grant revenue raised by faculty, thus providing researchers with high-powered incentives to remain productive even after they secure a tenured position.

Last, introspective accounts by practicing scientists indicate that collaboration plays a large role in both the creation and

diffusion of new ideas (Reese 2004). Knowledge and techniques often remain partially tacit until long after their initial discovery, and are transmitted within the confines of tightly knit research teams (Zucker and Darby 2008).

II.A. Superstar Sample

Our basic approach is to rely on the death of superstar scientists to estimate the magnitude of knowledge spillovers onto colleagues. From a practical standpoint, it is more feasible to trace back the careers of eminent scientists than to perform a similar exercise for less eminent ones. We began by delineating a set of 10,349 elite life scientists (roughly 5% of the entire relevant labor market), who are so classified if they satisfy at least one of the following criteria for cumulative scientific achievement: (1) highly funded scientists; (2) highly cited scientists; (3) top patenters; and (4) members of the National Academy of Sciences.

These four criteria will tend to select seasoned scientists, because they correspond to extraordinary achievement over an entire scientific career. We combine these measures with three others that capture individuals who show great promise at the early and middle stages of their scientific careers, whether or not these episodes of productivity endure for long periods of time: (5) NIH MERIT awardees; (6) Howard Hughes medical investigators; and (7) early career prize winners. Appendix I provides additional details regarding these seven metrics of “superstardom.”

We trace these scientists’ careers from the time they obtained their first positions as independent investigators (typically after a postdoctoral fellowship) until 2006. We do so through a combination of curriculum vitae, NIH biosketches, *Who’s Who* profiles, accolades/obituaries in medical journals, National Academy of Sciences biographical memoirs, and Google searches. For each one of these individuals, we record employment history, degree held, date of degree, gender, and up to three departmental affiliations. We also cross-reference the list with alternative measures of scientific eminence. For example, the elite subsample contains every U.S.-based Nobel Prize winner in medicine and physiology since 1975, and a plurality of the Nobel Prize winners in chemistry over the same time period.

Though we apply the convenient moniker of “superstar” to the entire group, it should be clear that there is substantial heterogeneity in intellectual stature within the elite sample. This

variation provides a unique opportunity to examine whether the effects we estimate correspond to vertical effects (spillovers from the most talented agents onto those who are less distinguished) or peer effects (spillovers between agents of roughly comparable stature).

The scientists who are the focus of this paper constitute a subset of this larger pool of 10,349. We impose several additional criteria to derive the final list. First, the scientist's death must occur between 1979 and 2003. This will enable us to observe at least four years' (resp. three years') worth of scientific output for every colleague before (resp. after) the death of the superstar collaborator. Second, he or she must be 67 years of age or less at the time of death (we will explore the sensitivity of our results to this age cutoff later). Third, we require evidence, in the form of published articles and/or NIH grants, that these scientists have not entered a preretirement phase of their careers prior to the time of death. This constraint is somewhat subjective, but we validate in the Online Appendix our contention that the final set is limited to scientists who are "research-active" at the time of their death. These sequential screens delineate a set of 248 scientists. Finally, we limit our attention to the subset of stars who died suddenly and unexpectedly. This is less difficult than it might seem, because the vast majority of obituaries mention the cause of death explicitly.⁴ After eliminating 136 scientists whose deaths could have been anticipated by their colleagues, we are left with 112 extinct superstars (their names, cause of death, and institutional affiliations are listed in Table W1 in the Online Appendix).

Table I provides descriptive statistics for the superstar sample. The average star received his degree in 1963, died at 57 years old, and worked with 47 coauthors during his lifetime. On the output side, the stars each received an average of roughly 11 million dollars in NIH grants (excluding center grants) and published 139 papers that had garnered 8,190 citations as of early 2008.

II.B. The Universe of Potential Colleagues

Information about the superstars' colleagues stems from the Faculty Roster of the Association of American Medical Colleges,

4. We exclude from the sample one scientist who took his own life, and a further two for whom suicide could not be ruled out. In ten other instances, the cause of death could not be ascertained from the obituaries and we contacted former collaborators individually to clarify the circumstances of the superstar's passing.

TABLE I
SUMMARY STATISTICS FOR SUPERSTAR SCIENTISTS (N = 112)

	Mean	Median	Std. Dev.	Min.	Max.
Birth age at death	57.170	58	7.042	37	67
Degree year	1962.741	1964	10.193	1942	1984
M.D.	0.420	0	0.496	0	1
Ph.D.	0.438	0	0.498	0	1
M.D./Ph.D.	0.143	0	0.351	0	1
Female	0.063	0	0.243	0	1
U.S. born	0.786	1	0.412	0	1
No. of collaborators	47.027	37	34.716	3	178
NIH review panel membership (past 5 yrs)	0.045	0	0.207	0	1
No. of collabs. in NIH review panels (past 5 yrs)	1.330	1	1.657	0	7
Career no. of publications	139.607	121	91.371	25	473
Career no. of citations	8,190	6,408	7,593	435	38,941
Career NIH funding	\$10,722,590	\$8,139,397	\$12,057,638	\$0	\$70,231,584

Notes. Sample consists of 112 superstar life scientists who died suddenly while still actively engaged in research. See Appendix I and Section IIA for details on sample construction. Degree year denotes the year of the most recent degree attained by the superstar. Number of collaborators is defined as the number of distinct coauthors within the scientists' cumulative stock of publications. NIH review panel membership denotes stars who were members of an NIH review panel in the five years prior to their death, and the number of collaborators in NIH review panels refers to the number of coauthors of each superstar who sat on NIH review panels in the five years prior to the star's death. We use the terms "star" and "superstar" interchangeably.

to which we secured licensed access for the years 1975 through 2006. The roster is an annual census of all U.S. medical school faculty in which each faculty is linked across yearly cross sections by a unique identifier.⁵ When all cross sections are pooled, we obtain a matched employee/employer panel data set. For each of the 230,000 faculty members that appear in the roster, we know the full name, the type of degrees received and the years they were awarded, gender, up to two departments, and medical school affiliation. An important implication of our reliance on the AAMC Faculty Roster is that the interactions we can observe in the data take place between faculty members, rather than between faculty members and trainees (graduate students or postdoctoral fellows).⁶

Because the roster only lists medical school faculty, however, it is not a complete census of the academic life sciences. For instance, it does not list information for faculty at institutions such as MIT, the University of California at Berkeley, Rockefeller University, the Salk Institute, or the Bethesda campus of the NIH; it also ignores faculty members in arts and sciences departments—such as biology and chemistry—if they do not hold joint appointments at local medical schools.⁷

Our interest lies in assessing the benefits of exposure to superstar talent that accrue through collaboration. Therefore, we focus on the one-degree, egocentric coauthorship network for the sample of 112 extinct superstars. To identify coauthors, we have developed a software program, the Stars/Colleague Generator, or S/CGEN.⁸ The source of the publication data is PubMed, an online resource

5. AAMC does not collect data from each medical school with a fixed due date. Instead, it collects data on a rolling basis, with each medical school submitting on a time frame that best meets its reporting needs. Nearly all medical schools report once a year, whereas many medical schools update once a semester.

6. To the extent that former trainees go on to secure faculty positions, they will be captured by our procedure even if the date of coauthorship predates the start of their independent career.

7. This limitation is less important than might appear at first glance. First, we have no reason to think that colleagues located in these institutions differ in substantive ways from those based in medical schools. Second, all our analyses focus on *changes* in research productivity over time for a given scientist. Therefore, the limited coverage is an issue solely for the small number of faculty who transition in and out of medical schools from (or to) other types of research employment. For these faculty, we were successful in filling career gaps by combining the AAMC Faculty Roster with the NIH data.

8. The software can be used by other researchers under an open-source (GNU) license. It can be downloaded, and detailed specifications accessed from the Web site <http://stellman-greene.com/SCGen/>. Note that the S/CGEN takes the AAMC Faculty Roster as an input; we are not authorized to share these data with third parties. However, they can be licensed from AAMC, provided a local IRB gives

from the National Library of Medicine that provides fast, free, and reliable access to the biomedical research literature. In a first step, S/CGEN downloads from the Internet the entire set of English-language articles for a superstar, provided they are not letters to the editor, comments, or other “atypical” articles. From this set of publications, S/CGEN strips out the list of coauthors, eliminates duplicate names, matches each coauthor with the Faculty Roster, and stores the identifier of every coauthor for whom a match is found. In a final step, the software queries PubMed for each validated coauthor, and generates publication counts as well as coauthorship variables for each superstar/colleague dyad, in each year. In the Online Appendix, we provide details on the matching procedure, how we guard against the inclusion of spurious coauthors, and our approach to addressing measurement error when tallying the publication output of coauthors with common names.

II.C. Identification Strategy

A natural starting point for identifying the effect of superstar death is to examine changes in collaborator research output after the superstar passes away, relative to when he or she was still alive, using a simple collaborator fixed effects specification. Because the extinction effect is mechanically correlated with the passage of time, as well as with the coauthor’s age, our specifications must include life-cycle and period effects, as is the norm in studies of scientific productivity (Levin and Stephan 1991). In this framework, the control group that pins down the counterfactual age and calendar time effects for the coauthors that currently experience the death of a superstar consists of coauthors whose associated superstars died in earlier periods, or will die in future periods. Despite its long pedigree in applied economics (e.g., Grogger [1995]; Reber [2005]), this approach may be problematic in our setting.

First, coauthors observed in periods after the deaths of their associated superstars are not appropriate controls if the event negatively affected the trend in their output; if this is the case, fixed effects will underestimate the true effect of superstar extinction. Second, collaborations might be subject to idiosyncratic life-cycle patterns, with their productive potential first increasing over time, eventually peaking, and thereafter slowly declining; if

its approval and a confidentiality agreement protects the anonymity of individual faculty members.

this is the case, fixed effects will overestimate the true effect of superstar extinction, at least if we rely on collaborators treated in earlier or later periods as an “implicit” control group.

To mitigate these threats to identification, our preferred empirical strategy relies on the selection of a matched control for each scientist who experiences the death of a superstar collaborator. These control scientists are culled from the universe of coauthors for the 10,000 superstars who do not die (see Section II.A). Combining the treated and control samples enables us to estimate the effect of superstar extinction in a difference-in-differences (DD) framework. Using a “coarsened exact matching” procedure detailed in Appendix II, the control coauthors are chosen so that (1) treated scientists exhibit no differential output trends relative to controls up to the time of superstar death; (2) the distributions of career age at the time of death are similar for treated and controls; (3) the time paths of output for treated and control coauthors are similar up to the time of death; and (4) the dynamics and main features of collaboration (number of coauthorships at the time of death; time elapsed since first and last coauthorship; status of the superstar collaborator as summarized by cumulative citations in the year of death) are balanced between treated and control groups.

However, adding this control group to the basic regression does not, by itself, yield a specification where the control group consists *exclusively* of matched controls. Figure A.1 displays the trends in average and median number of quality-adjusted publications, for treated and control collaborators respectively, without any adjustment for age or calendar time effects. This raw comparison is not without its problems, because it involves centering the raw data around the time of death, thus ignoring the lack of congruence between experimental and calendar time. Yet it is completely nonparametric, and provides early evidence that the loss of a superstar coauthor leads to a decrease in collaborators’ publication output. Furthermore, the magnitudes of the estimates presented below are very similar whether or not control scientists are added to the estimation sample.

Another potential concern with the addition of this “explicit” control group is that control coauthors could be affected by the treatment of interest. No scientist is an island. The set of coauthors for our 10,349 elite scientists comprises 65% of the labor market, and the remaining 35% corresponds in large part to clinicians who hold faculty appointments but do not publish

regularly. Furthermore, the death of a prominent scientist could affect the productivity of noncoauthors if meaningful interactions take place in “idea space,” as we propose. Thus, in robustness checks, we check whether eliminating from the estimation sample treated and control collaborators separated by small path lengths in the coauthorship network matters for the substance, or even the magnitude, of our main results.

III. DESCRIPTIVE STATISTICS

When applied to our sample of 112 extinct superstars, S/CGEN identifies 5,267 distinct coauthors with unique PubMed names.⁹ Our matching procedure can identify a control scientist for 5,064 (96%) of the treated collaborators. The descriptive statistics in Table II pertain to the set of $2 \times 5,064 = 10,128$ matched treated and control scientists. The covariates of interest are measured in the (possibly counterfactual) year of death for the superstar. We distinguish variables that are inherently dyadic (e.g., colocation at time of death) from variables that characterize the coauthor at a particular point of time (e.g., NIH R01 funding at the time of death).

III.A. Dyadic Variables

Of immediate interest is the distribution of coauthorship intensity at the dyad level. Although the average number of coauthorships is slightly less than three, the distribution is extremely skewed (Figure I). We define “casual” dyads as those that have two or fewer coauthorships with the star, “regular” dyads as those with three to nine coauthorships, and “close” dyads as those with ten or more coauthorships. Using these cutoffs, regular dyads correspond to those between the 75th and the 95th percentile of coauthorship intensity, whereas close dyads correspond to those above the 95th percentile.

We focus next on collaboration age and recency. On average, collaborations begin eleven years before the star’s death, and time since last coauthorship is slightly more than nine years. In other words, most of the collaborations in the sample do not involve active research projects at the time of death. Recent collaborations

9. Whenever a scientist collaborates with more than one extinct superstar (this is relevant for 10% of the sample), we take into account only the first death event. We have verified that limiting the estimation sample to collaborators with one and only one tie to a superstar who dies does not change the substance, or even the magnitudes, of our core result.

TABLE II
SUMMARY STATISTICS FOR COLLABORATORS IN THE YEAR OF SUPERSTAR DEATH

	Mean	Median	Std. dev.	Min.	Max.
Control collaborators (<i>N</i> = 5,064)					
No. of weighted publications	18,314	8	27.917	0	342
Cum. no. of weighted publications	327,330	187	409,098	0	3,968
Holds R01 grant	0.559	1	0.497	0	1
Colocated	0.144	0	0.351	0	1
Career age	23.698	23	9.963	1	59
Elite	0.093	0	0.290	0	1
Cum. no. of coauthorships	2,734	1	4,339	1	69
No. of other superstar collaborators	2,746	2	3,516	0	31
Years since first coauthorship	10.949	10	7.901	0	42
Years since last coauthorship	9.275	8	7.774	0	41
Former trainee of the star	0.070	0	0.255	0	1
"Accidental" collaborator	0.076	0	0.265	0	1
MeSH keyword overlap	0.265	0	0.162	0	1
Superstar citation count	10,083	7,245	8,878	99	90,136

TABLE II
(CONTINUED)

	Mean	Median	Std. dev.	Min.	Max.
	Treated collaborators ($N = 5,064$)				
No. of weighted publications	19,068	8	31,656	0	491
Cum. no. of weighted publications	334,905	187	436,927	0	4,519
Holds R01 grant	0.571	1	0.495	0	1
Colocated	0.123	0	0.328	0	1
Career age	23.761	23	9.969	0	59
Elite	0.077	0	0.266	0	1
Cum. no. of coauthorships	2,835	1	4,894	1	75
No. of other superstar collaborators	3,087	2	4,255	0	44
Years since first coauthorship	11.022	10	7,896	0	39
Years since last coauthorship	9.255	8	7,728	0	38
Former trainee of the star	0.084	0	0.278	0	1
“Accidental” collaborator	0.075	0	0.264	0	1
MeSH keyword overlap	0.259	0	0.157	0	1
Superstar citation count	10,228	7,239	7,952	397	34,746

Notes. The samples consist of faculty collaborators of 112 deceased superstar life scientists and an equal number of matched control coauthors. See Sections II.B and III for details on the sample construction and variable definitions and Appendix II for details on the matching procedure. All variables are measured as of the year of superstar death. Publications are JIF-weighted.

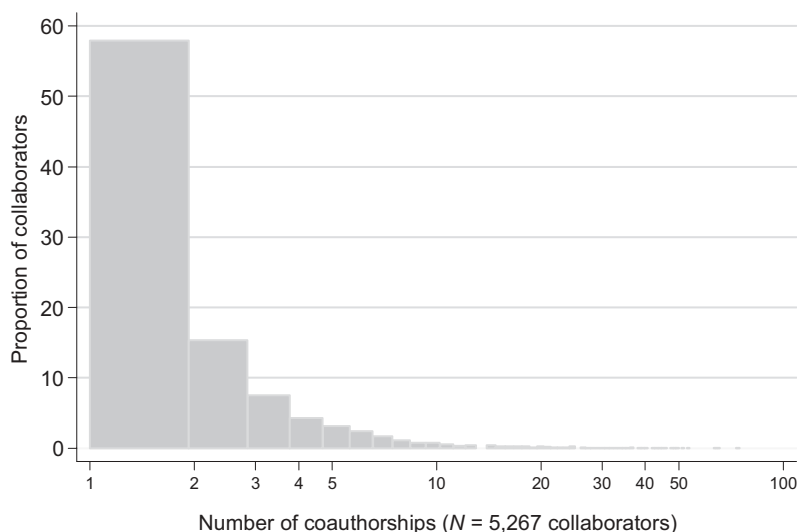


FIGURE I
Distribution of Coauthorship Intensity

(those that involve at least one coauthorship in the three years preceding the passing of the superstar) map into the top quartile of collaboration recency at the dyad level.

The research collaborations studied here occur between faculty members, who often run their own labs (a conjecture reinforced by the large proportion of coauthors with independent NIH funding). Yet it is interesting to distinguish collaborators who trained under a superstar (either in graduate school or during a postdoctoral fellowship) from those collaborations initiated at a time in which both nodes in the dyad already had a faculty appointment. Although there is no roster of mentor/mentee pairs, coauthorship norms in the life sciences provide an opportunity to identify former trainees. Specifically, we flag first-authored articles published within a few years of receipt of the coauthor's degree in which the superstar appears in last position on the authorship roster.¹⁰ Using this method, we find that roughly 8% of treated collaborators were former trainees of the associated superstar.

10. The purported training period runs from three years before graduation to four years after graduation for Ph.D.'s and M.D./Ph.D.'s, and from the year of graduation to six years after graduation for M.D.'s. Recall that we do not observe the population of former trainees, but only those trainees who subsequently went on to get full-time faculty positions in the United States. One concern is selection bias for the set of former trainees associated with superstars who died when they had just completed training. To guard against this potential source of bias, we

We now examine the spatial distribution of collaborations. Slightly more than 12% of collaborations correspond to scientists who were colocated at the time of superstar extinction; though this is not the focus of the paper, the proportion of local collaborations has declined over time, as many previous authors have documented (e.g., Rosenblat and Möbius [2004]). We also provide a measure of collaborators' proximity in ideas space. Every publication indexed by PubMed is tagged by a large number of descriptors, selected from a dictionary of approximately 25,000 MeSH (Medical Subject Headings) terms. Our measure of intellectual proximity between members of a dyad is simply the number of unique MeSH terms that overlap in their noncoauthored publications, normalized by the total number of MeSH terms used by the superstar's coauthor. The time window for the calculation is the five years that precede the passing of the superstar. The distribution of this variable is displayed in Figure II.¹¹

Finally, we create a measure of social proximity that relies not on the quantity of coauthored output, but on the degree of social interaction it implies. We focus on the pairs involving coauthors who, whenever they collaborate, find themselves in the middle of the authorship list. Given the norms that govern the allocation of credit in the life sciences, these coauthors are likely to share the least amount of social contact. Of the dyads in the sample, 7.5% correspond to this situation of "accidental coauthorship"—the most tenuous form of collaboration.

III.B. Coauthor Variables

We briefly mention demographic characteristics that do not play a role in the econometric results but are nonetheless informative. The sample is 20% female (only 10% of the superstars are women); approximately half of all coauthors are M.D.'s, 40% are Ph.D.'s, and the remainder are M.D./Ph.D.'s; and a third are affiliated with basic science departments (as opposed to clinical or public health departments). The coauthors are about eight years younger than the superstars on average (1971 vs. 1963 for the year of highest degree).

Coauthors lag behind superstars in terms of publication output at the time of death, but the difference is not dramatic (88

eliminated all former trainees from the sample with career age less than five at the time of death.

11. Further details on its construction are provided in the Online Appendix, Section II.

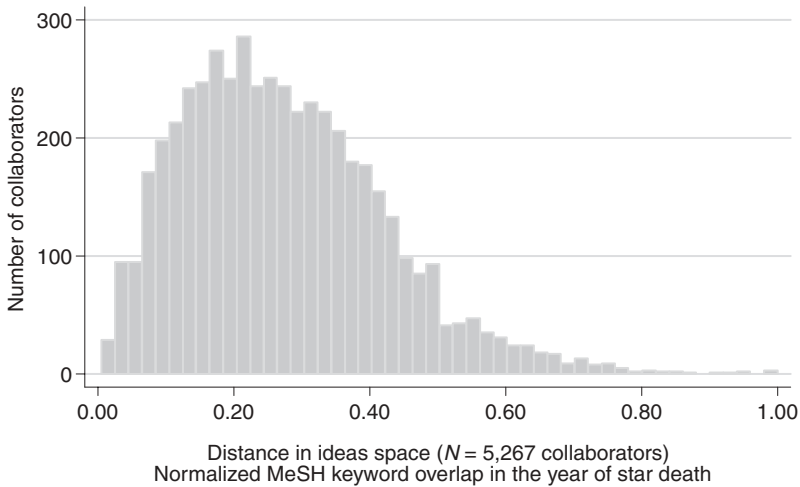


FIGURE II

Proximity in Ideas Space

Measure of distance in ideas space is defined as the number of unique MeSH terms that overlap between the colleague's and superstar's publications (excluding coauthored output), normalized by the total number of MeSH terms used in the colleague's total publications. This measure is calculated for articles published in the five years preceding superstar death. Calculation excludes coauthored publications.

vs. 140 articles, on the average). Assortative matching is present in the market for collaborators, as reflected by the fact that 2,852 (28.16%) of our 10,128 coauthors belong to the elite sample of 10,349 scientists. Of collaborators, 55% had served as PI on at least one NIH R01 grant when the superstar passed away, whereas about 8% of the treated collaborators (and 9% of the controls) belong to a more exclusive elite: Howard Hughes medical investigators, members of the NAS, or MERIT awardees.

The estimation sample pools observations between 1975 and 2006 for the dyads described above. The result is an unbalanced panel data set with 153,508 collaborator \times year observations (treated collaborators only) or 294,943 collaborator \times year observations (treated and control collaborators).

IV. RESULTS

The exposition of the econometric results proceeds in three stages. After a brief review of methodological issues, we provide results that pertain to the main effect of superstar exposure on

publication rates. Second, we examine whether this effect merely reflects the adverse impact of losing important skills within ongoing collaborative teams. Third, we attempt to explicate the mechanism, or set of mechanisms, responsible for the results. We do so by exploring heterogeneity in the treatment through the interaction of the postdeath indicator variable below with various attributes of the superstar, colleague, and dyad.

IV.A. Econometric Considerations

Our estimating equation relates colleague j 's output in year t to characteristics of j , superstar i , and dyad ij :

$$(1) \quad E[y_{jt} | X_{ijt}] = \exp[\beta_0 + \beta_1 \text{AFTER_DEATH}_{it} + f(\text{AGE}_{jt}) + \delta_t + \gamma_{ij}],$$

where y is a measure of research output, AFTER_DEATH denotes an indicator variable that switches to one the year after the superstar dies, $f(\text{AGE}_{jt})$ corresponds to a flexible function of the colleague's career age, the δ_t 's stand for a full set of calendar year indicator variables, and the γ_{ij} 's correspond to dyad fixed effects, consistent with our approach to analyze *changes* in j 's output following the passing of superstar i .

The dyad fixed effects control for many individual characteristics that could influence research output, such as gender or degree. Academic incentives depend on the career stage; given the shallow slope of posttenure salary increases, Levin and Stephan (1991) suggest that levels of investment in research should vary over the career life cycle. To flexibly account for life-cycle effects, we include seventeen indicator variables corresponding to different career age brackets, where career age measures the number of years since a scientist earned his/her highest degree (M.D. or Ph.D.).¹² In specifications that include an interaction between the treatment effect and some covariates, the models also include a set of interactions between the life-cycle effects and these covariates.

Estimation. The dependent variables of interest, including weighted or unweighted publication counts and NIH grants awarded, are skewed and nonnegative. For example, 24.80% of the

12. The omitted category corresponds to faculty members in the very early years of their careers (before age -3). It is not possible to separately identify calendar year effects from age effects in the "within" dimension of a panel in a completely flexible fashion, because one cannot observe two individuals at the same point in time who have the same (career) age but earned their degrees in different years (Hall, Mairesse, and Turner 2007).

collaborator/year observations in the data correspond to years of no publication output; the figure climbs to 87.40% if one focuses on the count of successful grant applications. Following a longstanding tradition in the study of scientific and technical change, we present conditional quasi-maximum likelihood (QML) estimates based on the fixed-effect Poisson model developed by Hausman, Hall, and Griliches (1984). Because the Poisson model is in the linear exponential family, the coefficient estimates remain consistent as long as the mean of the dependent variable is correctly specified (Gouriéroux, Montfort, and Trognon 1984).¹³

Inference. QML (i.e., “robust”) standard errors are consistent even if the underlying data-generating process is not Poisson. In fact, the Hausman et al. estimator can be used for any nonnegative dependent variables, whether integer or continuous (Santos Silva and Tenreiro 2006), as long as the variance/covariance matrix is computed using the outer product of the gradient vector (and therefore does not rely on the Poisson variance assumption). Further, QML standard errors are robust to arbitrary patterns of serial correlation (Wooldridge 1997), and hence immune to the issues highlighted by Bertrand, Duflo, and Mullainathan (2004) concerning inference in DD estimation. We cluster the standard errors around superstar scientists in the results presented below.

Dependent Variables. Our primary outcome variable is a coauthor’s number of publications. Because SC/GEN matches the entire authorship roster for each article, we can separate those publications coauthored with the superstar from those produced independent of him/her. We perform a quality adjustment by weighting each publication by its journal impact factor (JIF)—a measure of the frequency with which the average article in a journal has been cited in a particular year. One obvious shortcoming of this adjustment is that it does not account for differences in impact within a given journal. In the Online Appendix (Section V), we present additional results based on article-level citation outcomes.

IV.B. Main Effect of Superstar Extinction

Table III presents our core results. Column (1a) examines the determinants of the 5,267 treated coauthors’ JIF-weighted

13. In the Online Appendix (Section IV), we show that OLS yields results very similar to QML Poisson estimation for our main findings.

TABLE III
IMPACT OF SUPERSTAR DEATH ON COLLABORATORS' PUBLICATION RATES

	Panel A		Panel B	
	All JIF-weighted publications		JIF-weighted publications written with others	
	Without ctrls (1a)	With ctrls (1b)	Without ctrls (2a)	With ctrls (2b)
After death	-0.092** (0.022)	-0.086** (0.025)	-0.057** (0.022)	-0.054* (0.024)
Log pseudo-likelihood	-974,285	-1,832,594	-950,864	-1,783,958
No. of observations	153,508	294,943	153,508	294,943
No. of collaborators	5,267	10,128	5,267	10,128

Notes. Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. All models incorporate a full suite of year effects as well as seventeen age category indicator variables (career age less than -3 is the omitted category). Exponentiating the coefficients and differencing from one yield numbers interpretable as elasticities. For example, the estimates in column (1a) imply that collaborators suffer on average a statistically significant $(-1 - 1) \exp[-0.092] = 8.79\%$ decrease in the rate of publication after their superstar coauthor passes away. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

* $p < .05$.
** $p < .01$.

publication output. We find a sizable and significant 8.8% decrease in the yearly number of quality-adjusted publications coauthors produce after the star dies. Column (2b) adds the set of control coauthors to the estimation sample. This reduces our estimate of the treatment effect only slightly, to a statistically significant 8.2% decline.

Columns (1b) and (2b) provide the results for an identical set of specifications, except that we modify the dependent variable to exclude publications coauthored with the superstar when computing the JIF-weighted publication counts. The contrast between the results in Panels A and B elucidates scientists' ability to substitute toward new collaborative relationships upon the death of their superstar coauthor. The effects are now smaller, but they remain statistically significant.

We also explore the dynamics of the effects uncovered in Table III. We do so by estimating a specification in which the treatment effect is interacted with a set of indicator variables corresponding to a particular year relative to the superstar's death, and then graphing the effects and the 95% confidence interval around them (Figures IIIA and IIIB, corresponding to Table III, columns (1b) and (2b)). Following the superstar's death, the treatment effect increases monotonically in absolute value, becoming statistically significant three to four years after death. Two aspects of this result are worthy of note. First, we find no evidence of recovery—the effect of superstar extinction appears permanent. Though we will explore mechanisms in more detail below, this seems inconsistent with a bereavement-induced loss in productivity. Second, the delayed onset of the effect makes sense because it plausibly takes some time to exhaust the productive potential of the star's last scientific insights. In addition, the typical NIH grant cycle is three to five years, and the impact of a superstar's absence may not really be felt until it becomes time to apply for a new grant.

In all specifications, the results with and without controls are quite similar. In the remainder of the paper, the estimations sample always include the “explicit” control group, though the results without it are qualitatively similar.

IV.C. Imperfect Skill Substitution

Collaborative research teams emerge to pool the expertise of scientists, who, in their individual capacity, face the “burden of knowledge” problem identified by Jones (2009). Upon the death of

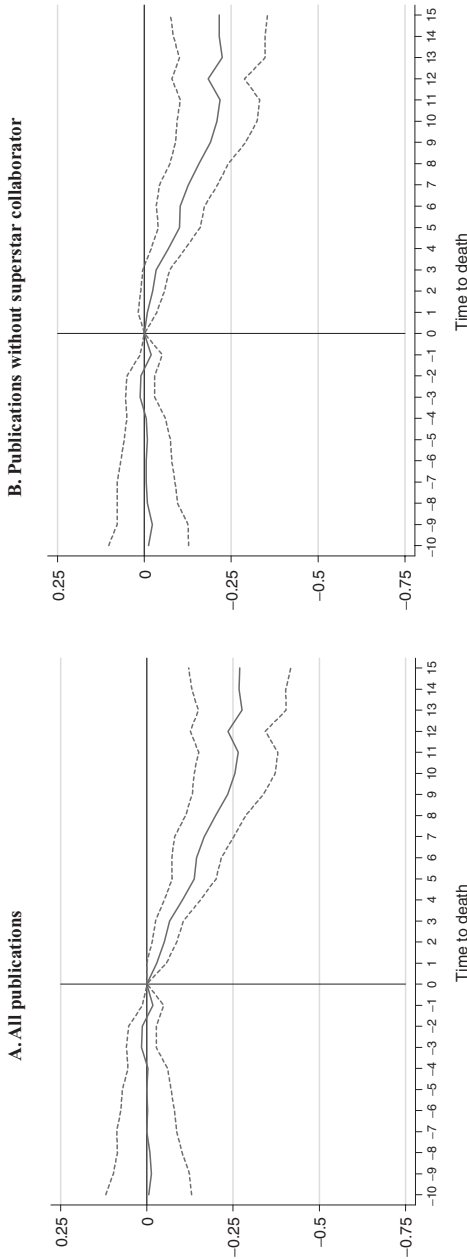


FIGURE III
Dynamics of the Treatment Effect

The solid lines in the above plots correspond to coefficient estimates of conditional fixed effects quasi-maximum likelihood Poisson specifications in which the weighted publication output of a collaborator is regressed onto year effects, seventeen indicator variables corresponding to different age brackets, and interactions of the treatment effect with 27 indicator variables corresponding to eleven years before the year of death and prior, ten years before the year of death, nine years before the year of death, ..., fourteen years after the year of death, and fifteen years after the year of death and above (the indicator variable for treatment status interacted with the year of death is omitted). The 95% confidence interval (corresponding to robust standard errors, clustered around superstars) around these estimates is plotted with dashed lines. The graph for all publications (A) uses column (1b) of Table III as a baseline (i.e., treated and control collaborators, the dependent variable, includes all of the collaborator's publications); the graph for publications without superstar collaborator (B) uses column (2b) of Table III as a baseline (i.e., treated and control collaborators, the dependent variable, is limited to the collaborator's publications in which the superstar does not appear on the authorship list).

a key collaborator, other team members might struggle to suitably replace the pieces of knowledge that were embodied in the star. Viewed in this light, the effects uncovered in Table III could be considered unsurprising—a mechanical reflection of the skill substitution process. The fact that publications with coauthors other than the superstar are adversely affected and the permanence of the treatment effect already suggests other forces are at play. The imperfect skill substitution (ISS) story carries additional testable implications. First, one would expect coauthors with closer relationships with the star to suffer steeper decreases in output; the same would be expected for recent or new collaborations, which are more likely to involve ongoing research efforts at the time of death. Table IV examines these implications empirically.

We find that regular and, to a lesser extent, close collaborators are indeed more negatively affected than casual collaborators, but these differential losses are relatively small in magnitude and statistically insignificant (column (1a)). The same holds true for recent collaborations (column (2a), at least one joint publication in the three years preceding the star's death) and for young collaborations (those for which the first coauthored publication appeared in the five years preceding the star's death, unreported results available from the authors). Columns (1b) and (2b) provide results for an identical set of specifications, but excluding publications coauthored with the superstar. The contrast between the results in columns (1a) and (1b) (resp. (2a) and (2b)) elucidates scientists' ability to substitute toward new collaborative relationships upon the death of their superstar coauthor. The estimates imply that close and, to a lesser extent, recent coauthors do manage to find replacement collaborators (or to intensify already existing collaborations). Close collaborators experience an imprecisely estimated 6.18% average *increase* in their quality-adjusted publications written independent of the star, but this is only a partial offset for the overall loss documented in column (1a). We find that casual collaborators and collaborators without a recent coauthorship see their independent output decline respectively by 5.54% (column (1b)) and 8.25% (column (2b)). Very similar results are obtained when all these covariates are combined into one specification (columns (3a) and (3b)).

Although the differential impacts on the closest and most recent collaborators are not statistically significant, they do appear to move in the direction that supports the skill substitution hypothesis. However, the inability of scientists to compensate fully

TABLE IV
COLLABORATOR PUBLICATION RATES AND IMPERFECT SKILL SUBSTITUTION

	Coauthorship intensity		Coauthorship recency		Coauthorship intensity & recency	
	All pubs. (1a)	Pubs. written with others (1b)	All pubs. (2a)	Pubs. written with others (2b)	All pubs. (3a)	Pubs. written with others (3b)
After death	−0.076** (0.026)	−0.057* (0.025)	−0.087** (0.024)	−0.074** (0.024)	−0.080** (0.024)	−0.075** (0.024)
After death × regular collaborator	−0.044 (0.041)	−0.020 (0.042)			−0.039 (0.042)	−0.018 (0.043)
After death × close collaborator	−0.026 (0.068)	0.117 (0.073)			−0.014 (0.069)	0.119 (0.074)
After death × at least one coauthorship in the three years preceding star's death			−0.022 (0.038)	0.032 (0.039)	−0.021 (0.039)	0.028 (0.039)
Log pseudo-likelihood	−1,831,987	−1,781,742	−1,822,664	−1,775,680	−1,821,791	−1,774,167
No. of observations	294,943	294,943	294,943	294,943	294,943	294,943
No. of collaborators	10,128	10,128	10,128	10,128	10,128	10,128

Notes. Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Regular and close collaborator are indicator variables for the number of publications coauthored by the superstar and colleague at the time of death (regular collaborations correspond to between three and nine coauthored publications; close collaborations correspond to ten or more coauthored publications; casual collaborations—the omitted category—corresponds to one or two coauthored publications). All models incorporate year effects and seventeen age category indicator variables (career age less than −3 is the omitted category), as well as seventeen interaction terms between the age effects and each covariate of interest (i.e., column (3b) includes a total of $3 \times 17 = 51$ age-specific interaction terms). Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

* $p < .05$.
** $p < .01$.

for the loss of expected future collaborations through alternative relationships, as well as the permanence of the extinction effect, demonstrate that something more than the star's skills disappears upon their death. Taken as a whole, these results suggest that the treatment effect from Table III cannot be fully explained by imperfect skill substitution within ongoing teams.

IV.D. Disentangling Mechanisms

We exploit the fine-grained level of detail in the data to sort between mechanisms which might underlie the extinction effect. Are collaborative ties with superstars conduits for tangible resources, or for knowledge and ideas? These two broad classes of explanations are not mutually exclusive, but ascertaining their relative importance matters because their welfare implications differ sharply. If superstars merely act as gatekeepers, then their deaths will lead to a reallocation of resources away from former collaborators, but may have little impact on social welfare. Conversely, if spillovers of knowledge were enabled by collaboration, their passing might result in significant welfare losses.

Superstars as Gatekeepers. Superstars may matter for their coauthors because they connect them to important resources either within their institution or in the scientific world at large. These resources might include funding, administrative clout, editorial goodwill, or other potential collaborators. We attempt to evaluate the validity of three particular implications of this story in Table V.

First, we examine whether the superstar's ties to the NIH funding apparatus moderate the magnitude of the extinction effect. Whereas social scientists sometimes emphasize the role that journal editors can have in shaping individual careers, life scientists are often more concerned that the allocation of grant dollars deviates from the meritocratic ideal. Therefore, we investigate whether the treatment effect is of larger magnitude when the star either sat on NIH review panels in the last five years, or has coauthorship ties with other scientists who sat on study sections in the recent past. In column (1), we find that this is not the case. The differential impacts are relatively small, positive in magnitude, and not statistically significant.

Second, we address the hypothesis that superstars matter because they broker relationships between scientists who would otherwise remain unaware of each other's expertise. We do so by

TABLE V
COLLABORATOR PUBLICATION RATES AND ACCESS TO RESOURCES

	Star's ties to NIH funding process (1)	Quartile of betweenness centrality (2)	Former trainee (3)	All covariates combined (4)
After death	−0.105** (0.037)	−0.067* (0.028)	−0.086** (0.025)	−0.089* (0.035)
After death × star sat on NIH review panel	0.042 (0.064)			0.024 (0.070)
After death × star's no. of coauthor ties to NIH review panelists	0.011 (0.013)			0.014 (0.015)
After death × star in fourth quartile of betweenness centrality		−0.031 (0.046)		−0.040 (0.051)
After death × coauthor is former trainee			0.056 (0.069)	0.048 (0.069)
Log pseudo-likelihood	−1,831,339	−1,831,779	−1,830,582	−1,828,754
No. of observations	294,943	294,943	294,943	294,943
No. of collaborators	10,128	10,128	10,128	10,128

Notes. Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Betweenness centrality is measured using the network of 10,349 superstar life scientists, former trainee indicates that the colleague was a graduate student or postdoctoral fellow in the laboratory of the superstar (7.69% of the collaborators). All models incorporate year effects and seventeen age category indicator variables (career age less than −3 is the omitted category), as well as seventeen interaction terms between the age effects and each covariate of interest. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

* $p < .05$.
** $p < .01$.

computing the betweenness centrality for the extinct superstars in the coauthorship network formed by the 10,349 elite scientists.¹⁴ We then rank the superstars according to quartile of betweenness and look for evidence that collaborators experience a more pronounced decline in output if their superstar coauthor was more central (column (2)). We find that collaborators with stars in the top quartile suffer additional losses, relative to collaborators of less central superstars, but this differential effect is statistically insignificant.

Finally, in column (3), we look for a differential effect of superstar death for coauthors that were also former trainees. It is possible that mentors continue to channel resources to their former associates even after they leave their laboratories, in which case one would expect these former trainees to exhibit steeper and more precipitous declines following the passing of their adviser. In fact, the differential effect is large and positive, though not statistically significant.

The evidence presented in Table V appears broadly inconsistent with the three particular gatekeeping stories whose implications we could test empirically. Our assessment of the gatekeeping mechanism must remain guarded for two reasons. First, the effect of variables used to proxy for the strength of social ties are subject to alternative interpretations. For instance, a former trainee effect could also be interpreted as providing evidence of knowledge spillovers, because mentorship can continue into the early faculty career and be extremely important for a young scholar's intellectual development. Furthermore, it is possible to think of alternative versions of the gatekeeping mechanism; as an example, superstars might be able to curry favors with journal editors on behalf of their protégés, or they might be editors themselves. We prefer to frame the findings contrapositively: it is hard to look at the evidence presented so far and conclude that access to resources is a potent way in which superstars influence their collaborators' scientific output.

Knowledge Spillovers. We now examine the possibility that stars generate knowledge spillovers onto their coauthors. In Table VI, we build a circumstantial case for the spillover view by

14. Betweenness is a measure of the centrality of a node in a network, and is calculated as the fraction of shortest paths between dyads that pass through the node of interest. In social network analysis, it is often interpreted as a measure of the influence a node has over the spread of information through the network.

TABLE VI
COLLABORATOR PUBLICATION RATES AND PROXIMITY IN GEOGRAPHIC AND INTELLECTUAL SPACE

	(1)	(2)	(3)	(4)	(5)
After death	-0.092** (0.027)	-0.067** (0.023)	-0.094** (0.022)	-0.081** (0.024)	-0.074** (0.026)
After death × colocated	0.042 (0.043)			0.037 (0.043)	0.042 (0.044)
After death × kwd. overlap in top quartile		-0.115* (0.059)		-0.114† (0.059)	-0.127* (0.057)
After death × “accidental” collaborator			0.104† (0.060)	0.111† (0.058)	0.077 (0.055)
After death × regular collaborator					-0.030 (0.044)
After death × close collaborator					0.002 (0.072)
After death × recent collaborator					-0.022 (0.038)
% of collaborators affected	13.33	25.35	7.53		
Log pseudo-likelihood	-1,831,900	-1,830,305	-1,831,787	-1,828,805	-1,817,667
No. of observations	294,943	294,943	294,943	294,943	294,943
No. of collaborators	10,128	10,128	10,128	10,128	10,128

Notes. Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Colocated indicates that the colleague and superstar were employed at the same institution at the time of superstar death. Keyword overlap is the normalized number of MeSH keywords that appear on both the colleague's and superstar's nonjoint publications. Accidental collaborators are those who only appear on coauthored publications with the superstar when both are in the middle of the authorship list. Regular and close collaborator are indicator variables for the number of publications coauthored by the superstar and colleague at the time of death (regular collaborators correspond to between three and nine coauthored publications; close collaborators correspond to ten or more coauthored publications; casual collaborators—the omitted category—correspond to one or two coauthored publications). All models incorporate year effects and seventeen age category indicator variables (career age less than -3 is the omitted category), as well as seventeen interaction terms between the age effects and the covariate of interest. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

† $p < .10$.
* $p < .05$.
** $p < .01$.

documenting evidence of additional output losses for collaborators who were more proximate with the superstar at the time of death, using two different meanings of proximity: physical and intellectual.

In column (1), we investigate the impact of physical proximity by interacting the treatment effect with an indicator variable for those collaborators who were colocated with the superstar at the time of death. We find essentially no difference between the fates of these coauthors and those of coauthors located further away—the interaction term is positive, small in magnitude, and imprecisely estimated. At first blush, this finding appears consistent with some recent work suggesting a fading role for geographic distance, both as a factor influencing the formation of teams (Rosenblat and Möbius 2004; Agrawal and Goldfarb 2008), and as a factor circumscribing the influence of peers (Kim, Morse, and Zingales 2009; Griffith, Lee, and Van Reenen 2007; Waldinger 2008). However, our estimate of the colocation interaction term conflates the effect of the loss of knowledge spillovers, the effect of the loss of help and protection provided by the star in the competition for internal resources (such as laboratory space), and the effect of any measure taken by the institution to compensate for the death of the superstar. As a result, it is unclear whether our results contradict the more conventional view that spillovers of knowledge are geographically localized (Zucker, Darby, and Brewer 1998; Ham and Weinberg 2008).¹⁵

In column (2), we investigate whether the death of a superstar coauthor has a disparate impact on the group of scientists who work on similar research problems. We proxy intellectual distance between the superstar and his/her coauthors with our measure of normalized keyword overlap. Coauthors in the top quartile of this measure at the time of death suffer output decreases that are particularly large in magnitude (-12.2%).¹⁶ This evidence is consistent with the existence of an “invisible college”—an elite of productive scientists highly visible in a research area, combined with a “scatter” of less eminent ones, whose attachment to the field may be more tenuous (de Solla Price and Beaver 1966; Crane 1972). Superstar scientists make their field of inquiry visible to others of lesser standing who might enter it; they replenish their field with

15. We thank an anonymous reviewer for making this point.

16. Specifications that include four different interactions corresponding to the four quartiles show that the treatment effect is monotonically increasing in intellectual distance, but we do not have enough statistical power to reject the hypothesis that the four coefficients are equal to one another.

fresh ideas, and their passing causes the processes of knowledge accumulation and diffusion to slow down, or even decline.

In this view, important interactions for the production of new scientific knowledge are not rigidly constrained by geographic or social space, but also take place in an ethereal “idea space.” But is the act of formal coauthorship necessary for a scientist to be brought into a superstar’s intellectual orbit? Because our sample is composed exclusively of coauthors, we cannot definitively answer this question. Yet one can use the norms of authorship in the life sciences to try to isolate collaborators whose coauthorship tie to the star is particularly tenuous: “accidental” collaborators—those who always find themselves in the middle of the authorship list. As seen in column (3), these accidental collaborators do not appear to experience net losses after the superstar’s death. This suggests that full membership in the invisible college may be difficult to secure in the absence of a preexisting social tie.

Column (4) provides evidence that the effects of physical and intellectual proximity are independent, because combining them in the same specification does not alter their magnitudes or statistical significance. Finally, column (5) demonstrates that these effects are robust to the inclusion of controls for coauthorship intensity and recency.

Table VII provides additional evidence in favor of the spillover view by examining the relationship between the magnitude of the treatment effect and the accomplishments of the star. We rank superstars according to two metrics of achievement: cumulative citations and cumulative NIH funding, and we focus on superstars in the top quartile of either distributions (where these quartiles are calculated using the population of 10,349 superstars in a given year). Column (1) shows that collaborators of heavily cited superstars suffer more following the superstar’s death, whereas column (2) shows that this is not true for collaborators of especially well-funded superstars. Column (3) puts the two effects in a single specification. Once again, it appears that it is the star’s citation impact that matters in shaping collaborators’ postextinction outcomes, rather than his/her control over a funding empire.¹⁷ We interpret these findings as buttressing our argument that it is the quality of ideas emanating from the stars, rather than simply the availability of the research funding they control, that goes missing

17. Table VII eliminates from the estimation sample the collaborators of eleven superstars who are NIH intramural scientists, and as such not eligible for extramural NIH funding.

TABLE VII
IMPACT OF SUPERSTAR STATUS ON COLLABORATORS' PUBLICATION RATES

	Superstar status citations (1)	Superstar status NIH funding (2)	Superstar status citations & NIH funding (3)
After death	-0.034 (0.036)	-0.070* (0.035)	-0.026 (0.039)
After death \times star in top quartile of cites	-0.082† (0.047)		-0.080† (0.048)
After death \times star in top quartile of NIH \$		-0.026 (0.050)	-0.016 (0.051)
Log pseudo-likelihood	-1,715,929	-1,716,213	-1,715,916
No. of observations	275,776	275,776	275,776
No. of collaborators	9,470	9,470	9,470

Notes. Estimates stem from conditional quasi-maximum likelihood Poisson specifications. Dependent variable is the total number of JIF-weighted articles authored by a collaborator of a superstar life scientist in the year of observation. Top quartiles of citations and career NIH funding are defined using the population of 10,009 superstar scientists with appointments compatible with extramural NIH funding. We exclude from the estimation sample the collaborators of eleven "intramural" NIH scientists who are *not* eligible to receive extramural funding. All models incorporate year effects and seventeen age category indicator variables (career age less than -3 is the omitted category), as well as seventeen (columns (1) and (2)) or 34 (column (3)) interaction terms between the age effects and the "top quartile" indicator variable. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

† $p < .10$.
* $p < .05$.

after their deaths. Furthermore, these results suggest that using the same empirical strategy, but applying it to a sample of “humdrum” coauthors who die, would not uncover effects similar in magnitude to those we observed in Table III. As such, they validate *ex post* our pragmatic focus on the effect of superstars.

The overall collection of results presented above help build a circumstantial case in favor of interpreting the effects of superstar extinction as evidence of missing spillovers. However, they do not enable us to reject some potentially relevant versions of the gatekeeping story—such as influence over the editorial process in important journals—nor do they allow us to learn about the effect on noncollaborators.

IV.E. Robustness and Sensitivity Checks

The Online Appendix provides additional evidence probing the robustness of these results. In Table W7, we interact the treatment effect with three indicators of collaborator status, to ascertain whether some among them are insulated from the effects of superstar extinction. Figure W3 provides evidence that the effect of superstar extinction decreases monotonically with the age of the collaborator at the time of death, becoming insignificantly different from zero after 25 years of career age. Table W8 performs a number of sensitivity checks. We verify that the effect (1) is not driven by a few stars with a large number of coauthors; (2) is robust to the inclusion of indicator variables for the age of the star; (3) is not overly sensitive to our arbitrary cutoff for the superstars’ age at death; and (4) is not sensitive to the problem of leakage through the coauthorship network between treated and control collaborators. Finally, we perform a small simulation study to validate the quasi-experiment exploited in the paper. We generate placebo dates of death for the control collaborators, where those dates are drawn at random from the empirical distribution of death events across years for the 112 extinct superstars. We then replicate the specification in Table III, column (1a), but we limit the estimation sample to the set of 5,064 control collaborators. Reassuringly, the effect of superstar extinction in this manufactured data is a precisely estimated zero.

V. CONCLUSIONS

We examine the role of collaboration in spurring the creation of new scientific knowledge. Using the premature and unexpected

deaths of eminent academic life scientists as a quasi-experiment, we find that their collaborators experience a sizable and permanent decline in quality-adjusted publication output following these events. Exploiting the rich heterogeneity in these collaborative relationships, we attempt to adjudicate between plausible mechanisms that could give rise to the extinction effect.

Neither a mechanical story whereby ongoing collaborative teams struggle to replace the skills that have gone missing, nor a gatekeeping story where stars merely serve as conduits for tangible resources is sufficient to explain our results. Rather, these effects appear to be driven, at least in part, by the existence of knowledge spillovers across members of the research team. When a superstar dies, part of the scientific field to which he or she contributed dies along with him or her, perhaps because the fount of scientific knowledge from which coauthors can draw is greatly diminished. The permanence and magnitude of these effects also suggests that even collaborations that produce a small number of publications may have long-term repercussions for the pace of scientific advance.

In the end, this paper raises as many questions as it answers. It would be interesting to know whether superstar extinction also impacts the productivity of noncoauthors proximate in intellectual space, and in which direction. The degree to which exposure to superstar talent benefits industrial firms is also potentially important and represents a fruitful area that we are pursuing in ongoing research. Future work could also usefully focus on identifying quasi-experiments in intellectual space. For instance, how do scientists adjust to sudden changes in scientific opportunities in their field? Finally, collaboration incentives and opportunities may be different when scientific progress relies more heavily on capital equipment; an examination of the generalizability of our findings to other fields therefore merits further attention.

Our results shed light on an heretofore neglected causal process underlying the growth of scientific knowledge, but they should be interpreted with caution. Although we measure the impact of losing a star collaborator, a full accounting of knowledge spillovers would require information on the benefits that accrued to the field while the star was alive. We can think of no experiment, natural or otherwise, that would encapsulate this counterfactual. Moreover, the benefits of exposure to star talent constitute only part of a proper welfare calculation. Scientific coauthorships also entail costs. These costs could be borne by low-status collaborators

in the form of lower wages, or by the stars, who might divert some of their efforts toward mentorship activities. Though some of these costs might be offset by nonpecuniary benefits, we suspect that the spillovers documented here are not fully internalized by the scientific labor market.

Finally, for every invisible college that contracts following superstar extinction, another might expand to slowly take its place. Viewed in this light, our work does little more than provide empirical support for Max Planck's famous quip: "science advances one funeral at a time."

APPENDIX I: CRITERIA FOR DELINEATING THE SET OF 10,349 "SUPERSTARS"

We present additional details regarding the criteria used to construct the sample of 10,349 superstars.

Highly funded scientists. Our first data source is the Consolidated Grant/Applicant File (CGAF) from the U.S. National Institutes of Health (NIH). This data set records information about grants awarded to extramural researchers funded by the NIH since 1938. Using the CGAF and focusing only on direct costs associated with research grants, we compute individual cumulative totals for the decades 1977–1986, 1987–1996, and 1997–2006, deflating the earlier years by the Biomedical Research Producer Price Index.¹⁸ We also recompute these totals excluding large center grants that usually fund groups of investigators (M01 and P01 grants). Scientists whose totals lie in the top ventile (i.e., above the 95th percentile) of either distribution constitute our first group of superstars. In this group, the least well-funded investigator garnered \$10.5 million in career NIH funding, and the most well-funded \$462.6 million.¹⁹

Highly cited scientists. Despite the preeminent role of the NIH in the funding of public biomedical research, the above indicator of "superstardom" biases the sample toward scientists conducting relatively expensive research. We complement this first group with a second composed of highly cited scientists identified by the

18. <http://officeofbudget.od.nih.gov/UI/GDPFFromGenBudget.htm>.

19. We perform a similar exercise for scientists employed by the intramural campus of the NIH. These scientists are not eligible to receive extramural funds, but the NIH keeps records of the number of "internal projects" each intramural scientist leads. We include in the elite sample the top ventile of intramural scientists according to this metric.

Institute for Scientific Information. A Highly Cited listing means that an individual was among the 250 most cited researchers for their published articles between 1981 and 1999, within a broad scientific field.²⁰

Top patenters. We add to these groups academic life scientists who belong in the top percentile of the patent distribution among academics—those who were granted seventeen patents or more between 1976 and 2004.

Members of the National Academy of Sciences. We add to these groups academic life scientists who were elected to the National Academy of Science between 1975 and 2007.

MERIT Awardees of the NIH. Initiated in the mid-1980s, the MERIT Award program extends funding for up to five years (but typically three years) to a select number of NIH-funded investigators “who have demonstrated superior competence, outstanding productivity during their previous research endeavors and are leaders in their field with paradigm-shifting ideas.” The specific details governing selection vary across the component institutes of the NIH, but the essential feature of the program is that only researchers holding an R01 grant in its second or later cycle are eligible. Further, the application must be scored in the top percentile in a given funding cycle.

Former and current Howard Hughes medical investigators. Every three years, the Howard Hughes Medical Institute selects a small cohort of mid-career biomedical scientists with the potential to revolutionize their respective subfields. Once selected, HHMIs continue to be based at their institutions, typically leading a research group of 10 to 25 students, postdoctoral associates, and technicians. Their appointment is reviewed every five years, based solely on their most important contributions during the cycle.²¹

Early career prize winners. We also included winners of the Pew, Searle, Beckman, Rita Allen, and Packard scholarships for the years 1981 through 2000. Every year, these charitable foundations provide seed funding to between twenty and forty young academic life scientists. These scholarships are the most prestigious

20. The relevant scientific fields in the life sciences are microbiology, biochemistry, psychiatry/psychology, neuroscience, molecular biology and genetics, immunology, pharmacology, and clinical medicine.

21. See Azoulay, Zivin, and Manso (2009) for more details and an evaluation of this program.

accolades that young researchers can receive in the first two years of their careers as independent investigators.

APPENDIX II: CONSTRUCTION OF THE CONTROL GROUP

We detail the procedure implemented to identify the control collaborators that help pin down the life-cycle and secular time effects in our DD specification. Because it did not prove possible to perfectly match treated and control collaborators on all covariates, the procedure is guided by the need to guard against two specific threats to identification.

First, collaborators observed in periods before the death of their associated superstar are more likely to work with a younger superstar; thus, they are not ideal as a control if research trends of collaborators differ by the age of the superstar. Collaborators observed in periods after the death of their associated superstar are only appropriate controls if the death of their superstar only affected the level of their output; if the death also negatively affected the trend, fixed effects estimates will be biased toward zero.

Second, fixed effects estimates might be misleading if collaborations with superstars are subject to idiosyncratic dynamic patterns. Happenstance might yield a sample of stars clustered in decaying scientific fields. More plausibly, collaborations might be subject to specific life-cycle patterns, with their productive potential first increasing over time, eventually peaking, and thereafter slowly declining. Relying solely on collaborators treated earlier or later as an implicit control group entails that this dyad-specific, time-varying omitted variable will not be fully captured by collaborator age controls.

To address these threats, the sample of control collaborators (to be recruited from the universe of collaborators for the 10,000 stars who do not die prematurely, regardless of cause) should be constructed such that the following four conditions are met:

1. treated collaborators exhibit no differential output trends relative to control collaborators up to the time of superstar death;
2. the distributions of career age at the time of death are similar for treated and controls;
3. the time paths of output for treated and controls are similar up to the time of death;
4. the dynamics of collaboration up to the time of death—number of coauthorships, time elapsed since first/last

coauthorship, superstar's scientific standing as proxied by his cumulative citation count—are similar for treated and controls.

Coarsened exact matching. To meet these goals, we have implemented a “Coarsened Exact Matching” (CEM) procedure (Iacus, King, and Porro 2008) to identify a control for each treated collaborator. As opposed to methods that rely on the estimation of a propensity score, CEM is a nonparametric procedure. This seems appropriate given that we observe no covariates that predict the risk of being associated with a superstar scientist who dies in a particular year.

The first step is to select a relatively small set of covariates on which the analyst wants to guarantee balance. In our example, this choice entails judgment but is strongly guided by the threats to identification mentioned above. The second step is to create a large number of strata to cover the entire support of the joint distribution of the covariates selected in the previous step. In a third step, each observation is allocated to a unique stratum, and for each observation in the treated group, a control observation is selected from the same stratum; if there are multiple choices possible, ties are broken randomly.

The procedure is coarse because we do not attempt to precisely match on covariate values; rather, we coarsen the support of the joint distribution of the covariates into a finite number of strata, and we match a treated observation if and only if a control observation can be recruited from this strata. An important advantage of CEM is that the analyst can guarantee the degree of covariate balance *ex ante*, but this comes at a cost: the more fine-grained the partition of the support for the joint distribution (i.e., the larger the number of strata), the larger the number of unmatched treated observations.

Implementation. We identify controls based on the following set of covariates (t denotes the year of death): collaborator's degree year, number of coauthorships with the star at t , number of years elapsed since last coauthorship with the star at t , JIF-weighted publication flow in year t , cumulative stock of JIF-weighted publications up to year $t - 1$, and the star's cumulative citation count at t . We then coarsen the joint distributions of these covariates by creating 51,200 strata. The distribution of degree years is coarsened using three year intervals; the distribution of coauthorship intensity is coarsened to map into our taxonomy of casual (one

and two coauthors), regular (between three and nine coauthors), and close collaborators (ten or more coauthors); the distribution of coauthorship recency is coarsened into quartiles (the first quartile corresponds to recent relationships, i.e., less than three years since the last coauthorship); the flow of publications in the year of death is coarsened into five strata (the three bottom quartiles; from the 75th to the 95th percentile, and above the 95th percentile); the stock of publications is coarsened into eleven strata (0 to 5th; 5th to 10th; 10th to 25th; 25th to 35th; 35th to 50th; 50th to 65th; 65th to 75th; 75th to 90th; 90th to 95th; 95th to 99th; and above the 99th percentile); and the distribution of citation count for the star is coarsened into quartiles.

We implement the CEM procedure year by year, without replacement. Specifically, in year t , we

1. eliminate from the set of potential controls all superstars who die, all coauthors of superstars who die, and all control coauthors identified for years of death k , $1979 \leq k < t$;
2. create the strata (the values for the cutoff points will vary from year to year for some of the covariates mentioned above);
3. identify within strata a control for each treated unit and break ties at random;
4. repeat these steps for year $t + 1$.

We match 5,064 of 5,267 treated collaborators (96.15%). In the sample of 5,064 treated + 5,064 controls = 10,128 collaborators, there is indeed no evidence of preexisting trends in output (Figure A.1); nor is there evidence of differential age effects in the years leading up to the death event (Figure A.2). As seen in Table II, treated and controls are very well-balanced on all covariates that pertain to the dynamics of the collaboration: number of coauthorships, time since last and first coauthored publication, and superstar's citation count. The age distributions are very similar as well. Furthermore, the CEM procedure balances a number of covariates that were not used as inputs, such as normalized keyword overlap and R01 NIH grantee status. For some covariates, we can detect statistically significant mean differences, though they do not appear to be substantively meaningful (e.g., 7% of controls vs. 8.4% of treated collaborators were former trainees of their associated superstars).

Sensitivity analyses. The analyst's judgment matters for the outcome of the CEM procedure insofar as she must draw a list

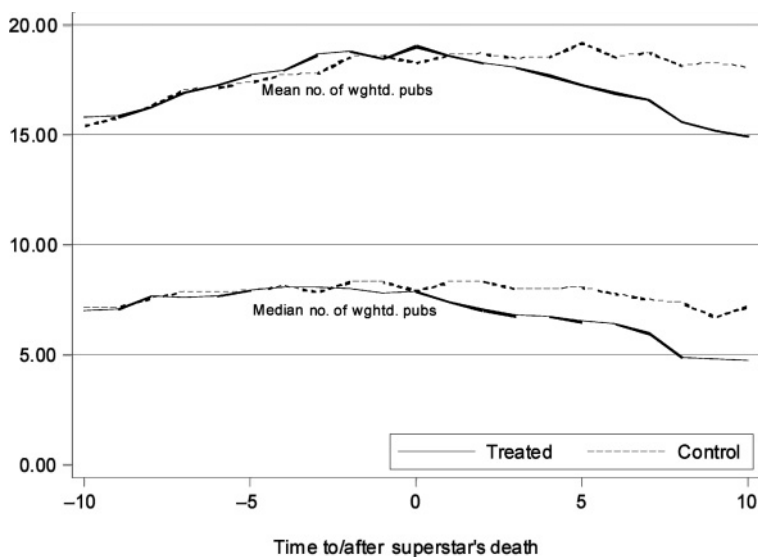


FIGURE A.1
Publication Trends for Treated and Control Collaborators

of “reasonable” covariates to match on, as well as decide on the degree of coarsening to impose. Therefore, it is reasonable to ask whether seemingly small changes in the details have consequences for how one should interpret our results.

Nonparametric matching procedures such as CEM are prone to a version of the “curse of dimensionality” whereby the proportion of matched units decreases rapidly with the number of strata. For instance, requiring a match on an additional indicator variable (i.e., doubling the number of strata from around 50,000 to 100,000) results in a match rate of about 70%, which seems unacceptably low. Conversely, focusing solely on degree age and the flow and stock of the outcome variables would enable us to achieve pairwise balance (as opposed to global balance, which ignores the one-to-one nature of the matching procedure) on this narrower set of covariates, but at the cost of large differences in the features of collaborations (such as recency and intensity) between treated and controls. This would result in a control sample that could address the first threat to identification mentioned above, but not the second.

However, we have verified that slight variations in the details of the implementation (e.g., varying slightly the number of cutoff points for the stock of publications; focusing on collaboration

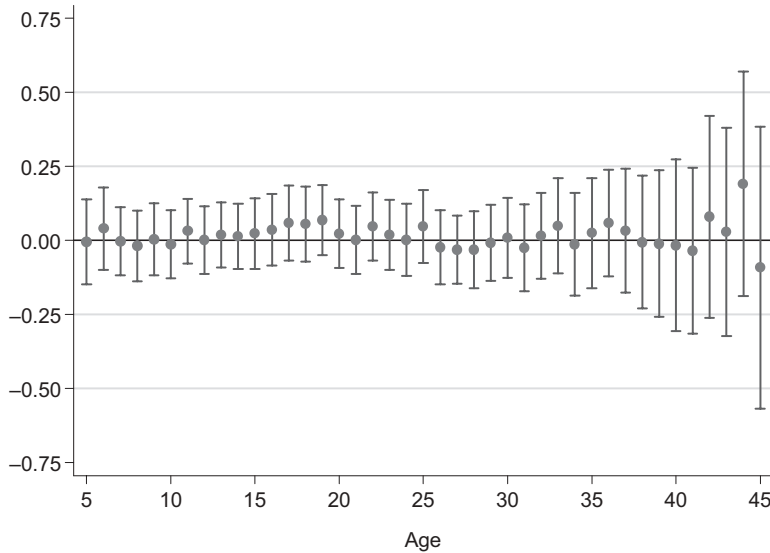


FIGURE A.2

Differential Age Trends for Treated vs. Control Collaborators

Each dot corresponds to the coefficient estimate for the interaction between an age indicator variable and treatment status in a Poisson regression of weighted publications onto a full suite of year effects, a full suite of age effects, and age by treatment status interaction terms. The population includes control and treated collaborators, but the estimation sample is limited to the years *before* the death of the associated superstar. The vertical brackets denote the 95% confidence interval (corresponding to robust standard errors, clustered around collaborators) around these estimates.

age as opposed to collaboration recency; or matching on superstar funding as opposed to superstar citations) have little impact on the results presented in Table III. To conclude, we feel that CEM enables us to identify a population of control collaborators appropriate to guard against the specific threats to identification mentioned in Section II.C.

MIT SLOAN SCHOOL OF MANAGEMENT AND NATIONAL BUREAU OF ECONOMIC RESEARCH
 UNIVERSITY OF CALIFORNIA, SAN DIEGO, AND NATIONAL BUREAU OF ECONOMIC RESEARCH
 MIT SLOAN SCHOOL OF MANAGEMENT

REFERENCES

Agrawal, Ajay K., and Avi Goldfarb, "Restructuring Research: Communication Costs and the Democratization of University Innovation," *American Economic Review*, 98 (2008), 1578–1590.

- Aizenman, Joshua, and Kenneth M. Kletzer, "The Life Cycle of Scholars and Papers in Economics—The 'Citation Death Tax,'" NBER Working Paper No. 13891, 2008.
- Azoulay, Pierre, Christopher Liu, and Toby Stuart, "Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs," MIT Sloan School Working Paper, 2009.
- Azoulay, Pierre, Joshua Graff Zivin, and Gustavo Manso, "Incentives and Creativity: Evidence from the Academic Life Science," NBER Working Paper No. 15466, 2009.
- Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Preferences and the Response to Incentives: Evidence from Personnel Data," *Quarterly Journal of Economics*, 120 (2005), 917–962.
- Bennedsen, Morten, Francisco Pérez-González, and Daniel Wolfenzon, "Do CEOs Matter?" New York University Working Paper, 2008.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119 (2004), 249–275.
- Cech, Thomas R., "Fostering Innovation and Discovery in Biomedical Research," *Journal of the American Medical Association*, 294 (2005), 1390–1393.
- Cockburn, Iain M., and Rebecca M. Henderson, "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery," *Journal of Industrial Economics*, 46 (1998), 157–182.
- Cole, Jonathan R., and Stephen Cole, "The Ortega Hypothesis," *Science*, 178 (1972), 368–375.
- Costa, Dora L., and Matthew E. Kahn, "Cowards and Heroes: Group Loyalty in the American Civil War," *Quarterly Journal of Economics*, 118 (2003), 519–548.
- Crane, Diana, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (Chicago: University of Chicago Press, 1972).
- de Solla Price, Derek J., *Little Science, Big Science* (New York: Columbia University Press, 1963).
- de Solla Price, Derek J., and Donald D. Beaver, "Collaboration in an Invisible College," *American Psychologist*, 21 (1966), 1011–1018.
- Fafchamps, Marcel, Sanjeev Goyal, and Marco van de Leij, "Matching and Network Effects," University of Oxford Working Paper, 2008.
- Gouriéroux, Christian, Alain Montfort, and Alain Trognon, "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 53 (1984), 701–720.
- Griffith, Rachel, Sokbae Lee, and John Van Reenen, "Is Distance Dying at Last? Falling Home Bias in Fixed Effects Models of Patent Citations," NBER Working Paper No. 13338, 2007.
- Grogger, Jeffrey, "The Effect of Arrests on the Employment and Earnings of Young Men," *Quarterly Journal of Economics*, 110 (1995), 51–71.
- Hall, Bronwyn H., Jacques Mairesse, and Laure Turner, "Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists," *Economics of Innovation and New Technology*, 16 (2007), 159–177.
- Ham, John C., and Bruce A. Weinberg, "Geography and Innovation: Evidence from Nobel Laureates," Ohio State University Working Paper, 2008.
- Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches, "Econometric Models for Count Data with an Application to the Patents–R&D Relationship," *Econometrica*, 52 (1984), 909–938.
- Henderson, Rebecca, Luigi Orsenigo, and Gary P. Pisano, "The Pharmaceutical Industry and the Revolution in Molecular Biology: Interactions Among Scientific, Institutional, and Organizational Change," in *Sources of Industrial Leadership*, David C. Mowery, and Richard R. Nelson, eds. (New York: Cambridge University Press, 1999).
- Iacus, Stefano M., Gary King, and Giuseppe Porro, "Matching for Causal Inference without Balance Checking," Harvard University Working Paper, 2008.
- Jones, Benjamin F., "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?" *Review of Economic Studies*, 76 (2009), 283–317.

- Jones, Benjamin F., and Benjamin A. Olken, "Do Leaders Matter? National Leadership and Growth since World War II," *Quarterly Journal of Economics*, 120 (2005), 835–864.
- Kim, E. Han, Adair Morse, and Luigi Zingales, "Are Elite Universities Losing Their Competitive Edge?" *Journal of Financial Economics*, 93 (2009), 353–381.
- Levin, Sharon G., and Paula E. Stephan, "Research Productivity over the Life Cycle: Evidence for Academic Scientists," *American Economic Review*, 81 (1991), 114–132.
- Lotka, Alfred J., "The Frequency Distribution of Scientific Productivity," *Journal of the Washington Academy of Sciences*, 16 (1926), 317–323.
- Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22 (1988), 3–42.
- Mairesse, Jacques, and Laure Turner, "Measurement and Explanation of the Intensity of Co-publication in Scientific Research: An Analysis at the Laboratory Level," NBER Working Paper No. 11172, 2005.
- Marshall, Alfred, *Principles of Economics* (New York: MacMillan, 1890).
- Mas, Alexandre, and Enrico Moretti, "Peers at Work," *American Economic Review*, 99 (2009), 112–145.
- Oettl, Alexander, "Productivity, Helpfulness and the Performance of Peers: Exploring the Implications of a New Taxonomy for Star Scientists," University of Toronto Working Paper, 2008.
- Reber, Sarah J., "Court-Ordered Desegregation," *Journal of Human Resources*, 40 (2005), 559–590.
- Reese, Thomas S., "My Collaboration with John Heuser," *European Journal of Cell Biology*, 83 (2004), 243–244.
- Romer, Paul M., "Endogenous Technological Change," *Journal of Political Economy*, 98 (1990), S71–S102.
- Rosenblat, Tanya S., and Markus M. Möbius, "Getting Closer or Drifting Apart?" *Quarterly Journal of Economics*, 119 (2004), 971–1009.
- Santos Silva, J.M.C., and Silvanna Tenreiro, "The Log of Gravity," *Review of Economics and Statistics*, 88 (2006), 641–658.
- Waldinger, Fabian, "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany," London School of Economics Working Paper, 2008.
- Weitzman, Martin L., "Recombinant Growth," *Quarterly Journal of Economics*, 113 (1998), 331–360.
- Wooldridge, Jeffrey M., "Quasi-likelihood Methods for Count Data," in *Handbook of Applied Econometrics*, M. Hashem Pesaran and Peter Schmidt, eds. (Oxford, UK: Blackwell, 1997).
- Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi, "The Increasing Dominance of Teams in Production of Knowledge," *Science*, 316 (2007), 1036–1039.
- Zucker, Lynne G., and Michael R. Darby, "Defacto and Deeded Intellectual Property Rights," NBER Working Paper No. 14544, 2008.
- Zucker, Lynne G., Michael R. Darby, and Marilyn B. Brewer, "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88 (1998), 290–306.