NBER WORKING PAPER SERIES

SUPERSTAR EXTINCTION

Pierre Azoulay
Joshua S. Graff Zivin
Jialan Wang

Working Paper 14577
http://www.nber.org/papers/w14577

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2008

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 (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 N01-OD-3-1015. The usual disclaimer applies. Please send correspondence to pazoulay@mit.edu.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2008 by Pierre Azoulay, Joshua S. Graff Zivin, and Jialan Wang. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Superstar Extinction
Pierre Azoulay, Joshua S. Graff Zivin, and Jialan Wang
NBER Working Paper No. 14577
December 2008, Revised July 2009
JEL No. O3,O31,O43

ABSTRACT

We estimate the magnitude of spillovers generated by 112 academic "superstars" who died pre- maturely 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."

Pierre Azoulay
MIT Sloan School of Management
50 Memorial Drive E522-555
Cambridge, MA 02142
and NBER
pazoulay@mit.edu

Jialan Wang
60 Wadsworth St Apt 7E
Cambridge, MA 02142
jialanw@mit.edu

Joshua S. Graff Zivin
International Relations & Pacific Studies
University of California, San Diego
9500 Gilman Drive, MC 0519
La Jolla, CA 92093-0519
and NBER
jgraffzivin@ucsd.edu

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

1 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 et al. 2007). The growth of scientific collaboration has important implications for the optimal allocation of public R&D funds, the apportionment of credit amongst scientists, the formation of scientific reputations, and ultimately 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 largely due to the empirical challenges inherent to 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 et al. 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]

[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 et al. 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

Specifically, we analyze changes in the research output of collaborators for 112 eminent life scientists who die 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 dataset 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]

The study's focus on the scientific elite can be justified both on 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

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 die 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 et al. (2009).

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? Since 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 which 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.

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 3 provides descriptive statistics at the coauthor and dyad level. We report the results in section 4. Section 5 concludes.

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

2 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 US GDP. Much biomedical innovation is science-based (Henderson et al. 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 et al. 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.

Lastly, 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).

2.1 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, since 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 back these scientists' careers from the time they obtain their first position as independent investigators (typically after a postdoctoral fellowship) until 2006. We do so through a combination of curriculum vitæs, 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) rather than 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 intervene between 1979 and 2003. This will enable us to observe at least 4 years' (resp. 3 years') worth of scientific output for every colleague before (resp. after) the death

of their superstar collaborator. Second, they must be 67 years of age or less at the time of their passing (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 pre-retirement phase of their career prior to the time of death. This constraint is somewhat subjective, but we validate in the on-line 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 that it might seem, since the vast majority of obituaries mention the cause of death explicitly.[4] After eliminating 136 scientists whose death 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 on-line 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 garnered 8,190 citations as of early 2008.

2.2 The Universe of Potential Colleagues

Information about the superstars' colleagues stems from the Faculty Roster of the Association of American Medical Colleges, 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 dataset. For each of the 230,000 faculty members that appear in the roster, we know the full name, the type of degrees

[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 10 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.

[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, while many medical schools update once a semester.

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 post-doctoral 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, University of California at Berkeley, Rockefeller University, the Salk Institute, or the Bethesda campus of the NIH; and it also ignores faculty members in Arts and Sciences departments — such as biology and chemistry — if they do not hold joint appointments at a local medical school.[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 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

[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 this data with third-parties. However, it can be licensed from AAMC, provided a local IRB gives its approval and a confidentiality agreement protects the anonymity of individual faculty members.

validated coauthor, and generates publication counts as well as coauthorship variables for each superstar/colleague dyad, in each year. In the on-line 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.

2.3 Identification Strategy

A natural starting point to identify the effect of superstar death is to examine changes in collaborator research output after the superstar passes away, relative to when s/he was still alive, using a simple collaborator fixed effects specification. Since the extinction effect is mechanically correlated with the passage of time, as well as with 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 superstar died in earlier periods, or will die in future periods. Despite its long pedigree in applied economics (e.g., Grogger 1995; Reber 2006), this approach may be problematic in our setting.

First, coauthors observed in periods after the death of their associated superstar 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 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 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 2.1). Combining the treated and control

samples enables us to estimate the effect of superstar extinction in a difference in differences 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 controls.

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 A1 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, since 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 non-parametric, and provides early evidence that the loss of a superstar coauthor leads to a decrease in collaborators' publication output. Furthermore, the magnitude 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 non-coauthors 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 magnitudes, of our main results.

3 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 year of the (possibly counterfactual) year of death for the superstar. We distinguish between variables that are inherently dyadic (e.g., co-location 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).

Dyadic variables. Of immediate interest is the distribution of coauthorship intensity at the dyad level. While 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 75^{th} and the 95^{th} percentile of coauthorship intensity, while "close" dyads correspond to those above the 95^{th} percentile.

We focus next on collaboration age and recency. On average, collaborations begin 11 years before the star's death, and time since last coauthorship is slightly more than 9 years. In other words, most of the collaborations in the sample do not involve active research projects at the time of death. Recent collaborations (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

[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.

(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. While 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.

We now examine the spatial distribution of collaborations. Slightly more than 12% of collaborations correspond to scientists who were co-located 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 which overlap in their non-coauthored 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. 7.5% of the dyads in

[10]The purported training period runs from 3 years before graduation to 4 years after graduation for PhDs and MD/PhDs; and from the year of graduation to 6 years after graduation for MDs. Recall that we do not observe the population of former trainees, but only those trainees that 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 eliminated all former trainees from the sample with career age less than 5 at the time of death.

[11]Further details on its construction are provided in the on-line appendix, Section II.

the sample correspond to this situation of "accidental coauthorship" — the most tenuous form of collaboration.

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 MDs, 40% are PhDs, and the remainder are MD/PhDs; and a third are affiliated with basic science departments (as opposed to clinical or public health departments). The coauthors are about 8 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 vs. 140 articles, on 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. 55% of collaborators had served as PI on at least one NIH R01 grant when the superstar passes away, while 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 dataset with 153,508 collaborator×year observations (treated collaborators only) or 294,943 collaborator×year observations (treated and control collaborators).

4 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 post-death indicator variable below with various attributes of the superstar, colleague, and dyad.

4.1 Econometric Considerations

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

$$E\left[y_{jt}|X_{ijt}\right] = exp\left[\beta_0 + \beta_1 AFTER_DEATH_{it} + f(AGE_{jt}) + \delta_t + \gamma_{ij}\right] \quad (1)$$

where y is a measure of research output, $AFTER\ DEATH$ denotes an indicator variable that switches to one the year after the superstar dies, $f(AGE_{jt})$ 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 post-tenure 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 (MD or PhD).[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 non-negative. For example, 24.80% of the 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.

[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 that have the same (career) age but earned their degrees in different years (Hall et al. 2007).

Following a long-standing tradition in the study of scientific and technical change, we present conditional quasi-maximum likelihood estimates based on the fixed-effect Poisson model developed by Hausman et al. (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 et al. 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 non-negative dependent variables, whether integer or continuous (Santos Silva and Tenreyro 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 et al. (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. Since SC/Gen matches the entire authorship roster for each article, we can separate those publications coauthored with the superstar from those produced independently 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 on-line appendix (section V), we present additional results based on article-level citation outcomes.

4.2 Main effect of superstar extinction

Table III presents our core results. Column 1a examines the determinants of the 5,267 treated coauthors' JIF-weighted publication output. We find a sizable and significant 8.8% decrease

[13] In the on-line appendix (section IV), we show that OLS yields very similar results to QML Poisson estimation for our main findings.

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 only slightly our estimate of the treatment effect, 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 towards 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.

4.3 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 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 suggest 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 (at least one joint publication in the three years preceding the star's death, column 2a) 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 towards 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 independently 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 coau-

thorship see their independent output decline respectively by 5.54% (column 1b) and 8.25% (column 2b). Very similar results are obtained when combining all these covariates into one specification (columns 3a and 3b).

While 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 fully compensate 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.

4.4 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 5 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 that would otherwise remain unaware of each other's expertise. We do so by 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 labs, 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, since mentorship can continue into the early faculty career and be extremely

[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.

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 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 co-located 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 et al. 2006; Griffith et al. 2007; Waldinger 2008). However, our estimate of the co-location 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 et al. 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 intellec-

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

tual 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 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? Since 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, since 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

[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 five coefficients are equal to one another.

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, while 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' post-extinction 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 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 non-collaborators.

4.5 Robustness and Sensitivity Checks

The on-line 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 twenty five 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

[17]Table VII eliminates from the estimation sample the collaborators of 11 superstars who are NIH intramural scientists, and as such not eligible for extramural NIH funding.

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.

5 Conclusion

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 are 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 contributed dies along with him, 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 which 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 non-coauthors 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. While 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 towards mentorship activities. Though some of these costs might be offset by non-pecuniary 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."*

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
MD	0.420	0	0.496	0	1
PhD	0.438	0	0.498	0	1
MD/PhD	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
Nb. of Collaborators	47.027	37	34.716	3	178
NIH Review Panel Membership (past 5 yrs)	0.045	0	0.207	0	1
Nb. of Collabs. in NIH Review Panels (past 5 yrs)	1.330	1	1.657	0	7
Career Nb. of Publications	139.607	121	91.371	25	473
Career Nb. 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 II.A 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 5 years prior to the star's death. We use the terms "star" and "superstar" interchangeably.

Table II
Summary Statistics for Collaborators in the Year of Superstar Death

		Mean	Median	Std. Dev.	Min.	Max.
Control Collabs. (N=5,064)	Nb. of weighted Publications	18.314	8	27.917	0	342
	Cum. Nb. of weighted Publications	327.330	187	409.098	0	3,968
	Holds R01 grant	0.559	1	0.497	0	1
	Co-Located	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. Nb. of Coauthorships	2.734	1	4.339	1	69
	Nb. of Other Superstar Collabs.	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
Treated Collabs. (N=5,064)	Nb. of weighted Publications	19.068	8	31.656	0	491
	Cum. Nb. of weighted Publications	334.905	187	436.927	0	4,519
	Holds R01 grant	0.571	1	0.495	0	1
	Co-Located	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. Nb. of Coauthorships	2.835	1	4.894	1	75
	Nb. of Other Superstar Collabs.	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 extinct superstar life scientists 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.

Table III
Impact of Superstar Death on Collaborators' Publication Rates

	Panel A: All JIF-Weighted Publications		Panel B: JIF-Weighted Pubs. Written with others	
	Without Ctrls (1a)	With Ctrls (1b)	Without Ctrls (2a)	With Ctrls (2b)
After Death	-0.092**	-0.086**	-0.057**	-0.054*
	(0.022)	(0.025)	(0.022)	(0.024)
Log Pseudo-Likelihood	-974,285	-1,832,594	-950,864	-1,783,958
Nb. of Obs.	153,508	294,943	153,508	294,943
Nb. 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 17 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-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.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table IV
Collaborator Publication Rates and Imperfect Skill Substitution

	Coauthorship Intensity		**Coauthorship Recency**		**Coauthorship Intensity & Recency**	
	All Pubs.	Pubs. written with others	All Pubs.	Pubs. written with others	All Pubs.	Pubs. written with others
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
After Death	-0.076**	-0.057*	-0.087**	-0.074**	-0.080**	-0.075**
	(0.026)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
After Death × Regular Collab.	-0.044	-0.020			-0.039	-0.018
	(0.041)	(0.042)			(0.042)	(0.043)
After Death × Close Collab.	-0.026	0.117			-0.014	0.119
	(0.068)	(0.073)			(0.069)	(0.074)
After Death × At least 1 coauthorship in the three years preceding star's death			-0.022	0.032	-0.021	0.028
			(0.038)	(0.039)	(0.039)	(0.039)
Log Pseudo-Likelihood	-1,831,987	-1,781,742	-1,822,664	-1,775,680	-1,821,791	-1,774,167
Nb. of Obs.	294,943	294,943	294,943	294,943	294,943	294,943
Nb. of Collabs.	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 3 and 9 coauthored pubs.; close collaborations correspond to 10 or more coauthored pubs; casual collaborations — the omitted category — corresponds to 1 or 2 coauthored pubs.). All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 interaction terms between the age effects and each covariate of interest (i.e., column (3b) includes a total of 3×17=51 age-specific interaction terms). Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table V
Collaborator Publication Rates and Access to Resources

	Star's Ties to NIH Funding Process	**Quartile of Betweenness Centrality**	**Former Trainee**	**All Covariates Combined**
	(1)	(2)	(3)	(4)
After Death	-0.105**	-0.067*	-0.086**	-0.089*
	(0.037)	(0.028)	(0.025)	(0.035)
After Death × Star Sat on NIH Review Panel	0.042			0.024
	(0.064)			(0.070)
After Death × Star's Nb. of Coauth. Ties to NIH Review Panelists	0.011			0.014
	(0.013)			(0.015)
After Death × Star in 4th Quartile of Betweenness Centrality		-0.031		-0.040
		(0.046)		(0.051)
After Death × Coauthor is Former Trainee			0.056	0.048
			(0.069)	(0.069)
Log Pseudo-Likelihood	-1,831,339	-1,831,779	-1,830,582	-1,828,754
Nb. of Obs.	294,943	294,943	294,943	294,943
Nb. of Collabs.	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 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 interaction terms between the age effects and each covariate of interest. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table VI
Collaborator Publication Rates and Proximity in Geographic & Intellectual Space

	(1)	(2)	(3)	(4)	(5)
After Death	-0.092**	-0.067**	-0.094**	-0.081**	-0.074**
	(0.027)	(0.023)	(0.022)	(0.024)	(0.026)
After Death × Co-Located	0.042			0.037	0.042
	(0.043)			(0.043)	(0.044)
After Death × Kwd. Overlap in Top Quartile		-0.115*		-0.114†	-0.127*
		(0.059)		(0.059)	(0.057)
After Death × "Accidental" Collaborator			0.104†	0.111†	0.077
			(0.060)	(0.058)	(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 Collabs. 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
Nb. of Obs.	294,943	294,943	294,943	294,943	294,943
Nb. of Collabs.	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. Co-located 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 which appear on both the colleague and superstars non-joint 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 collaborations correspond to between 3 and 9 coauthored pubs.; close collaborations correspond to 10 or more coauthored pubs; casual collaborations — the omitted category — corresponds to 1 or 2 coauthored pubs.). All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 interaction terms between the age effects and the covariate of interest. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table VII
Impact of Superstar Status on Collaborators' Publication Rates

	Superstar Status Citations	**Superstar Status NIH Funding**	**Superstar Status Citations & NIH Funding**
	(1)	(2)	(3)
After Death	-0.034	-0.070*	-0.026
	(0.036)	(0.035)	(0.039)
After Death ×	-0.082†		-0.080†
Star in Top Quartile of Cites	(0.047)		(0.048)
After Death ×		-0.026	-0.016
Star in Top Quartile of NIH $		(0.050)	(0.051)
Log Pseudo-Likelihood	-1,715,929	-1,716,213	-1,715,916
Nb. of Obs.	275,776	275,776	275,776
Nb. of Collabs.	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 11 "intramural" NIH scientists who are not eligible to receive extramural funding. All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 (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.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Figure I
Distribution of Coauthorship Intensity

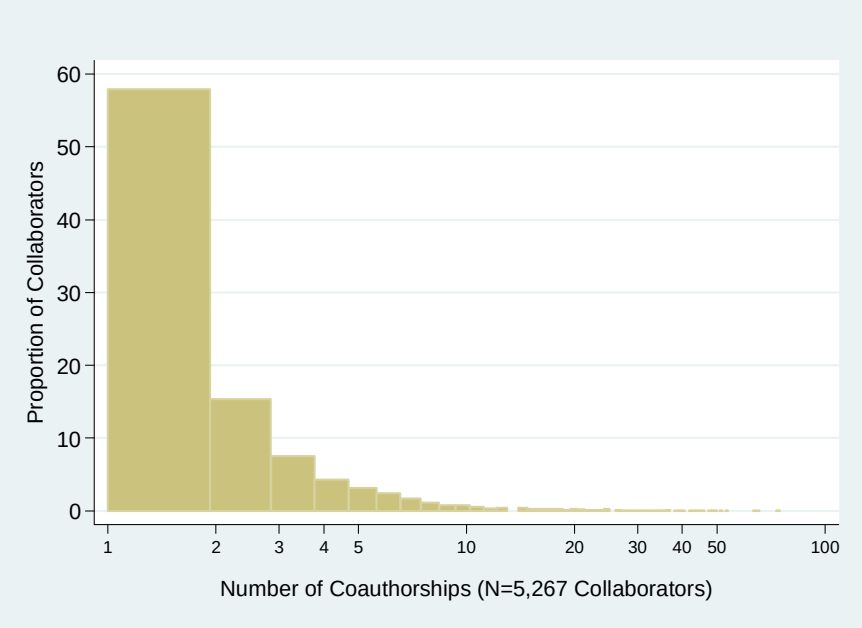
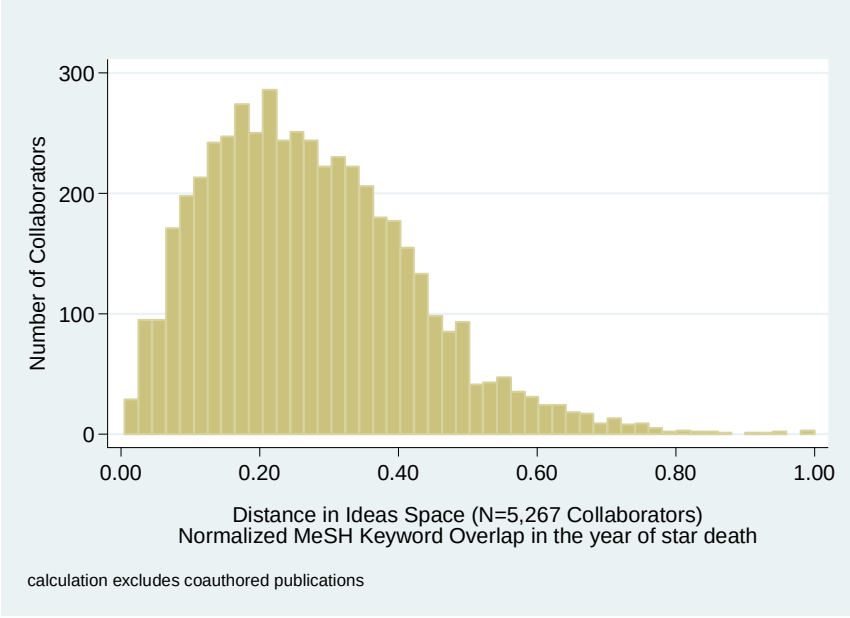


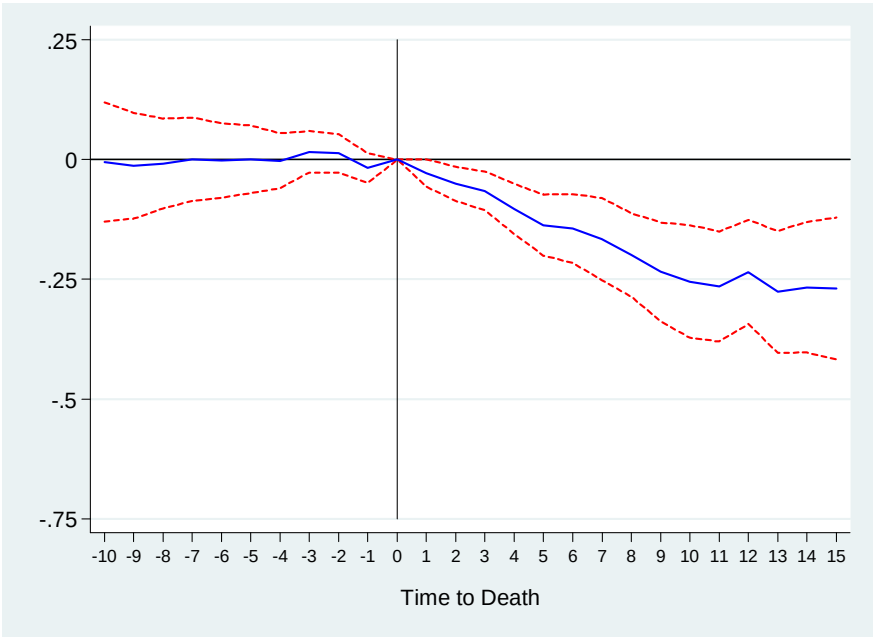
Figure II
Proximity in Ideas Space



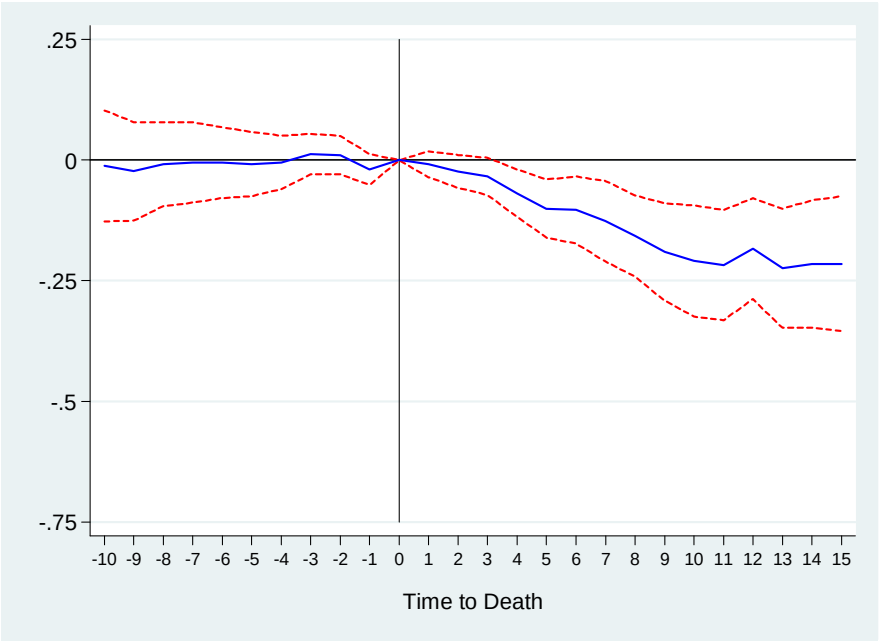
Notes: Measure of distance in ideas space is defined as the number of unique MeSH terms which overlap between the colleague's and superstar's publications (excluding coauthored output), normalized by the total number of MeSH terms used in the colleague's total publications. This measure is calculated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

Figure III
Dynamics of the Treatment Effect

A. All Publications



B. Publications without Superstar Collaborator



Notes: The solid blue 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, 17 indicator variables corresponding to different age brackets, and interactions of the treatment effect with 27 indicator variables corresponding to 11 years before the year of death and prior, 10 years before the year of death, 9 years before the year of death,..., 14 years after the year of death, and 15 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 red lines. Figure IIIA uses column (1b) of Table III as a baseline (i.e., treated and control collaborators, the dep. var. includes all of the collaborator's publications); Figure IIIB uses column (2b) of Table III as a baseline (i.e., treated and control collaborators, the dep. var. is limited to the collaborator's publications in which the superstar does not appear on the authorship list).

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 dataset 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 95^{th} 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 towards scientists conducting relatively expensive research. We complement this first group with a second composed of highly cited scientists identified by the 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 17 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 5 years (but typically 3 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 20 and 40 young academic life scientists. These scholarships are the most prestigious accolades that young researchers can receive in the first two years of their careers as independent investigators.

[18] `http://officeofbudget.od.nih.gov/UI/GDPFromGenBudget.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.

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

[21] See Azoulay et al. (2009) for more details and an evaluation of this program.

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 difference-in difference (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 towards 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 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 et al. 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 non-parametric 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 judgement, 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 strata, and for each observation in the treated group, a control observation is selected from the same strata; 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 higher 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 (1 and 2 coauth.), regular (between 3 and 9 coauth.), and close collaborators (10 or more coauth.); the distribution of coauthorship recency is coarsened into quartiles (the first quartile corresponds to recent relationships, i.e. less than 3 years since the last coauthorship); the flow of publications in the year of death is coarsened into 5 strata (the three bottom quartiles; from the 75^{th} to the 95^{th} percentile, and above the 95^{th} percentile); the stock of publications is coarsened into eleven strata (0 to 5^{th}; 5^{th} to 10^{th}; 10^{th} to 25^{th}; 25^{th} to 35^{th}; 35^{th} to 50^{th}; 50^{th} to 65^{th}; 65^{th} to 75^{th}; 75^{th} to 90^{th}; 90^{th} to 95^{th}; 95^{th} to 99^{th}; and above the 99^{th} 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 the some of the covariates mentioned above);
3. identify within strata a control for each treated unit; break ties at random;
4. repeat these steps for year $t+1$.

We match 5,064 out 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 prexisting trends in output (Figure A1); nor is there evidence of differential age effects in the years leading up to the death event (Figure A2). 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 judgement matters for the outcome of the CEM procedure insofar as she must draw a list 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.

Non-parametric 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 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 2.3.

Figure A1: Publication Trends for Treated and Control Collaborators

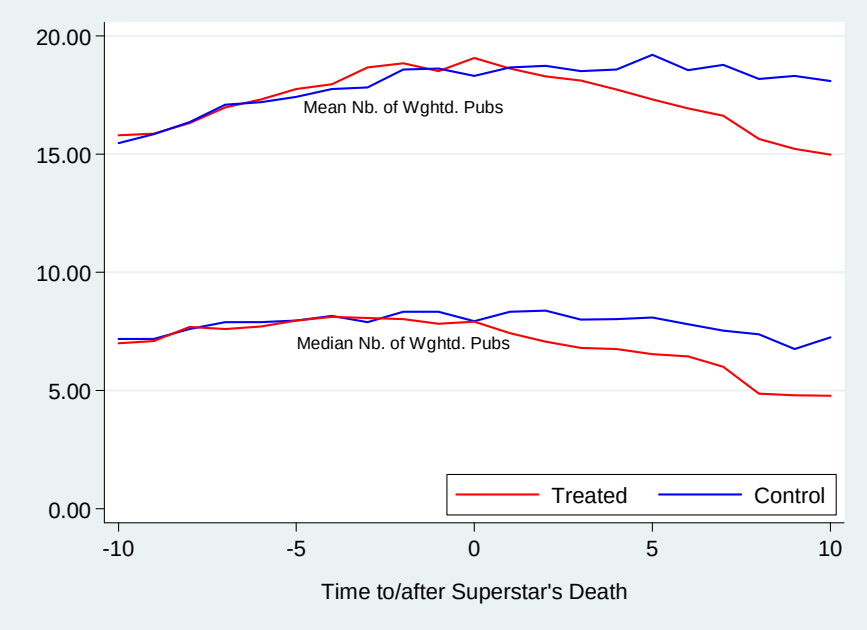
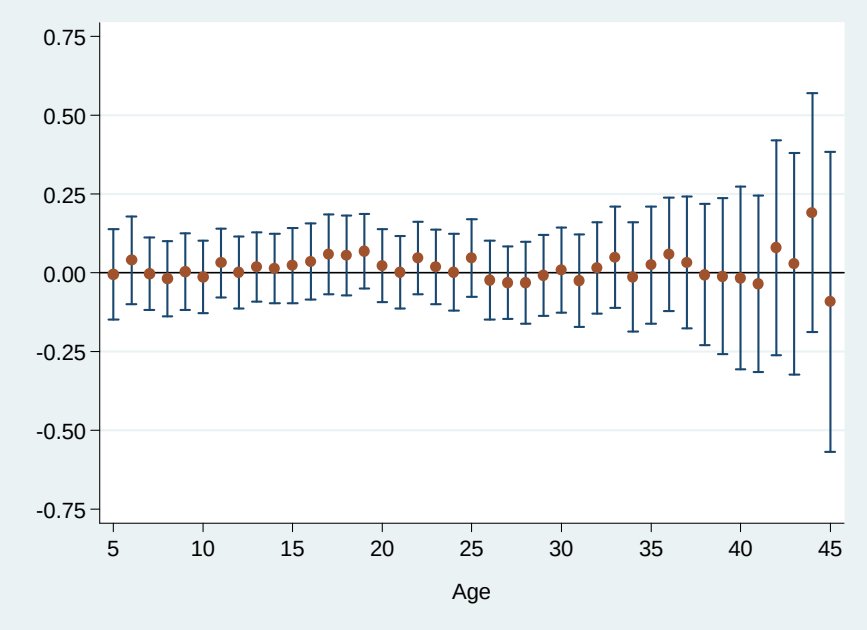


Figure A2: Differential Age-Trends for Treated vs. Control Collaborators



Note: Each dot cooresponds 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 limied to the years before the death of the associated superstar. The blue vertical brackets denote the 95% confidence interval (corresponding to robust standard errors, clustered around collaborators) around these estimates.

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 #13891 (2008).

Azoulay, Pierre, Christopher Liu, and Toby Stuart, "Social Influence Given (Partially) Deliberate Matching: Career Imprints in the Creation of Academic Entrepreneurs," Working Paper, MIT Sloan School (2009).

Azoulay, Pierre, Joshua Graff Zivin, and Gustavo Manso, "Incentives and Creativity: Evidence from the Academic Life Sciences," Working Paper, MIT Sloan School (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?," Working Paper, New York University (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, IL: University of Chicago Press, 1972).

de Solla Price, Derek J., *Little Science, Big Science* (New York, NY: 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," Working Paper, University of Oxford (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 #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," Working Paper, Ohio State University (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 David C. Mowery, and Richard R. Nelson, eds., *Sources of Industrial Leadership* (New York, NY: Cambridge University Press, 1999), pp. 267-311.

Iacus, Stefano M., Gary King, and Giuseppe Porro, "Matching for Causal Inference Without Balance Checking," Working Paper, Harvard University (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?," NBER Working Paper #12245 (2006).

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 #11172 (2005).

Marshall, Alfred, *Principles of Economics* (New York, NY: 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," Working Paper, University of Toronto (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 of Drifting Apart?," *Quarterly Journal of Economics*, 119 (2004), 971-1009.

Santos Silva, J.M.C., and Silvanna Tenreyro, "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," Working Paper, London School of Economics (2008).

Weitzman, Martin L., "Recombinant Growth," *Quarterly Journal of Economics*, 113 (1998), 331-360.

Wooldridge, Jeffrey M., "Quasi-Likelihood Methods for Count Data," in M. Hashem Pesaran, and Peter Schmidt, eds., *Handbook of Applied Econometrics* (Oxford: Blackwell, 1997), pp. 352-406.

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., Michael R. Darby, and Marilynn B. Brewer, "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88 (1998), 290-306.

Zucker, Lynne G., and Michael R. Darby, "Defacto and Deeded Intellectual Property Rights," NBER Working Paper #14544 (2008).

On-Line Appendix

I — Matching Superstars and Collaborators

We designed the Stars/Colleague Generator (S/CGen) to harvest coauthors' names from a superstar's bibliome. S/CGen identifies colleagues to the extent that (a) they coauthor at least once; and (b) they can be matched (based on a combination of a last name and up to two initials) with the AAMC Faculty Roster. We will describe the matching process using as an example one of our extinct superstars, Jeffrey M. Isner, MD. Isner, a pioneer of gene therapy for Peripheral Artery Diseases, and a faculty member at the Tufts University School of Medicine, died in 2001 from a heart attack, at the age of 54.

The matching process begins with the creation of a customized PubMED search query for each superstar. In the case of Isner, the query is `("isner jm"[au] OR "isner j"[au]) AND 1977:2006[dp]`, and it returns 373 original publications (the query also returns 24 letters, editorials, interviews, etc., which we ignore). The process of harvesting bibliomes from PubMED using name variations and queries as inputs is facilitated by the use of PubHarvester, a software program we specifically designed for this purpose (Azoulay et al. 2006).

Spurious Coauthors. Jeff Isner's PubMED query accounts for his inconsistent use of the middle initial, but is otherwise quite simple. For other scientists, queries might factor in their inconsistent use of the suffix "Jr.," or name variations coincident with changes in marital status. For yet many others with frequent names, the queries are more involved, and make use of CV information such as scientific keywords, institutional affiliation, frequent coauthors' names, etc. This is essential, since errors of commission will tend to generate spurious coauthor matches. We guarded against this source of error by devoting hundreds of person-hours to the design of accurate search queries for each of our 10,349 superstars. This degree of labor-intensive customization ensures that a superstar's bibliome excludes publications belonging to homonymous scientists.

Matching process. The second step is to extract the name of coauthors from the star's bibliome and to match them with the AAMC Faculty Roster. Unfortunately, PubMED does not record authors' full names, nor does it record their institutional affiliations; it only keeps track of authors by using a combination of last name, two initials, and a suffix (where the suffix and the second initial fields can be empty). The matching process is automated by SC/Gen, and its outcome in the case of a sample publication authored by Jeff Isner is illustrated in Figure W1. S/CGen cannot generate a match for each coauthor. Some coauthors are technicians or undergraduate students; others are graduate students or postdocs who do not go on to faculty positions; yet others are located in foreign institutions; others still publish under names that differ from the Faculty Roster listing (for instance by being inconsistent with the use of middle initials, suffixes, or hyphens). In total, SC/Gen generates 355 matches with the AAMC Faculty Roster for Isner.

Ambiguous Coauthors. Often, SC/Gen can match a given PubMED name with more than one faculty in the Roster. Notice the case of `ramaswamy k` on Figure W1. Does it correspond to K. Ramaswamy (University of Illinois–Chicago), to Karthik Ramaswamy (UMASS School of Medicine), or to Krishna Ramaswamy (Tufts University School of Medicine)? Several options are available to deal with these ambiguous matches. We could discard the first two matches, since the third one corresponds to an individual who shared Isner's institutional affiliation. Alternatively, we could retain all three matches, but assign each a weight of $\frac{1}{3}$, incorporating a guess on the probability that each match is genuine. Finally, we could simply discard all three matches, and focus instead on those matches that are unambiguous. This is the approach we have followed to generate the results we present in the paper.[1] Out of the 355 matches mentioned above, only

[1] Trajtenberg et al. (2006) propose algorithms to automate the process of name disambiguation in patent data. Adapting their approach to publication data lies far beyond the scope of this paper. To fix ideas, `Lechleiter JD` is an example of unique

177 correspond to coauthors with unambiguous PubMED names. For the set of 112 superstars, S/CGen identifies 5,267 distinct coauthors with unambiguous PubMED names — an average of 47 coauthors per superstar (the median is 37).

Coauthors' Publication Output. The publication output of coauthors with frequent names will be measured with error. This source of error is less worrisome, since it involves a dependent variable. Nonetheless, we have taken several steps to ascertain the extent to which it biases our results. First, our decision to eliminate from the sample coauthors with ambiguous PubMED names means that it is almost entirely composed of individuals with relatively rare names. Second, we have experimented with deleting from the estimation sample observations corresponding to coauthors with unique PubMED names, but popular last names.[2] Specifically, we dropped from the main analysis all coauthors whose last name appear 160 or more times in the roster (the 99^{th} percentile of the distribution of last name frequency, which correspond to names such as `Greenwald`, `McKee`, `O'Malley`, or `Fu`). This hardly affected the main results. Third, we limit the estimation sample to elite coauthors (i.e., coauthors who belong to the set of 10,349 "superstars"). Because we designed custom PubMED queries for these individuals, their output is measured with little (if any) error. The magnitude of the treatment effect is very similar to the one obtained on the full sample of coauthors.

II — Measuring Proximity in Ideas Space

We describe the construction of our variable to measure distance (or rather, proximity) in intellectual or "ideas space" between nodes in a dyad of scientists. The boundaries around scientific fields are difficult to delineate since most scientific research can be classified in numerous ways, and agreement among scientists regarding the categorization of specific bits of knowledge is often elusive. Our approach is predicated on the inadequacy of measures based on shared department affiliation, or on coarse distinctions between scientific fields (e.g., cell vs. molecular biology). Instead of attempting to position individual scientists relative to some fixed address in ideas space, we provide a method to cheaply and conveniently measure relative position in this space.

An essential input is provided by the Medical Subject Headings (MeSH) thesaurus, a controlled vocabulary produced by the National Library of Medicine whose explicit statement of purpose is to *"provide a reproducible partition of concepts relevant to biomedicine for the purpose of organizing knowledge and information."* The MeSH vocabulary consists of 24,767 terms arranged in a hierarchical structure, and these terms are used by NLM staff to tag all the articles indexed by the PubMED database.[3] From our standpoint, one of the MeSH system's most attractive feature is its fine-grained level of detail. For instance, the initial draft of the public human genome project (Lander et al. 2001) is tagged by 26 distinct descriptors, which run the gamut from the very general ("Humans", "RNA/Genetics") to the very specific ("Repetitive Sequences, Nucleic Acid", "CpG Islands", "DNA Transposable Elements").[4]

The procedure followed to generate our dyadic measure of intellectual proximity is best explained through a concrete example. We will focus on a two scientists, Andrew Schally (from Tulane University in New Orleans, LA) and Roger Guillemin (from the Salk Institute in San Diego, CA). Throughout the 1960s and 1970s, this pair of eminent neuro-endocrinologists was locked in a very public (and often acrimonious) rivalry

PubMED name. In contrast, `Weinstein SL` corresponds to two distinct faculty in the roster, `Miller MJ` to ten, and `Wang Y` to thirty six.

[2]For instance, `Miller CR` is a unique PubMED name, though `Miller` is the last name for 800 distinct individuals in the AAMC Faculty Roster.

[3]At the highest level of the hierarchical structure are very broad headings such as "Anatomy" or "Mental Disorders." More specific headings are found at lower levels of the eleven-level hierarchy, such as "Ankle" and "Conduct Disorder." See `http://www.nlm.nih.gov/mesh/` for more details.

[4]This stands in sharp contrast to the coarse partition of technological space provided by patent classes, which are often used in the study of *involuntary* knowledge spillovers (Benner and Waldfogel 2007).

whose ultimate goal was the synthesis of peptide hormones produced by the brain. Together with Rosalyn Yalow, the Nobel committee awarded them both the Prize in Medicine and Physiology in 1977 (details of this celebrated case of a scientific race can be found in Nicholas Wade's The Nobel Duel). We will focus on the five-year window that preceded the award of the Prize, i.e., 1973-1977. During this period, Guillemin and Schally did not collaborate at all, and according to Wade (1981), even actively sought to undermine each other's progress.

The calculation is illustrated in Table W2; it is automated by SciDist, an open-source software program we specifically designed for this purpose.[5] Between 1973 and 1977, Schally published 240 articles, and Guillemin "only" 60. We extract from these publications all MeSH terms, regardless of their position in the descriptor hierarchy. There are a total of 607 unique MeSH terms tagging the two scientists' publications, 147 of which overlap. Table W2 lists the Top 10 overlapping terms with highest and lowest combined use, respectively.[6]

To compute the proximity of Guillemin to Schally, we simply divide the number of overlapping MeSH terms (147), by the total number of unique MeSH terms tagging Guillemin's 60 publications (220). In contrast, the proximity of Schally to Guillemin is given by 147 divided by 534 (the total number of unique MeSH terms tagging Schally's 240 publications). We view this lack of symmetry as an attractive feature of our approach, since Schally's research agenda during this period was significantly broader, and in fact encompassed most of Guillemin's. In contrast, many of the distance concepts used to date in the literature — for example to position firms' research portfolio in technology space — use an Euclidean (hence symmetric) concept of distance (e.g., Jaffe 1986).

III — Pre-existing Trends in Output for the Superstars

In Table W3, we present results for specifications in which the superstars' quality-adjusted publication output is regressed onto a series of indicator variables corresponding to the timing of death: 5 years before the year of death, 4 years before the year of death, and so on, up until two years after the year of death (a scientist can, and often does, publish after his death because his/her coauthors will typically steward articles through the pipeline on his behalf). All models include superstar scientist fixed effects, and we use as a control group the set of superstars who collaborate with the sample of control collaborators. The inclusion of controls is important insofar as it enables us to pin down the effect of age and calendar time, which might be correlated with the death effect.

We use two definitions of the dependent variable. In the first (column 1), all of the stars' publications participate in the calculation of the JIF-weighted count; in the second (column 2), only the publications in which the star appears in last position on the authorship roster are considered (last author status is reserved to the heads of laboratory/research group in the life sciences). In both specifications, we find no evidence that the superstars' output trends downward even before their death. In fact, the coefficient estimates turn negative in sign only in the year that follows the year of death, and reach statistical significance only two years after the death. In light of these results, we feel confident that our informal screen for research activity yields a set of 112 extinct superstars still actively engaged in science at the time of their deaths.

[5] SciDist is available for download at `http://www.stellman-greene.com/ScientificDistance/`.

[6] An open question is whether one should weight each term by its frequency of use, or whether it is the number of unique terms that matters. In practice, these alternatives yield two measures of proximity that are heavily correlated, and the distinction does not affect the substance of our results.

IV — Main Results with OLS Estimation

In Table W4, we replicate the results in Table III using linear collaborator fixed effects specifications. This robustness check is informative insofar as linear specifications enable us to completely saturate the specifications with age effects (a total of 54 indicator variables, vs. 17 in the QML Poisson specifications presented in the main body of the paper). The results in column 1a imply that coauthors suffer a 1.55 yearly decline in JIF-weighted publication output following the death of their superstar collaborator. This represents a 8.14% decrease relative to the mean of the dependent variable at the time of death. In contrast, the estimate of the treatment effect in column 1a of Table III corresponds to a 8.79% decline in the JIF-weighed publication rate. The magnitudes observed in colums 1b through 2b in Table III and W4 are likewise very similar.

V — Publication-level Quality Adjustment using Citation Data

The quality adjustment used to produce JIF-weighted publication counts is crude. It does not allow us to learn whether the research that does not get published as a consequence of superstar death is more likely to be of great vs. marginal significance. Table W5 answers this question by modeling the effect of superstar extinction for the production of articles falling above various quantiles of the citation distribution. An important caveat is that the results pertain only to the set of 1,436 controls+$1,416$ treated=$2,852$ collaborators who are also part of our elite group of 10,349 scientists, since this is the set for which article-level citation data is available. These 2,852 scientists account for 28.16% of the collaborators in the overall sample.

Citation data suffer from a well known truncation problem: older articles have had more time to be cited, and hence are more likely to reach the tail of the citation distribution, ceteris paribus. To overcome this issue, we compute a different empirical cumulative distribution for the article-level distribution of citations *in each publication year*.[7] For example, in the life sciences broadly defined, an article published in 1980 would require at least 98 citations to fall into the top ventile of the distribution; an article published in 1990, 94 citations; and an article published in 2000, only 57 citations (this is illustrated in Figure W2). With these empirical distributions in hand, it becomes meaningful to count the number of articles that fall, for example, in the top quartile of citations for a given scientist in a particular year. These counts in turn provide the dependent variables used in Table W5.

We begin by replicating the results of Table III, Model 2b on this restricted set of collaborators. The treatment effect is slightly lower in magnitude, but remains highly statistically significant (column 1). The same result obtains when using the raw (i.e., not JIF-weighted) number of publications as the outcome variable (column 2). We then find that the magnitude of the treatment effect increases as one restricts the dependent variable to publications that fall in higher quantiles of the citation distribution. It hovers between -6 and -9% when we examine the effect of superstar extinction on publications that fall in the bottom quartile, below the median, above the median, or in the top quartile of the citation distribution. It increases to -9% for publications in the top ventile, and still further to -15% when focusing on "blockbuster" publications — those falling in the top percentile of the citation distribution. At the very least, these results suggest that superstar exposure is not limited to the production of relatively less significant scientific knowledge.

[7] We thank Stefan Wuchty and Ben Jones from Northwestern University for performing the computations. These vintage-specific distributions are not based on in-sample article data, but use the universe of articles published since 1970 in biomedical and chemical journals indexed by the Web of Science.

VI — Effect of Superstar Extinction on Receipt of NIH Funding

We present evidence on the effect of superstar extinction on receipt of NIH funding. Grants are typically awarded for a period of years (three to five is typical), and disbursed in equal yearly amounts over this period. Only the first of these payments is indicative of successful grantsmanship. We exclude from the calculation non-research grants (fellowships, training grants, and infrastructure grants), as well as large center research grants. The CGAF dataset only lists principal investigators (PIs) for each grant; as a result, we are unable to separate the grants in which coauthor and superstars are co-investigators from those that do not entail a formal research collaboration. This limitation must be borne in mind when interpreting the results of specifications relying on grant data. We also eliminate from the estimation sample 318 treated and 260 control collaborators who are NIH employees at some point during their career, and as such not eligible for receipt of extramural NIH funding.

Table W6 presents the results, using two different dependent variables: the number of research grants (columns 1a and 1b), and the probability of receiving at least one grant in a given year (columns 2a and 2b). The first two models are estimated using conditional collaborator fixed effects quasi-maximum likelihood. In these specifications, the 3,669 collaborators (38.72% of the controls vs. 36.55% of the treated) who never receive a grant during the observation period drop out of the observation sample. The last two models are estimated using a collaborator fixed effects linear probability model, on the entire sample of grant-eligible collaborators, including 37% among them who never receive (and may not even have applied for) a grant from the NIH.

The magnitudes of the effects in columns 1a and 1b are strikingly similar to those observed for publication output, though they are only statistically significant at the 10% level. In contrast, the magnitudes of the extinction effect for the linear probability models are quite small: they suggest that the probability of receiving a grant falls by a statistically significant 1% after the scientist loses a superstar collaborator. We must interpret these results with caution: there is obviously large heterogeneity in the quality and importance of research grants, and our dependent variable does not account for this.

VII — Treatment Effect Heterogeneity: Impact of Collaborator Status and Age at the Time of Superstar Death

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 documented earlier. Column 1 focuses on faculty members whose sole elite collaborator was the superstar who died. For these coauthors with relatively poor substitution opportunities (they account for roughly 27.66% of the dyads in the sample), the consequences of the superstar's loss are particularly severe, with an overall 15.3% decline in publication output. Columns 2 asks whether scientists who are PIs on a NIH R01 grant at the time of their superstar coauthor's death are shielded from the adverse effects documented earlier. With independent funding of this type, these investigators (who account for more than half of the sample) are likely to be less dependent on the goodwill of their collaborators, but we find no evidence supporting this conjecture: the differential effect is an small and imprecisely estimated; Independent NIH funding is not enough to insulate scientists from the loss of an eminent collaborator. In column 3, we present evidence that the "elite among the elite" (members of the National Academy of Science, Howard Hughes Medical Investigators, and NIH MERIT awardees who together account for 8.5% of the total number of collaborators) is relatively unaffected by the loss of a "peer superstar." The differential impact on elite coauthors is positive, large, and statistically significant; it offsets almost exactly the main treatment effect.

We conclude that the effect of superstar extinction is heterogeneous with respect to coauthor status, but the heterogeneity stems from the tails of the status distribution. The loss of a prominent collaborator adversely

impacts the productivity of investigators even if they are independently funded, unless they have already achieved great renown at the time of the star's death.[8]

We also investigate whether the magnitude of the treatment effect varies with the collaborator's age at the time of death for the superstar. To do so, we interact the treatment effect with 8 indicator variables corresponding to different career age brackets: 5 to 10 years, 10 to 15 years, 15 to 20 years, 20 to 25 years, 25 to 30 years, 30 to 35 years, 35 to 40 years, and more than 40 years of career age at death. We then plot the corresponding coefficient estimates in Figure W3, along with their 95% confidence interval. The effect decreases monotonically with the age of the collaborator at death, becoming insignificantly different from 0 after twenty five years of career age. Therefore, researchers appear particularly vulnerable to the loss of a superstar coauthor in the early part of their scientific career.

VIII — Robustness Checks

In Table W8, we present the results of a number of robustness checks, using Model (1b) of Table III as a benchmark specification. In column 2, we examine whether a small number of stars with many collaborators drive the main results. We drop all collaborators for the 7 superstars with the highest number of collaborators (120 or more) from the estimation sample. The magnitude of the treatment effect drops only slightly, and remains highly statistically significant. In column (3), we add to the specification 10 indicator variables for the superstar's imputed career age. This decreases the magnitude of the treatment effect by from -0.086 to -.066. In columns (4a) and (4b), we explore the sensitivity of our results to changes in our arbitrary age cutoff for the the superstar's age at death. In (4a), we limit the sample to 71 stars who were 60 years old or younger at the time of their death. This results in an even higher magnitude for the extinction effect (-.113 instead of -.092). In contrast, we obtain a much smaller magnitude (-0.051) when we focus on the collaborators of 38 eminent scientists who die beyond the creative stages of their career — at 75 years of age or older (column 4b). This effect is also imprecisely estimated.

We then examine the possibility that our control group is contaminated because some of the control collaborators are separated from treated collaborators by a only few degrees in the coauthorship network. Specifically, we keep in the estimation sample only those scientists that are at least 3 degrees apart in the coauthorship network formed by all 10,349 superstars. These scientists represent 75% of the overall sample. In column 5, we find that the treatment effect increases in magnitude, which is consistent with the hypothesis that the effect of superstar extinction extends beyond the set of direct coauthors, but decays quickly with social distance.

Finally, we perform a small simulation study to validate the quasi-experiment exploited in the paper (column 6). 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 (based on 500 replications) is a precisely estimated 0.

[8] As seen in Table W5, taken as a whole, the set of elite coauthors suffers a decline in output similar to the one observed for the universe of all coauthors (i.e., in Table III). At the risk of repeating ourselves, the elite sample is very heterogeneous, and does include young, old, and fading stars.

IX — Main Results for Anticipated Death

In Table W9 and Figure W4, we present some results pertaining to the 6,515 collaborators of 136 superstars who died prematurely, but whose particular circumstances imply that their passing was anticipated. The vast majority of these anticipated deaths are due to cancer. Since coauthors might alter their collaboration strategies even before the superstar's death, the case for exogeneity of the extinction event is weaker in this case.[9]

The results in Table W9 parallel exactly those presented in Table III. We find that the treatment effect is of lower magnitude than in the sudden case (especially when the estimation sample includes control collaborators), and less precisely estimated. We also find very little evidence of impact on the publication output without the superstar (columns 2a and 2b). Figure W4 mirrors Figure IIIA. The trajectory of output appears to begin its monotonic decline prior the superstar's death (though the corresponding interaction terms are very small in magnitude, and statistically insignificant). The treatment effect, though consistently negative in sign, reaches statistical significance only in the long run — 10 years or so after the superstar's death.

These findings suggest that collaborators and quite possibly the superstar him/herself adjust their behavior in anticipation of the star's impending death. Though the determinants and particular form of these endogenous responses are certainly worthy of study, they are beyond the scope of the present paper.

X — An Alternative Interpretation Based on a Sociological Mechanism: Ascription

Sociological studies of the scientific reward system have provided some evidence supporting the existence of the "Matthew Effect,"[10] whereby scientists receive differential recognition for a particular scientific contribution depending on their location in the status hierarchy (Merton 1968; Cole 1970). It is possible that editors and reviewers ascribe positive qualities to research they are charged with evaluating because of the mere presence of the superstar's name on the authorship roster, regardless of the contribution's intrinsic merits.

The relevance of this dynamic in our setting is doubtful for two reasons. First, we observe a decline in the output written independently of the star (Table III); second, the treatment effect is not driven by the collaborators who have recent, or many collaborations; third, its onset is delayed until after the death of the star. These facts argue against an interpretation of the effect based on ascription.

[9] Most of the anticipated deaths are due to conditions with relatively short life expectancies; those with longer ones are not necessarily viewed as terminal until the final stages. Six scientists who died from a neurodegenerative disease constitute an exception. They were included in the sample because their obituaries implied they had remained actively engaged in research until a short period before their death. We verified that our results are robust to the omission of these six superstars.

[10] *"For unto every one that hath shall be given, and he shall have abundance; but from him that hath not shall be taken away even that which he hath"* [Matthew 25:29]

References

Azoulay, Pierre, Andrew Stellman, and Joshua Graff Zivin, "PublicationHarvester: An Open-source Software Tool for Science Policy Research," *Research Policy*, 35 (2006), 970-974.

Benner, Mary, and Joel Waldfogel, "Close to You? Bias and Precision in Patent-based Measures of Technological Proximity," *Research Policy*, 37 (2008), 1556-1567.

Cole, Stephen, "Professional Standing and the Reception of Scientific Discoveries," *The American Journal of Sociology*, 76 (1970), 286-306.

Jaffe, Adam B., "Technological Opportunity and Spillovers from R&D: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, 76 (1986), 984-1001.

Lander, Eric S., et al., "Initial Sequencing and Analysis of the Human Genome," *Nature*, 409 (2001), 934-941.

Merton, Robert K., "The Matthew Effect in Science," *Science*, 159 (1968), 56-63.

Trajtenberg, Manuel, Gil Shiff, and Ran Melamed, "The 'Names Game': Harnessing Inventors' Patent Data for Economic Research," NBER Working Paper #12479 (2006).

Wade, Nicholas, *The Nobel Duel: Two Scientists' 21-year Race to Win the World's Most Coveted Research Prize* (Garden City, NY: Anchor Press/Doubleday, 1981).

Table W1
Superstar Sample

			Cause of Death	Institutional Affiliation	Field
Henry G. Kunkel	(1916-1983)	MD	complications after vascular surgery	Rockefeller University	identification of MHC Class II molecules
John P. Merrill	(1917-1984)	MD	drowned	Harvard Medical School/Brigham & Women's Hospital	role of the immune system in kidney transplantation
Merton F. Utter	(1917-1980)	PhD	heart attack	Case Western Reserve University School of Medicine	structure and function of pep carboxykinase isozymes
Abraham I. Braude	(1917-1984)	MD/PhD	heart attack	UCSD	pathogenesis and treatment of life-threatening septic shock
E. Jack Wylie	(1918-1982)	MD	heart attack	UCSF	development of techniques for the treatment and management of chronic visceral ischemia
Abraham M. Lilienfeld	(1920-1984)	MD	heart attack	Johns Hopkins University School of Public Health	epidemiological methods for the study of chronic diseases
Sidney Riegelman	(1921-1981)	PhD	drowned while scuba diving	UCSF	intersubject variation in first pass effect of drugs
Susumu Hagiwara	(1922-1989)	PhD	bacterial infection	UCLA	evolutionary and developmental properties of calcium channels in cell membranes
Lucille S. Hurley	(1922-1988)	PhD	complications from open heart surgery	University of California — Davis	genetic and nutritional interactions in development
Lewis W. Wannamaker	(1923-1983)	MD	heart attack	University of Minnesota Medical School	clinical and epidemiologic aspects of streptococcal infections
Eugene C. Jorgensen	(1923-1981)	PhD	murdered	UCSF	structure/activity relationships of compounds related to thyroxin
James M. Felts	(1923-1988)	PhD	heart failure	UCSF	synthesis and processing of plasma lipoproteins
Josiah Brown	(1923-1985)	MD	tragic accident	UCLA	biochemical studies of lipid and carbohydrate metabolism
Thomas R. Johns, 2nd	(1924-1988)	MD	refractory arrhythmia	University of Virginia School of Medicine	physiological studies of myasthenia gravis
Robert J. Stoller	(1924-1991)	MD	killed by a reckless teenage driver	UCLA	clinical studies of gender identity
Lucien J. Rubinstein	(1924-1990)	MD	ruptured intracranial aneurysm	University of Virginia School of Medicine	differentiation and stroma-induction in neural tumors
William H. Oldendorf	(1925-1992)	MD	complications from heart disease	UCLA	x-ray shadow radiography and cerebral angiography
Margaret O. Dayhoff	(1925-1983)	PhD	heart attack	Georgetown University Medical Center	computer study of sequences of amino acids in proteins
Norman Geschwind	(1926-1984)	MD	heart attack	Harvard Medical School/Beth Israel Medical Center	relationship between the anatomy of the brain and behavior
Norbert Freinkel	(1926-1989)	MD	heart attack	Northwestern University	metabolic regulation in normal and diabetic pregnancies
Edward V. Evarts	(1926-1985)	MD	heart attack	NIH	electrophysiological activity of in vivo neurons in waking and sleeping states
Zanvil A. Cohn	(1926-1993)	MD	aortic dissection	Rockefeller University	macrophage in cell biology and resistance to infectious disease
Daniel Rudman	(1927-1994)	MD	complications from brain surgery	Medical College of Wisconsin	adipokinetic substances of the pituitary gland
Gerald P. Rodnan	(1927-1983)	MD	complications after vascular surgery	University of Pittsburgh	renal transport if uric acid and protein
Gustavo Cudkowicz	(1927-1982)	MD	brief illness	SUNY Buffalo	controls of proliferation specific for leukemias
Gerald D. Aurbach	(1927-1991)	MD	hit in a head by a stone	NIH	bone metabolism and calcium homeostasis
George Streisinger	(1927-1984)	PhD	scuba-diving accident	University of Oregon	genetic mutations and the nervous system development in lower vertebrates
Carl Monder	(1928-1995)	PhD	brief illness, acute fulminating leukemia	Population Council	corticosteroid metabolism in juvenile hypertension
Lucien B. Guze	(1928-1985)	MD	sudden cardiac arrest	UCLA	pathogenesis of experimental pyelonephritis
Edgar C. Henshaw	(1929-1992)	MD	complications from early-stage cancer treatment	University of Rochester	intermediary metabolism in animals and in man
Donald J. Magilligan, Jr.	(1929-1989)	MD	short illness	Henry Ford Health Sciences Center	natural history and limitations of porcine heart valves
Lubomir S. Hnilica	(1929-1986)	PhD	automobile accident	Vanderbilt University	nuclear antigens in human colorectal cancer
Laurence M. Sandler	(1929-1987)	PhD	heart attack	University of Washington School of Medicine	cytogenetics of meiosis and development in drosophila
DeWitt S. Goodman	(1930-1991)	MD	pulmonary embolism	Columbia University	lipid metabolism and its role in the development of heart and artery disease
George B. Craig, Jr.	(1930-1995)	PhD	heart attack	University of Notre Dame	genetics and reproductive biology of aedes mosquitoes
Hymie L. Nossel	(1930-1983)	MD/PhD	heart attack	Columbia University	causes of thrombosis and the nature of hemostasis
James W. Prahl	(1931-1979)	MD/PhD	rock climing accident	University of Utah	structural basis of the functions of human complement
Harold A. Baltaxe	(1931-1985)	MD	heart attack	University of California — Davis	development of new coronary angiographic techniques
George J. Schroepfer, Jr.	(1932-1998)	MD/PhD	heart attack	Rice University	regulation of the formation and metabolism of cholesterol
Philip J. Fialkow	(1933-1996)	MD	trekking accident in Nepal	University of Washington	origins of myeloid leukemia tumors
John C. Seidel	(1933-1988)	PhD	heart attack	Boston Biomedical Research Institute	actin-myosin interaction in pulmonary smooth muscle
Issa Yaghmai	(1933-1992)	MD	sudden cardiac arrest	UCLA-Olive View Medical Center	radiological diagnosis of musculoskeletal disorders
Donald C. Shreffler	(1933-1994)	PhD	heart attack	Washington University in St. Louis	organization and functions of H-2 gene complex
Paul B. Sigler	(1934-2000)	MD/PhD	heart attack	Yale University	structural analysis of biological macromolecules
Kenneth L. Melmon	(1934-2002)	MD	heart attack	Stanford University	autacoids as pharmacologic modifiers of immunity
Gerald P. Murphy	(1934-2000)	MD	heart attack	Roswell Park Cancer Institute/SUNY Buffalo	detection, immunotherapy, and prognostic indicators of prostate cancer
Demetrios Papahadjopoulos	(1934-1998)	PhD	adverse drug reaction/multi-organ failure	UCSF	phospholipid-protein interactions, lipid vesicles, and membrane function
Takis S. Papas	(1935-1999)	PhD	unexpected and sudden	Medical University of South Carolina	characterization of ETS genes and retroviral onc genes
Donald T. Witiak	(1935-1998)	PhD	stroke	University of Wisconsin	stereochemical studies of hypocholesterolemic agents
Shu-Ren Lin	(1936-1979)	MD	plane crash	University of Rochester	imaging studies of cerebral blood flow after cardiac arrest
James R. Neely	(1936-1988)	PhD	heart attack	Penn State University	effects of diabetes and oxygen deficiency in regulation of metabolism in the heart
D. Martin Carter	(1936-1993)	MD/PhD	dissecting aortic aneurysm	Rockefeller University	susceptibility of pigment and cutaneous cells to DNA injury by UV
Dale E. McFarlin	(1936-1992)	MD	heart attack	NIH	neuroimmunological studies of multiple sclerosis
Roy D. Schmickel	(1936-1990)	MD	died tragically	University of Pennsylvania	isolation and characterization of human ribosomal DNA
John J. Jeffrey, Jr.	(1937-2001)	PhD	stroke	Albany Medical College	mechanism of action and the physiologic regulation of mammalian collagenases
Victor J. Ferrans	(1937-2001)	MD/PhD	complications from diabetes	NIH	myocardial and vascular pathobiology
Sandy C. Marks, Jr.	(1937-2002)	DDS/PhD	heart attack	UMASS	bone cell biology
A. Arthur Gottlieb	(1937-1998)	MD	pulmonary embolus following surgery	Tulane University School of Medicine	role of macrophage nucleic acid in antibody production
Patricia S. Goldman-Rakic	(1937-2003)	PhD	struck by a car	Yale University	development and plasticity of the primate frontal lobe

Thomas P. Dousa	(1937-2000)	MD/PhD	heart attack	Mayo Clinic	cellular action of vasopressin in the kidney
William L. McGuire	(1937-1992)	MD	scuba-diving accident	University of Texas HSC at San Antonio	mechanisms of hormonal control and growth and regression of mammary carcinoma
Roland L. Phillips	(1937-1987)	MD/PhD	glider plane accident	Loma Linda University School of Medicine	role of lifestyle in cancer and cardiovascular disease among Adventists
Emil T. Kaiser	(1938-1988)	PhD	complications from kidney transplant	Rockefeller University	mechanism of carboxypeptidase action
John H. Walsh	(1938-2000)	MD	heart attack	UCLA	gastrointestinal hormones, gastric acid production and peptic ulcer disease
Harold A. Menkes	(1938-1987)	MD	car accident	Johns Hopkins University	occupational and environmental lung disease
Thomas F. Burks, II	(1938-2001)	PhD	heart attack	University of Texas HSC at Houston	central and peripheral neuropeptide pharmacology
Verne M. Chapman	(1938-1995)	PhD	died suddenly while attending meeting	Roswell Park Cancer Institute/SUNY Buffalo	development of cumulative multilocus map of mouse chromosomes
Samuel A. Latt	(1938-1988)	MD/PhD	heart attack	Harvard Medical School/Children's Hospital	genetic and cytogenetic studies of mental retardation
Walter F. Heiligenberg	(1938-1994)	PhD	plane crash	UCSD	neuroethological studies of electrolocation
Dolph O. Adams	(1939-1996)	MD/PhD	heart attack	Duke University	Development and regulation of macrophage activation
James N. Davis	(1939-2003)	MD	airplane crash	SUNY HSC at Stony Brook	mechanisms underlying neuronal injury after brain ischemia
Raymond R. Margherio	(1940-2000)	MD	aneurysm	Wayne State University School of Medicine	clinical studies in age-related eye diseases
Robert M. Macnab	(1940-2003)	PhD	accidental fall	Yale University	sequence analysis and function of bacterial flagellar motor
D. Michael Gill	(1940-1990)	PhD	heart attack	Tufts University	biochemistry of cholera toxin and other pathogenic toxins
Anthony Dipple	(1940-1999)	PhD	heart attack	NIH	metabolic activation and DNA interactions of polycyclic aromatic hydrocarbon carcinogens
Ronald G. Thurman	(1941-2001)	PhD	massive heart attack	University of North Carolina	hepatic metabolism, alcoholic liver injury and toxicology
Richard E. Heikkila	(1942-1991)	PhD	murder	UMDNJ Robert Wood Johnson Medical School	oxidation-reduction reactions and the dopamine receptor system
Julio V. Santiago	(1942-1997)	MD	heart attack	Washington University in St. Louis	role of social factors, lifestyle practices, and medication in the onset of type II diabetes
Pokar M. Kabra	(1942-1990)	PhD	plane crash	UCSF	application of liquid chromatography to therapeutic drug monitoring
Simon J. Pilkis	(1942-1995)	MD/PhD	heart attack	University of Minnesota	carbohydrate metabolism and diabetes
Christopher A. Dawson	(1942-2003)	PhD	heart attack	Medical College of Wisconsin	pulmonary hemodynamics
Bruce M. Achauer	(1943-2002)	MD	gastrointestinal bacterial infection	University of California — Irvine	non-invasive methods to assess the depth of burn wounds
Roland D. Ciaranello	(1943-1994)	MD	heart attack	Stanford University	molecular neurobiology and developmental disorders
Fredric S. Fay	(1943-1997)	PhD	heart attack	UMASS	generation and regulation of force in smooth muscle
Thomas A. McMahon	(1943-1999)	PhD	complications from routine surgery	Harvard University	orthopedic biomechanics
William D. Nunn	(1943-1986)	PhD	sudden cardiac arrest	University of California — Irvine	regulation of fatty acid/acetate metabolism in e. coli
Ahmad I. Bukhari	(1943-1983)	PhD	heart attack	Cold Spring Harbor Laboratory	life cycle of mutator phage μ
James S. Seidel	(1943-2003)	MD/PhD	bacterial infection	Harbor-UCLA Medical Center	clinical studies in pediatric life support and cardiopulmonary resuscitation
Jonathan M. Mann	(1943-1998)	MD	plane crash	Harvard University School of Public Health	AIDS prevention
Lonnie D. Russell, Jr.	(1944-2001)	PhD	swimming accident	Southern Illinois University School of Medicine	filament regulation of spermatogenesis
Don C. Wiley	(1944-2001)	PhD	accidental fall	Harvard University	viral membrane and glycoprotein structure
Roger R. Williams	(1944-1998)	MD	airplane crash	University of Utah	genetics and epidemiology of coronary artery diseases
G. Scott Giebink	(1944-2003)	MD	heart attack	University of Minnesota	pathogenesis of otitis media and immunizations
Joaquim Puig-Antich	(1944-1989)	MD	asthma attack	University of Pittsburgh	psychobiology and treatment of child depression
Peter M. Steinert	(1945-2003)	PhD	heart attack	NIH	structures and interactions of the proteins characteristic of epithelial cells
John P. Merlie	(1945-1995)	PhD	heart failure	Washington University in St. Louis	molecular genetics of the acetylcholine receptor
Howard S. Tager	(1945-1994)	PhD	heart attack	University of Chicago	biochemical structure, action, regulation and degradation of the insulin and glucagon molecules
John J. Wasmuth	(1946-1995)	PhD	heart attack	University of California — Irvine	human-hamster somatic cell hybrids/localization of Hnyington's disease gene
Stanley R. Kay	(1946-1990)	PhD	heart attack	Albert Einstein College of Medicine	symptoms and diagnostic tests of schizophrenia
Mary Lou Clements	(1946-1998)	MD	airplane crash	Johns Hopkins University	development of AIDS vaccines
Ronald E. Talcott	(1947-1984)	PhD	automobile accident	UCSF	carboxylesterases of toxicologic significance
Lynn M. Wiley	(1947-1999)	PhD	plane crash	University of California — Davis	morphogenesis in early mammalian embryos
John B. Penney, Jr.	(1947-1999)	MD	heart attack	Harvard Medical School/MGH	receptor mechanisms in movement disorder pathophysiology
Jeffrey M. Isner	(1947-2001)	MD	heart attack	Tufts University	therapeutic angiogenesis in vascular medicine, cardiovascular laser phototherapy
Trudy L. Bush	(1949-2001)	PhD	heart attack	University of Maryland School of Medicine	postmenopausal estrogen/progestins interventions
Neil S. Jacobson	(1949-1999)	PhD	heart attack	University of Washington	marital therapy, domestic violence, and the treatment of depression
Tsunao Saitoh	(1949-1996)	PhD	murdered	UCSD	altered protein kinases in alzheimer's disease
Gary J. Miller	(1950-2001)	MD/PhD	heart attack	University of Colorado HSC	vitamin D receptors in the growth regulation of prostate cancer cells
Elizabeth A. Rich	(1952-1998)	MD	traffic accident	Case Western Reserve University School of Medicine	natural history of lymphocytic alveolitis in hiv disease
Matthew L. Thomas	(1953-1999)	PhD	died while travelling	Washington University in St. Louis	function and regulation of leukocyte surface glycoproteins
Mu-En Lee	(1954-2000)	MD/PhD	complications from routine surgery	Harvard Medical School/MGH	characterization of vascular smooth muscle LIM protein
Alan P. Wolffe	(1959-2001)	PhD	car accident	NIH	role of DNA methylation in regulating gene expression in normal and pathological states

Table W2
Measuring Proximity in Ideas Space

	Andrew Schally	Roger Guillemin	Dyad
Top 10 overlapping MeSH terms with highest combined use			
Animals	170	49	
Rats	127	33	
Male	131	23	
Gonadotropin-Releasing Hormone	121	9	
Luteinizing Hormone	121	8	
Humans	94	23	
Female	106	8	
Follicle Stimulating Hormone	81	6	
Pituitary Gland	65	19	
Time Factors	54	8	
Top 10 overlapping MeSH terms with lowest combined use			
Molecular Weight	1	1	
Somatomedins	1	1	
Peptide Chain Termination, Translational	1	1	
Steroids	1	1	
Arginine Vasopressin	1	1	
Propylthiouracil	1	1	
Neural Pathways	1	1	
Electric Stimulation	1	1	
Cerebellum	1	1	
Fatty Acids, Nonesterified	1	1	
Number of Publications	240	60	
Number of MeSH Terms (freq.-unweighted)	534	220	
Number of MeSH Terms (freq.-weighted)	3,035	750	
Number of Ovrlp. MeSH Terms (freq.-unweighted)			147
Number of Ovrlp. MeSH Terms (freq.-weighted)			609
Proximity of Guillemin to Schally (freq.-unweighted)			0.668
Proximity of Schally to Guillemin (freq.-unweighted)			0.275
Proximity of Guillemin to Schally (freq.-weighted)			0.812
Proximity of Schally to Guillemin (freq.-weighted)			0.201

Table W3
Trends in Stars' Publication Output Around the Time of Death

	All Pubs.	Last-Authored Pubs.
2 years after year of death	-0.988**	-1.238**
	(0.159)	(0.246)
1 year after year of death	-0.103	-0.143
	(0.111)	(0.165)
year of death	0.047	0.079
	(0.123)	(0.167)
1 year before year of death	0.125	0.209
	(0.101)	(0.129)
2 years before year of death	0.103	0.063
	(0.093)	(0.129)
3 years before year of death	0.201*	0.200
	(0.099)	(0.129)
4 years before year of death	0.077	0.057
	(0.100)	(0.147)
5 years before year of death	0.130†	0.141
	(0.075)	(0.095)
Log Quasi-Likelihood	-831,596	-626,787
Nb. of Obs.	104,154	103,959

Notes: The estimates above are taken from a conditional fixed effects Poisson specification that also include 54 career age indicator variables and a full suite of calendar year effects (estimates not reported). The estimate in column (1) implies a statistically significant (1-exp[-.988]))=62.77% decrease in the rate of publication two years after a superstar scientist passes away (regardless of cause of death), relative to years prior to the last 5 years of his/her life. The dependent variable in column 1 is the weighted article count for the superstar. Columns 2 restricts the count to publications in which the superstar appears in last position on the authorship roster. The weights used to create these counts are Journal Impact Factors (JIF) published by the Institute for Scientific Information. Robust (QML) standard errors are reported in parentheses.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table W4
Impact of Superstar Death on Collaborators' Publication Rates – OLS

	All JIF-Weighted Publications		JIF-Weighted Pubs. Written with others	
	Without Ctrls (1a)	With Ctrls (1b)	Without Ctrls (2a)	With Ctrls (2b)
After Death	-1.553**	-1.528**	-0.947†	-0.980†
	(0.513)	(0.531)	(0.501)	(0.514)
R^2	0.574	0.566	0.576	0.569
Nb. of Obs.	153,508	294,943	153,508	294,943
Nb. of Collaborators	5,267	10,128	5,267	10,128

Notes: Estimates stem from collaborator fixed effects linear 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 54 age category indicator variables (career age less than -3 is the omitted category). Robust standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table W5
Impact of Superstar Death on Collaborators' Citation Impact [Elite Subsample]

	JIF-wghtd. Pubs	**All Pubs**	**Pubs in Bottom Quartile**	**Pubs below the Median**	**Pubs above the Median**	**Pubs in Top Quartile**	**Pubs in Top Ventile**	**Pubs in Top Percentile**
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
After Death	-0.062*	-0.068**	-0.089†	-0.071	-0.069**	-0.073**	-0.105**	-0.161**
	(0.028)	(0.024)	(0.052)	(0.044)	(0.023)	(0.022)	(0.028)	(0.044)
Log Pseudo-Likl.	-773,535	-226,069	-46,236	-98,925	-207,452	-185,282	-125,897	-69,291
Nb. of Obs.	86,457	86,457	86,457	86,457	86,457	86,457	86,457	86,457
Nb. of Collabs.	2,852	2,852	2,852	2,852	2,852	2,852	2,852	2,852

Notes: Conditional fixed effects quasi-maximum likelihood estimates for the determinants of publication rates among coauthors of "superstar" academic life scientists. We bin their publications according to the various quantiles of the vintage-specific, article-level distribution of citations they fall into. For instance, an article that garnered 100 citations by 2008 would fall above the top ventile of the 1980 citation distribution, but above the top percentile of the 2000 distribution. The underlying empirical distributions were computed using the universe of publications and citations in the biomedical and chemical journals indexed by ISI/Web of Science (Table W2). Because article-level citation data is only available for scientists in the elite subsample (n=10,349), we restrict the estimation sample to elite coauthors, which account for a 28.16% of the collaborators in the overall sample. All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category). Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table W6
Impact of Superstar Death on Receipt of NIH Funding

	Nb. of Grants		At Least One Grant	
	Without Ctrls (1a)	With Ctrls (1b)	Without Ctrls (2a)	With Ctrls (2b)
After Death	-0.096†	-0.095†	-0.010**	-0.010**
	(0.050)	(0.050)	(0.004)	(0.004)
Log Quasi-Likelihood/R^2	-44,166	-83,644	0.230	0.221
Nb. of Obs.	92,014	175,062	143,727	277,922
Nb. of Collaborators	3,140	5,965	4,949	9,574

Notes: Estimates stem from collaborator fixed effects QML Poisson specifications (columns (1a) and (1b)) and collaborator fixed effects linear probability model specifications (columns (2a) and (2b)). Estimates stem from conditional quasi-maximum likelihood Poisson specifications. The dependent variable in columns 1a and 1b is the total number of NIH research grants and contracts (R, U, N, or K codes, new grants or competitive renewals) awarded in the year of observation. In columns (2a) and (2b), the dependent variable is an indicator for award of at least one such grant or contract. All models incorporate a full suite of year effects as well as 17 age category indicator variables (career age less than -3 is the omitted category). Robust standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table W7
Impact of Collaborator Status at the Time of Superstar Death

	No Other Elite Coauthor	R01 Grantee	MERIT, NAS, or HHMI	All Covariates Combined
	(1)	(2)	(3)	(4)
After Death	-0.077**	-0.096*	-0.109**	-0.091*
	(0.026)	(0.040)	(0.029)	(0.042)
After Death × Coauthor has no other superstar collaborator	-0.089*			-0.069
	(0.039)			(0.042)
After Death × Coauthor Holds R01 Grant		0.014		-0.015
		(0.037)		(0.037)
After Death × Coauthor "Elite"			0.118**	0.114**
			(0.042)	(0.043)
% of Collabs. Affected	27.66	56.50	8.48	
Log Pseudo-Likelihood	-1,832,458	-1,832,586	-1,832,104	-1,832,022
Nb. of Obs.	294,943	294,943	294,943	294,943
Nb. of Collabs.	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. We interact the treatment variable with indicator variables capturing various aspects of coauthor status: poor substitution opportunities, i.e., coauthors with no other elite coauthor save the extinct superstar; R01 grantee status at the time of death; and a composite "Elite" indicator variable combining membership in the National Academy of Science, MERIT Award from the NIH, and HHMI investigatorship. All models incorporate year effects and 17 age category indicator variables (career age less than -3 is the omitted category), as well as 17 interaction terms between the age effects and the covariate of interest. Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table W8
Sensitivity Checks

	Benchmark Specification	**Without "Gregarious" Superstars**	**Superstar Age Ctrls.**	**Superstar Age at Death Cutoff**		**Leakage through Coauthorship Network**	**Placebo Test**
	(1)	(2)	(3)	(4a)	(4b)	(5)	(6)
After Death	-0.086**	-0.084**	-0.066**	-0.113**	-0.051	-0.096**	-.00009
	(0.025)	(0.022)	(0.024)	(0.032)	(0.048)	(0.025)	(0. 0162)
Robustness	Table III, Column (1b)	w/o stars with 120 coauthors or more	with star age indic. vars.	<60 yrs old	>75 yrs old	path length>2	Only Controls
Log Likelihood	-1,832,594	-1,517,842	-1,827,615	-1,048,026	-627,820	-1,338,509	-914,468
Nb. of Obs.	294,943	246,405	294,943	165,635	106,696	218,400	147,339
Nb. of Collabs.	10,128	8,488	10,128	5,694	3,594	7,510	5,064

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 year effects and 17 age category indicator variables (career age less than -3 is the omitted category). Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Table W9
Impact of Superstar Death on Collaborators' Publication Rates
(136 stars whose premature deaths could be anticipated)

	All JIF-Weighted Publications		JIF-Weighted Pubs. Written with others	
	Without Ctrls (1a)	With Ctrls (1b)	Without Ctrls (2a)	With Ctrls (2b)
After Death	-0.057**	-0.038†	-0.016	0.001
	(0.021)	(0.021)	(0.020)	(0.020)
Log Pseudo-Likelihood	-1,186,953	-2,213,193	-1,159,119	-2,152,735
Nb. of Obs.	191,771	370,166	191,771	370,166
Nb. of Collaborators	6,515	12,592	6,515	12,592

Notes: Estimates stem from conditional quasi-maximum likelihood Poisson specifications. 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-exp[-0.092])=8.79% decrease in the rate of publication after their superstar coauthor passes away. All models incorporate a full suite of year effects as well as 17 age category indicator variables (career age less than -3 is the omitted category). Robust (QML) standard errors in parentheses, clustered at the level of the superstar.

$^{\dagger}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$

Figure W1
Coauthor Matching for a Sample Publication

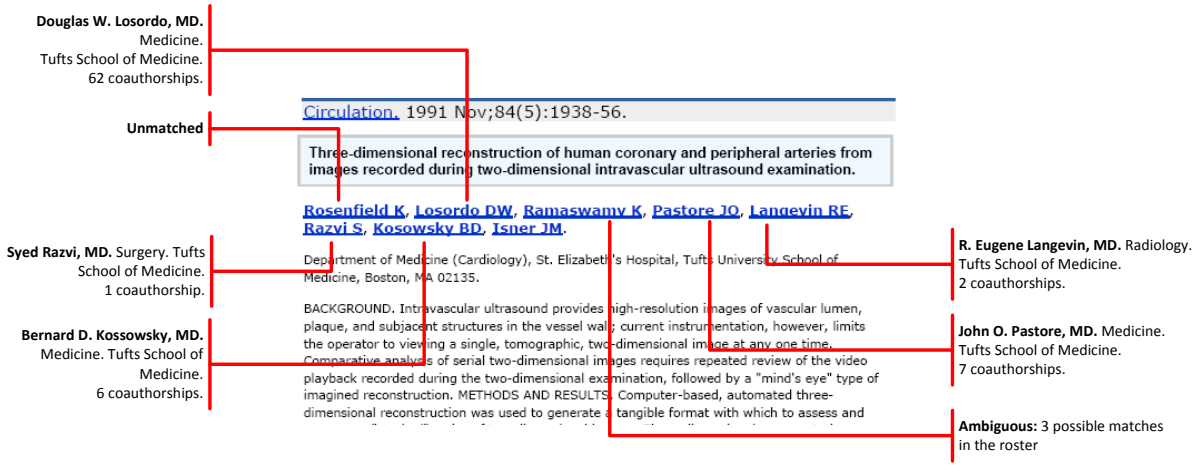


Figure W2
Vintage-specific Empirical Distributions of Citations at the Article Level

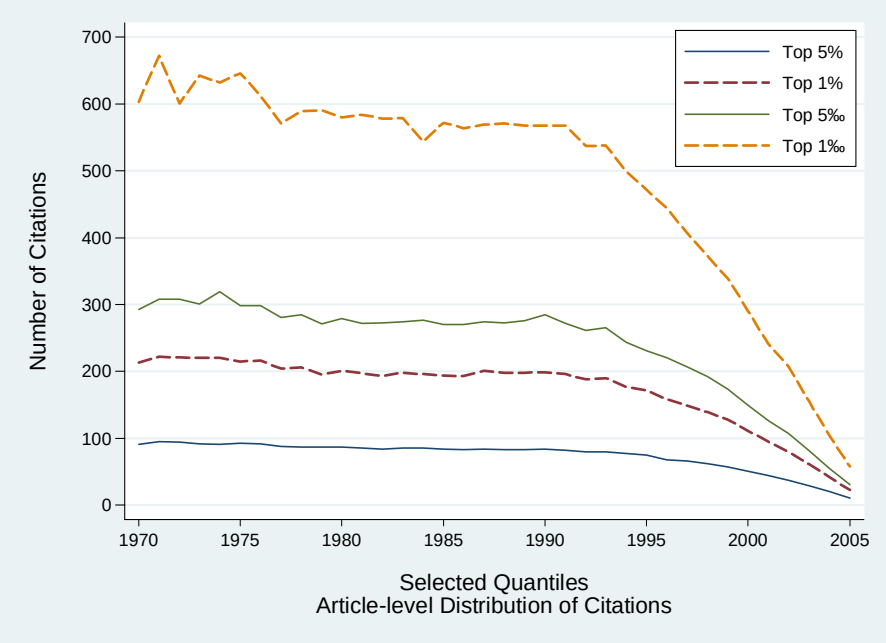
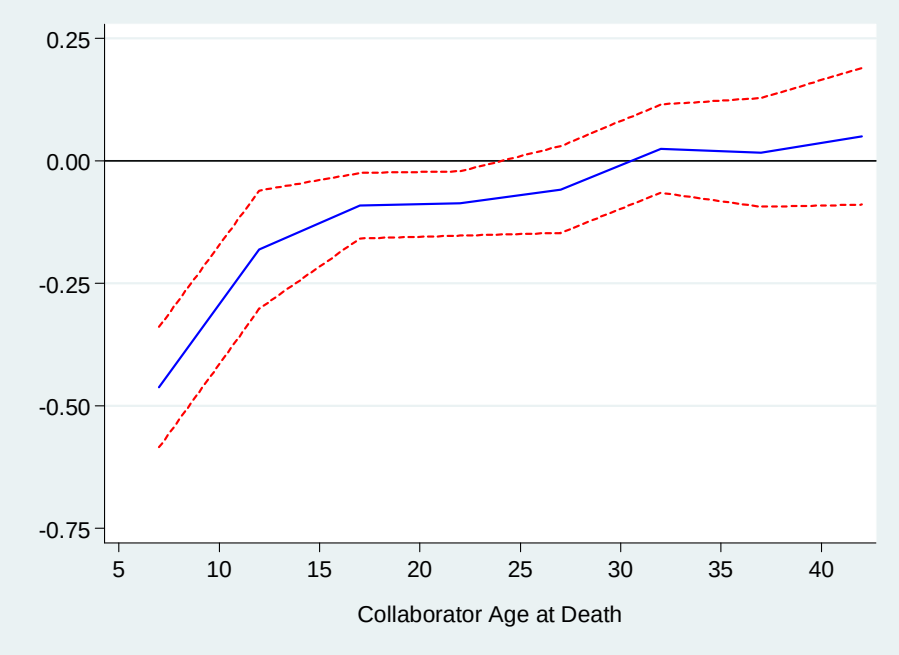
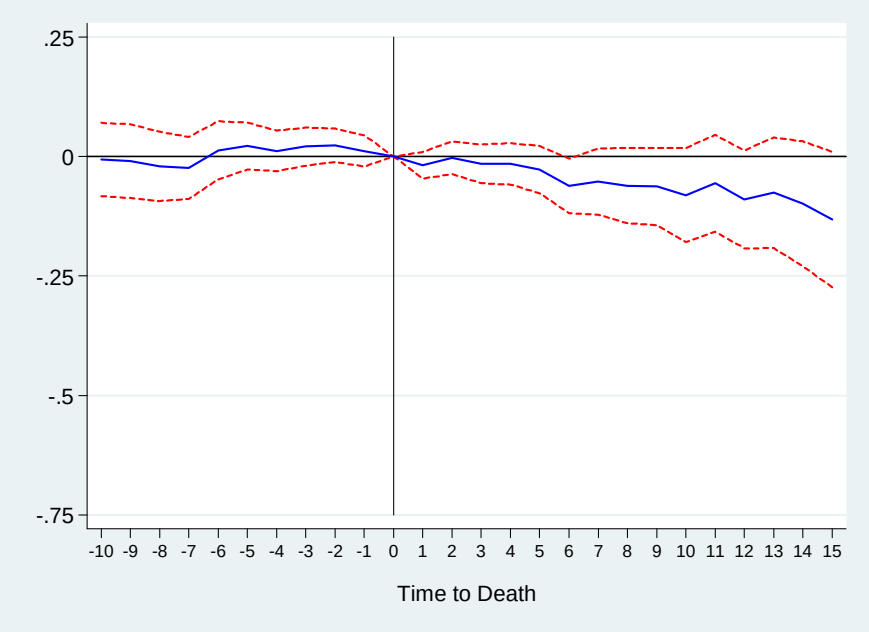


Figure W3
Magnitude of the Treatment Effect as a Function of Collaborator Age at Time of Superstar Death



Notes: The solid blue lines in the above plot correspond to coefficient estimates of a conditional fixed effects quasi-maximum likelihood Poisson specification in which the weighted publication output of a collaborator is regressed onto year effects, 17 indicator variables corresponding to different age brackets, and interactions of the treatment effect with 8 indicator variables corresponding to different brackets for the career age of the collaborator at the time of superstar death: 5 to 10 years, 10 to 15 years, 15 to 20 years, 20 to 25 years, 25 to 30 years, 30 to 35 years, 35 to 40 years, and more than 40 years of career age. The 95% confidence interval (corresponding to robust standard errors, clustered around superstars) around these estimates is plotted with dashed red lines. The baseline specification is that of Table III, Column (1b).

Figure W4
Dynamics of the Treatment Effect
(136 stars whose premature deaths could be anticipated)



The solid blue lines in the above plot correspond to coefficient estimates of a conditional fixed effects quasi-maximum likelihood Poisson specification in which the weighted publication output of a collaborator is regressed onto year effects, 17 indicator variables corresponding to different age brackets, and interactions of the treatment effect with 27 indicator variables corresponding to 11 years before the year of death and prior, 10 years before the year of death, 9 years before the year of death,..., 14 years after the year of death, and 15 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 red lines. The figure uses Column (1b) of Table W9 as a baseline (i.e., treated and control collaborators, the dep. var. includes all of the collaborator's publications)