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

Send correspondence to pazoulay@mit.edu. Part of the work was performed while the first author was an Alfred P. Sloan Industry Studies Fellow. We thank 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 Faculty Data Systems and Studies). The National Institutes of Health partially supports the AAMC Faculty Roster under contract N01-OD-3-1015. The usual disclaimer applies. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

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

ABSTRACT

We estimate the magnitude of spillovers generated by 161 academic "superstars" onto their collaborators' research output. These life scientists died while still being actively engaged in science, 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 10% decline in their quality-adjusted publication rates. By exploring interactions of the treatment effect with a wide range 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 ideas space, but not in physical or social space. 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
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts
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 2005 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 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 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

Specifically, we analyze changes in the research output of collaborators for 161 eminent scientists who die prematurely. 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 younger scientists exhibiting sudden bursts of productivity. Using the AAMC Faculty Roster as an input — a comprehensive, longitudinal, matched employee-employer database pertaining to 222,555 faculty members in all U.S. medical schools between 1975 and 2006, we construct a panel dataset of 8,220 collaborator-star dyads, and we examine how coauthors' scientific output — as measured by publications, citations, and 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 5 to 10% decrease in the quality-adjusted publication output of coauthors in response to the sudden and unexpected loss of a superstar. When the superstar death is anticipated and thus less plausibly exogenous, our results are weaker but generally consistent with the effects due to unanticipated losses. Furthermore, the impact of star death extends across coauthors of varying talent — only those who had achieved wide recognition at the time of death appear immune to the effect of superstar extinction. We also show that the magnitude of the effect is increasing in the star's scientific eminence at the time of death.

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

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. (2008).

by Alfred Marshall (1890) and Robert Lucas (1988).[3] Should the effects of exposure to superstar talent be interpreted as laying bare the presence (or in our case, absence) of knowledge spillovers? Evidence of adverse coauthor outcomes is not sufficient to seal the case. Since we identify 51 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.

One intuitive story centers around skill substitution within ongoing collaborative teams. However, we find that recent collaborations — for which this concern is presumably most relevant — are not driving the effect, a fact at odds with this class of explanations. A jaundiced view of the academic reward system provides the backdrop for a second 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 co-located coauthors, or on coauthors of stars well-connected to the NIH funding apparatus. 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. We also find that former trainees — who may be more vested in the star's domain of expertise — suffer a loss twice as large in magnitude, relative to that experienced by non-trainees. 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. Section 3 provides descriptive

[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.

statistics at the coauthor and dyad level. We lay out the econometric methodology and report the results in section 4. Section 5 concludes.

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).

We now provide a detailed description of the process through which the matched coauthor/superstar data used in the econometric analysis was assembled. In order, we describe (1) the criteria used to select our sample of "extinct" superstar life scientists, along with basic demographic information; (2) the universe of potential colleagues for these superstars; and (3) the matching procedure implemented to identify colleagues from coauthorship records.

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 — though still surprisingly difficult — 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 8,963 "elite" life scientists (roughly 5% of the entire relevant labor market) who are so classified if they satisfy at least one of the following seven criteria for scientific achievement:

- **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 years 1977 to 2006, deflating the earlier years by the biomedical research producer price index. 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.[4]

- **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

[4]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.

250 most cited researchers for their published articles between 1981 and 1999, within a broad scientific field.[5]

- **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.** Finally, we add to these groups academic life scientists who were elected to the National Academy of Science between 1975 and 2007.

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 bursts of productivity endure for long periods of time.

- **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

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

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.[6]

- **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.

Many among these 8,963 scientists achieve elite status according to more than one metric. We trace back their 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 161 scientists who are the focus of this paper constitute a subset of this larger pool of 8,963. We impose several additional criteria to derive the final list. First, the scientist's death must intervene between 1982 and 2003. This will enable us to observe at least 3 years'

[6]See Azoulay et al. (2008) for more details and an evaluation of this program.

worth of scientific output for every colleague 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). Finally, 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 screen is somewhat subjective, but we will validate below our contention that the final set is limited to scientists that are "research-active" at the time of their death.

Appendix Tables 1A and 1B present individual-level details on the extinct superstar sample, broken down by sudden and anticipated deaths, respectively. Heart attack is the most frequent cause of sudden death, while the vast majority of anticipated deaths are due to cancer. Of course, the case for exogeneity is weakest in the anticipated case, since coauthors might alter their collaboration strategies even before the superstar's passing.[7]

Table 1A provides descriptive statistics for the superstar sample. The average star received his degree in 1963 (min.=1940; max.=1986), died at 58 years old (min.=38; max.=67) and worked with 51 coauthors during his lifetime (min.=5; max.=177; the histogram for the distribution of this variable can be found in Figure 2). On the output side, the stars each received an average of roughly 9.5 million dollars in NIH grants (excluding center grants; min.=$0; max.=$65 million), and published 156 papers (min.=20; max.=528) that garnered 9,041 citations (min.=282; max.=34,625) as of early 2008.[8] Though we do not display separately these descriptive statistics for the scientists whose deaths were anticipated and for those whose deaths were sudden, the two subsamples are virtually identical.

Table 1B provides additional information about the extinct superstar sample. The sample is approximately 10% female and 84% US-born. 45% of our stars hold an MD degree, 42% a PhD, and the remainder hold dual MD/PhD degrees. Keeping in mind that our met-

[7]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.

[8]We also compute the h index due to Hirsch (2005), which is commonly used by bibliometricians: h is the highest integer such that an individual has h publications cited at least h times. In our sample, h is approximately 50 (min.=9; max.=111).

rics of superstardom are not mutually exclusive, roughly 6.8% are Howard Hughes Medical Investigators, 31.7% are MERIT awardees, and 20.5% are members of the National Academy of Sciences.

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.[9] When all cross-sections are pooled, we obtain a matched employee/employer panel dataset. For each of the 222,478 faculty members that appear in the roster, we know the full name, the type of degrees received and the years they were awarded, gender, up to two departments, and medical school affiliation. An important implication of our reliance on the AAMC Faculty Roster is that the interactions we can observe in the data take place between faculty members, rather than between faculty members and trainees (graduate students or post-doctoral fellows).[10]

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.[11]

[9] Although 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.

[10] We do not mean to suggest that mentor effects — within-lab interactions — are unimportant or less worthy of study. But the study of mentor imprinting effects will require alternative empirical approaches and data sources (see Azoulay et al. [2008] for a relevant example). 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.

[11] 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

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 161 extinct superstars. An alternative approach to identify potential recipients of spillovers is to rely on shared departmental affiliation (e.g., Kim et al. 2006). The use of departmental boundaries to identify "relevant" peers is fraught with interpretive difficulties in our setting. First, department affiliations are not fixed over time for most faculty — this is apparent in our sample of superstars, who frequently switch departments when they hop from one employer to another. It is also reflected in the fact that many collaborations span departmental boundaries. Second, new departments were created during the time period we study (e.g., neuroscience, genetics, or biomedical engineering), while others were phased out or dramatically shrunk (e.g., anatomy). Third, the merging or survival of many departments is often a reflection of internal political struggles, rather than characteristics of the research conducted within them. For example, in some medical schools, orthopedic surgeons are housed in a separate department while in others, they are part of a large surgery department; in the basic sciences, many faculty would feel equally at home in cell biology, microbiology, or immunology. Finally, three large departments (internal medicine, pediatrics, and surgery) tend to account for a large proportion of medical school employment, but their size masks enormous heterogeneity (e.g., neurosurgeons vs. cardiothoracic surgeons; endocrinologists vs. infectious diseases specialists, etc.).

From a substantive standpoint, the ways in which scientists conceive of their epistemic community reflects the advanced degree of specialization that is prevalent in most fields of the life sciences. A physiologist working on protein trafficking between intracellular organelles will not necessarily be influenced by developments in physiology writ large, but will certainly follow closely the research of those working in his/her specific area of inquiry. In our setting, the set of past and present coauthors of a prominent scientist represents a reasonable first-order approximation of the relevant intellectual community. Since not all coauthors share

schools from (or to) other types of research employment. For these faculty, we were successful in filling career gaps by combining the AAMC Roster with the NIH data.

identical attachment to the same topic area, we will also measure distance in intellectual space to refine our definition of coauthors.

2.3 Coauthor Matching

To identify coauthors, we have developed a software program, the Stars/Colleague Generator, or S/CGen.[12] 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 validated coauthor, and generates publication counts as well as coauthorship variables for each superstar/colleague dyad, in each year. In Appendix I, we provide a great deal of detail on the matching procedure, how we guard against the inclusion of spurious coauthors, and how we deal with measurement error when tallying the publication output of coauthors with common names.

2.4 From One to Two Levels of Difference: Control Superstars and Control Coauthors

Our original research design called for identifying the effect of superstar exposure by examining changes in coauthor research output after the superstar passes away. With a single level of difference, we rely on the coauthors of stars who have not yet died as an implicit control group to pin down life cycle and calendar year effects. This will provide estimates that can be given a causal interpretation under fairly general assumptions regarding the exogeneity of the

[12]The software can be used by other researchers under a open-source (GNU) license. It can be downloaded, and detailed specifications accessed, from the SC/Gen web site `http://stellman-greene.com/SCGen/`. Note that the S/CGen takes the faculty roster as an input; we are not authorized to share this data with third-parties. However, it can be licensed (for a fee) from AAMC, provided a local IRB gives its approval and a confidentiality agreement protects the anonymity of individual faculty members.

death event. However, the simple before/after contrast 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. With a single level of difference, this dyad-specific, time-varying omitted variable will not be fully captured by collaborator age controls. The standard approach to this type of estimation challenge is to rely on both a treatment and a control group, and to estimate the effect using a difference-in-differences (hereafter DD) framework. This is the approach we implement here, though the construction of the control group of superstar/coauthor dyads faces two serious obstacles, the first practical, the second substantive.

From a practical standpoint, the most natural procedure would be to match — using a set of observable dyad-level variables — the treatment dyads with control dyads formed by superstar scientists who do not die and their coauthors. In practice, millions of such dyads exist, and the task of culling from this vast universe a relatively small number to serve as controls is computationally non-trivial. We resolve the issue by proceeding in two steps. First, we select a set of 161 control superstars from among our elite sample of 8,963, by matching on star demographic and time-varying observables. Second, we add to the estimation samples all the dyads formed by these control stars and their coauthors. We provide details on the procedure in Appendix II.

From a substantive standpoint, the members of a valid control group should be unaffected by the treatment of interest. This is an ideal that we can only hope to approximate in this setting. No scientist is an island. The set of coauthors for our 8,963 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 "ideas space." As a result, we take three pragmatic steps to avoid obvious sources of contamination. First, when matching on observables to recruit control superstars, we do so in the subsample of 3,979 elite scientists who do not have any

coauthorship tie with any of the 161 extinct superstars. Second, the list of observables we rely on for matching does not include variables that could be correlated with scientific field, such as department. Third, when estimating the treatment effects in the DD framework, we eliminate from the estimation sample (a) treatment coauthors who also collaborate with control superstars; and (b) control coauthors who also collaborate with treatment superstars. As a result, the minimum path length in the coauthorship network between a control coauthor and a treated coauthor is 3 if we constrain the paths to pass through at least one of our 8,963 superstars (see Figure 1). More details are provided in Appendix II.

3 Descriptive Statistics

It is appropriate to describe the data at two different levels — that of the individual coauthor, and that of the superstar/coauthor dyad.

3.1 Colleague characteristics

When applied to our sample of 161 extinct superstars, S/CGen identifies 7,111 distinct coauthors with unique PubMED names. When applied to the set of $161 + 161 = 322$ extinct and control superstars, S/CGen identifies 12,162 distinct coauthors, 2,166 (17.11%) of whom are problematic in the sense that they collaborate with at least one extinct superstar and one control superstar. The descriptive statistics we provide in Tables 2A, 2B, and 2C pertain to the set of $12,162 - 2,166 = 10,496$ coauthors we can unambiguously assign to either the treatment or the control group.

Demographic characteristics for the coauthors are presented in Table 2A. 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; slightly less than two-thirds serve as a principal investigator on at least one NIH grant over the course of their careers, and a third are affiliated with basic science departments (as opposed to clinical or public health departments). The coauthors are about 10 years younger than the superstars on average (1974 vs. 1964 for the year of highest degree).

Table 2B shows that coauthors lag behind superstars in terms of publication output, but the difference is not dramatic (92 vs. 155 articles, on average). Relative to the superstars, their collaborators have had more time to accumulate publications, since their academic careers were not cut short by an untimely death. Furthermore, assortative matching is present in the market for collaborators, as reflected by the fact that 2,101 (20%) of our 10,496 coauthors belong to the elite sample of 8,963 scientists.

3.2 Dyad characteristics

The 7,111 treatment coauthors mentioned above participate in 8,220 extinct superstar-coauthor dyads. Conversely, the 10,496 treatment and control coauthors generate 11,570 dyads (5,564 control dyads and 6,006 treatment dyads). Table 3 provides descriptive statistics on these 11,570 dyads in the year of superstar extinction.[13] Further, 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).

Dyad-level 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 3). We define "casual" dyads as those that have two or fewer coauthorships with the star, "regular" dyads as those with three to ten 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 10 years before the star's death, and time since last coauthorship is slightly more than 8 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

[13]For control dyads, the year selected is the year of death for the extinct superstar who is the control superstar's "nearest neighbor." See Appendix II for further details.

in the three years preceding the passing of the superstar) map into the top quartile of collaboration recency at the dyad level (Figure 4).

The research collaborations studied here occur between faculty members, who often run their own labs (a conjecture reinforced by the large proportion of coauthors with independent NIH funding). Yet, it is interesting to distinguish collaborators who trained under a superstar (either in graduate school or during a postdoctoral fellowship) from those collaborations initiated at a time in which both nodes in the dyad already had a faculty appointment. 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.[14] Using this method, we find that slightly more than 9% of dyads involve a former trainee.

We now examine the spatial distribution of collaborations. Slightly less than a quarter of collaborations correspond to scientists who shared an institutional affiliation 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 5. Further details on its construction are provided in Appendix III.

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 dyads

[14]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.

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. Slightly fewer than 8% of the dyads correspond to this situation of "accidental coauthorship" — the most tenuous form of collaboration.

Coauthor-level variables. We mentioned above that a large number of coauthors were themselves quite eminent, but this assessment was based on their career accomplishments. Here, we focus specifically on their achievements as of the year of superstar extinction. 55% of dyads include a collaborator who has served as PI on at least one NIH R01 grant when the superstar passes away, while about 6% include a collaborator who belongs to a much more exclusive elite: Howard Hughes Medical Investigators, members of the NAS, or MERIT awardees. As explained in Appendix II, the coauthors of extinct superstars are slightly more accomplished than the coauthors of control superstars (72 vs. 65 publications at the time of death), but half of the gap reflects the deletion from the sample of coauthors associated with both extinct and control superstars.

The estimation sample pools observations between 1975 and 2006 for the dyads described above. The result is an unbalanced panel dataset with 216,746 dyad/year observations (treatment dyads only) or 294,463 dyad/year observations (treatment and control dyads).

4 Results

The exposition of the econometric results proceeds in five stages. After a brief review of methodological issues, we validate our earlier contention that the sample of extinct superstars is composed of individuals who remained actively engaged in science at the time of their death. Second, we provide results that pertain to the main effect of superstar exposure on publication rates, citations, and NIH grants. Third, we use the heterogeneity within the set of extinct superstars to ascertain whether the treatment effect is in fact a "superstar effect," in the sense that it would disappear if we focused instead on coauthors for a sample

of unexceptional scientists who die. Fourth, we investigate whether some coauthors — those of especially high or low status — are more affected by, or completely immune to, the effect of superstar extinction. Fifth, focusing solely on publication rates, we attempt to explicate the mechanism, or set of mechanisms, responsible for the main effect of superstar extinction. We do so by exploring heterogeneity in the treatment effect; in practice, we interact 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:

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

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 seven indicator variables corresponding to different career age brackets, where career age measures the number of years since a scientist earned his/her highest doctoral degree (MD or PhD).[15]

[15]The omitted category corresponds to faculty members in the very early years of their careers (age 0 to 4). 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. 2005).

Estimation. The dependent variables of interest, including weighted or unweighted publication counts and NIH grants awarded, are skewed and non-negative. For example, 20.82% of the dyad/year observations in the data correspond to years of no publication output; the figure climbs to 87.39% if one focuses on the count of successful grant applications. 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).[16]

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. Yet, the residuals for the observations corresponding to dyads involving different coauthors but the same superstar at a point in time cannot be considered independent. As a result, we cluster the standard errors around superstar scientists in the results presented below.[17]

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. Contrasting the effects of superstar extinction on both measures of output provides

[16] In Appendix Table 5, we show that OLS yields very similar results to QML Poisson estimation for our main findings.

[17] In contrast, we ignore clustering around coauthors, since of 91% of coauthors collaborate with a single superstar *in the sample*, though they might well collaborate with other elite scientists (see Table 2C).

insight on the extent to which collaborators substitute towards other relationships following the passing of their prominent coauthor.

We perform a crude 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. Although we do not have access to article-level citation data for the universe of all coauthors, this fine-grained level of detail is available for the sample of 8,963 elite scientists. These individuals represent 20% of coauthors, and a third of the superstar-coauthor dyads in the full 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*.[18] 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 A2). 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.

We also rely on NIH grant information. We focus on the number of research grants, rather than their amounts. 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. In addition, funding levels are strongly influenced by scientific specialization and the associated costs of research in that area, thus making them an inferior measure of success in our context. 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

[18]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.

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.

4.2 Trends in publication output in the years immediately preceding a superstar's death

In Table 4, we present results for specifications in which the quality-adjusted publication output of our superstars is regressed onto a series of indicator variables corresponding to the timing of death: 4 years before the year of death, 3 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). We stack the deck in favor of finding preexisting trends in the data by using solely the publications in which the star appears as last author (last author status is invariably reserved to the head of a laboratory/research group in the life sciences). All models include superstar scientist fixed effects, and we use as a control group the set of control superstars described above. 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.

All estimates are presented in the form of incidence rate ratios; the formula $(e^{\beta}-1)\times 100\%$ (where β denotes an estimated coefficient) provides a number directly interpretable in terms of elasticity. Model (1), for instance, implies that output falls $1 - 0.76 = 24\%$ in the year following the year during which the superstar passed away. Column (1) uses all the data, and uncovers no evidence of decline in output prior to the superstar's death. In column (2), we drop the 92 colleagues whose death was anticipated, and obtain similar results. Conversely, column (3) excludes 69 superstars whose death was sudden and unexpected. The indicator variables are therefore identified solely off of the superstars whose deaths were anticipated. In this case, we find statistically significant evidence that scientists who died exhibited *higher* output in the year immediately preceding their death. As a sensitivity check, column (4) focuses on a group of 30 older superstars who died on or after they had reached 75 years of age, but before retirement. For these scientists, we observe declines in the number of publications

prior to the scientist's demise; although the magnitudes are economically meaningful, these effects are not estimated precisely. In light of these results, we feel confident that our 161 extinct scientists were still actively engaged in science at the time of their deaths. However, the analysis also calls into question the assumption that the anticipated death of a superstar is an exogenous event from their coauthors' point of view.

4.3 Main effect of superstar extinction

Table 5 presents our core results. Column 1a in Panel A examines coauthors' total JIF-weighted publication output, regardless of cause of death for the superstar. We find a sizable and significant 7.4% decrease in the total number of quality-adjusted publications coauthors produce after the star dies. Columns 2a and 3a break down the effect by underlying cause of death. In both cases, we find statistically significant effects, though the coefficient estimate for the sudden death case is roughly twice the magnitude of that corresponding to the anticipated death case (10.5% vs. 4.7%). Columns 1b, 2b, and 3b examine whether coauthorship intensity moderates the size of the main effects described above. We find that close and regular collaborators suffer larger declines in output, relative to casual collaborators, but these differential impacts, while often sizable in magnitude, are not themselves statistically significant.

Panel B provides 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 results imply that "close" and, to a lesser extent, "regular" coauthors do manage to find replacement collaborators (or to intensify already existing collaborations). Close collaborators experience between a $(1-0.958)+0.152 = 11\%$ (column 6b) and $(1-0.927)+0.211 = 13.8\%$ (column 5b) increase in their quality-adjusted publications written independently of the star, but this is only a partial offset for the overall loss documented in Panel A. In every case, we find that casual coauthors see their indepen-

dent output decline between 4.2% (column 6b) and 7.3% (column 5b). Though we defer the interpretation of the results to the next section, we note that this last result strongly hints at the presence of spillovers.

Panels C and D present specifications that mirror those in Panels A and B in all respects, except that the estimation sample includes the set of control dyads. The inclusion of the control group attenuates slightly the magnitudes of the treatment effect, especially in the unanticipated death case, which is no longer statistically significant. However, the results are qualitatively similar to those relying on a single level of difference.

We also explore the timing of the effects uncovered in Panel A of Table 5. 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 before or after the superstar's death, and then graphing the effects and the 95% confidence interval around them (Figure 6, corresponding to Table 5A, column 1a). We find evidence of a very slight preexisting output trend, but it is imprecisely estimated.[19] 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 presumably 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.

As mentioned in section 4.1, 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 6 answers this question by modeling the effect of superstar extinction for the production of articles falling above various quantiles of the citation distribution. An

[19] In contrast, for the case of sudden death (not displayed), there is absolutely no hint of a preexisting trend — statistically significant or not.

important caveat is that the results pertain only to the set of coauthors that are part of our elite group of 8,963 scientists, since this is the set for which article-level citation data is available. We find remarkable stability of the treatment effect as one goes further up the tail of the citation distribution to compute the dependent variable. The magnitude of the effect remains in the 3.5% to 6.1% range, whether we examine the effect of superstar extinction on raw number of publications, publications that fall above the median number of citations, publications in the top quartile, or publications in the top ventile. The magnitude of the effect increases to 10% when focusing on "blockbuster" publications — those falling in the top percentile of the citation distribution. Adding the control group produces estimates of very similar magnitudes. At the very least, these results suggest that superstar exposure is not limited to the production of relatively less significant scientific knowledge.

Our exploration of the main effect of superstar extinction concludes with an analysis of NIH grant outcomes (Table 7). The number of observations decreases by about a third, since the coauthors who receive no grants during the observation period fall out of the estimation sample. The magnitudes of the effects are strikingly similar to those observed in Table 5, but the statistical significance of the estimates is weaker. When broken down by underlying cause of death for the superstar, the main effects are not significant (columns 2a and 3a); they are statistically significant when pooling all superstars (column 1a), but only at the 10% level. When adding the control dyads, we find no effect for casual and regular coauthors, though we find a surprisingly large effect for close coauthors (column 4b). 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. Furthermore, some of these grants might list the superstar scientist as a co-investigator, but the data do not allow us to distinguish independent from collaborative grantsmanship.

4.4 Does the treatment effect increase with the star's scientific achievements?

Table 8 investigates whether the magnitude of the treatment effect depends on the accomplishments of the star. We find that the (negative) impact of star death is increasing in

quartiles of the superstars citation count at the time of their death (columns 1a and 1b), a pattern that becomes clearer when citations are adjusted for career length (columns 2a and 2b). In contrast, there is no monotonic pattern of this type when we rank the superstars by quartile of cumulative NIH funding (columns 3a and 3b) or by quartile of cumulative NIH funding normalized by years of career (columns 4a and 4b). Together, these results suggest 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. They also 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 observe here. As such, the results in Table 8 validate *ex post* our pragmatic focus on the effect of superstars.

4.5 Impact of coauthor status

In Table 9, we interact the treatment effect with three indicators of coauthor status, to ascertain which collaborators, if any, are insulated from the effects of superstar extinction documented earlier. Columns 1a and 1b focus on faculty members whose sole elite collaborator was the superstar who died. For these coauthors with relatively poor substitution opportunities (they account for roughly 20% of the dyads in the sample), the consequences of the superstar's loss are disastrous, with an overall 30 to 35% decline in publication output. Columns 2a and 2b 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. In column 2a, for example, the differential effect is an imprecisely estimated 0. Of course, this also means that independent NIH funding is not enough to insulate scientists from the loss of an eminent collaborator. In columns 3a and 3b, we present evidence that the "elite among the elite" — members of the National Academy of Science, Howard Hughes Medical Investigators, and NIH MERIT awardees — is relatively unaffected by the loss of a "peer superstar." These eminent coauthors account for 9% of the treatment dyads (column 3a) and 6% of the treatment and control dyads combined

(column 3b). The differential impact on elite coauthors is always positive and statistically significant (though only at the 10% level when control dyads are excluded), and the overall treatment effect ranges from −2.8% (column 3a) to +4.9% (column 3b).

We conclude that the effect of superstar extinction is heterogeneous with respect to coauthor status, but also very broad. 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.[20]

4.6 Disentangling Mechanisms

In Tables 10, 11, and 12, we exploit the richness and fine-grained level of detail in the data to sort between alternative mechanisms which might underlie the superstar extinction effect. We seek to adjudicate between four broad classes of explanations.

Imperfect Skill Substitution [ISS]. The basis for ISS is the idea that collaborative research teams emerge to pool the expertise of scientists, who, in their individual capacity, face the "burden of knowledge" problem identified by Jones (2005). 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.[21] The ISS story carries three testable implications. First, one would expect the effect to be temporary: even if the projects that necessitate the extinct superstar's intellectual input are slowed down or, in the worst case, aborted, coauthors should eventually turn their attention to new projects and/or new collaborators. Second, we would expect the effect to be more pronounced in the case of sudden death; if the star's passing is anticipated, s/he might be able to lay the ground work necessary for his/her coauthors to effectively plug the knowledge gaps opened by his impending exit. This assessment is somewhat tempered by the fact that life scientists generally have apprentices (graduate students and postdocs) who might be of considerable assistance to ensure the continuity of

[20] As seen in Table 6, taken as a whole, the set of elite coauthors suffers a decline in output very similar to the one observed for the universe of all coauthors (i.e., in Table 5). At the risk of repeating ourselves, the elite sample is very heterogeneous, and does include young, old, and fading stars.

[21] We acknowledge that, under this scenario, using the term "spillover" to characterize the effect of the superstar would stretch the common meaning of the term.

the team's research efforts. Third, and most importantly, the ISS mechanism pertains only to ongoing or recent collaborations.[22]

Table 10 addresses the last of these implications, by examining how the age and recency of collaboration moderates the treatment effect. The results are unambiguous: not only are young and recent collaborations not driving the effect, the reverse is in fact true. Columns 1a and 1b look at the differential effect corresponding to dyads that have produced at least one joint publication in the three years preceding the star's death (the top quartile of collaboration recency). The interaction effect has a positive sign, and implies that recent collaborators see a slight uptick in their output following the star's death. Columns 2a and 2b focuses on the interaction with collaboration age, i.e., the length of time between the dyad's first coauthorship and the superstar's death. The results are almost identical. It is important to note that these findings are not mere artifacts of collaborator age. In fact, columns 3a and 3b show that no strong pattern emerges when the main effect is interacted with the age of the coauthor.[23] Taken as a whole, the results in this table are strikingly inconsistent with the ISS mechanism.

Superstars as Gatekeepers. An alternative story is that superstars matter for their coauthors because they connect them to important resources, such as funding, editorial goodwill, or other potential collaborators. We attempt to evaluate the validity of three particular versions of this story in Table 11, but we acknowledge that many other versions of it are equally plausible.

In a first step, we examine whether shared physical location plays any role in heightening the effect of the superstar's death. We might expect stars' protective reach to be geographically circumscribed. For instance, in the case of anticipated death, the star might advocate

[22] An additional implication is that it is coauthors with the least amount of intellectual overlap with the star who should experience the most adverse effects, since their own expertise is a poor substitute for the star's. As we will see below, the reverse is in fact true.

[23] If anything, the magnitude of the effects are slightly larger for younger coauthors, though we hesitate to make too much of the apparent non-monotonicity with respect to coauthor age. One might also worry that relationships that lay fallow correspond to adversely selected coauthors, who have been written off by the star as having low added value. In unreported specifications, we verified that the dynamics of the treatment effect (i.e., Figure 6) are not appreciably different if we exclude from the sample recent collaborations.

on behalf of a local collaborator with senior administrators. If stars are generally involved in the management of their "succession," we might expect the effect of superstar extinction on co-located coauthors to be more pronounced. We use two alternative definitions of physical proximity: co-location (i.e., same institutional affiliation; columns 1a and 1b), or less than 10 miles separating the nodes of the dyad at the time of death (columns 2a and 2b). In both cases, the differential impacts are very small and statistically insignificant. Physical proximity might be important to explain the initiation of collaboration; but once established, the deletion of a local coauthorship tie appears just as debilitating as the deletion of a distant one.

Whereas social scientists sometimes emphasize the role that benevolent 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 study sections, or has coauthorship ties with other scientists who sit on study sections. In columns 3a and 3b, we find that this is not the case. Once again, the differential impacts are small and imprecisely estimated.

Finally, 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 our extinct superstars in the coauthorship network formed by the 8,963 elite scientists.[24] 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 (columns 4a and 4b). Although we find that collaborators with stars in the bottom quartile of centrality are insulated from any adverse impact, the other interaction terms are not ordered monotonically.

[24]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. Empirically, we find that betweenness is heavily correlated with publication output. Therefore, our measure of betweenness is the residual of a regression of betweenness onto the cumulative number of publications for the star and year effects.

The differential effects for stars in the second quartile are larger in magnitude, relative to those corresponding to the third and fourth quartile.

The evidence presented in Table 11 appears broadly inconsistent with the three particular access stories whose implications we could test empirically. Our assessment of the gatekeeping mechanism must remain guarded, since it is possible to think of variations on the same theme. For instance, 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 is the most relevant way in which superstars influence their collaborators' scientific output.

Ascription. Sociological studies of the scientific reward system have provided some evidence supporting the existence of the "Matthew Effect,"[25] 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 for our setting is doubtful for two reasons. First, in the case of casual coauthors (who account for 75% of the dyads), we observe a decline in the output written independently of the star (Tables 5B and 5D). Second, the treatment effect on total publication output is driven by dyads that do not have recent collaborations, but its onset is delayed until after the death of the star. These two facts argue against an interpretation of the effect based on ascription.

The "Invisible College" hypothesis. Finally, we examine the possibility that stars generate spillovers on the entire group of scientists who work on similar research problems, which we refer to as the "invisible college" hypothesis, in reference to early studies of scientific communication by sociologists (e.g., de Solla Price and Beaver 1966; Crane 1972). In this literature, the invisible college refers to an elite of productive scientists highly visible in a

[25] *"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]

research area, combined with a "scatter" of less eminent ones, whose attachment to the field may be more fleeting. The empirical question we tackle in Table 12 is whether the death of superstar scientists provide a unique opportunity to reveal the boundaries of the invisible colleges to which they belong. To do so, we construct two dyad-level variables that help us measure the proximity of the superstar and his/her coauthors in ideas space.

First, in columns 1a and 1b, we look for a differential effect of superstar death for coauthors that were also former trainees. In many fields of science, the mastery of experimental techniques and other methodologies are difficult to learn without direct contact with those already skilled in them. Scientists who trained under a superstar, either in graduate school or during a postdoctoral fellowship, are likely to be more deeply vested in the scientific problems that have informed the star's scientific trajectory. The data indeed provide evidence of a strong former trainee effect, with a magnitude double that of non-trainees. This suggests that mentorship continues into the early faculty career and is extremely important in intellectual development. Of course, an interpretation of this effect based on social — rather than intellectual — proximity could be advanced.

Our next step is to measure proximity in "ideas space" between members of a collaboration dyad, and to ascertain whether coauthors who are intellectually proximate in the years immediately preceding the star's death suffer more severe consequences following the event, relative to intellectually distant ones. Using quartiles of normalized keyword overlap, we find that this is the case. In columns 2a and 2b, the effect corresponding to the top quartile of overlap is highly statistically significant, the largest in absolute value, and roughly double the size of the effects corresponding to the bottom three quartiles.

This evidence is very consistent the invisible college hypothesis, but leaves an important question unresolved: is the act of formal coauthorship necessary for a scientist to be brought into a superstar's intellectual orbit? It may not be necessary to be socially acquainted with a particular member of the college in order to be influenced by her. Since our sample is composed exclusively of coauthors, we cannot answer this question in a definitive fashion. Yet, we have used the norms of authorship in the life sciences to try to isolate collaborators

whose coauthorship tie to the star is particularly tenuous. These "accidental" coauthors are those who, whenever they collaborate, find themselves in the middle of the authorship list. In its purest form, the invisible college hypothesis would imply no differential effect for these colleagues, and this is what we find in columns 3a and 3b.[26]

We conclude that the overall collection of results presented above are most consistent with the invisible college hypothesis. Superstar scientists make their field of inquiry visible to others of lesser standing who might enter it, however briefly; 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, the most important interactions for the production of new scientific knowledge are not constrained by geographic or social space. Rather, they take place in an ethereal, amorphous, but essential space of ideas.

4.7 Sensitivity Checks

In Table 13, we present the results of a number of robustness and sensitivity checks. We begin by manipulating the age cutoff we imposed when constructing the sample of extinct superstars. In columns 1a and 1b, we limit the sample to the 86 stars who were 60 years old or younger at the time of their death. We find very similar results to those displayed in Table 5A. The same holds true when we use a 70 years of age cutoff (columns 2a and 2b). Conversely, columns 3a and 3b show no effect of superstar extinction when examining the impact of 30 eminent scientists who die beyond the highly creative stages of their career — at 75 years of age or older.

Finally, we perform a small simulation study to dispel any remaining doubt regarding the validity of the quasi-experiment exploited in the paper. We generate placebo dates of death among our control superstars, where those dates are drawn at random from the empirical distribution of death events across years for the 161 extinct superstars. We then replicate the specifications in Tables 5A and 5C, column 1a, but we limit the estimation sample to the

[26] An important caveat is that "accidental" dyads necessarily imply collaboration between three or more laboratories; the matching process generating these accidental matches might differ from the process giving rise to collaborations which imply a higher degree of social closeness.

set of 5,564 control dyads. Column 4 presents the average treatment effect, z-statistic, and log quasi-likelihood over 100 replications. Reassuringly, the effect of superstar extinction in this manufactured data is an imprecisely estimated 0.

5 Conclusion

We examine the role of collaboration in spurring the creation of new scientific knowledge. Using the premature deaths of "superstar" academic life scientists as a quasi-experiment, we find that their coauthors experience a sizable and permanent decline in quality-adjusted publication output following the event. Only coauthors who themselves had achieved wide recognition at the time of death appear insulated from the effect of superstar extinction. In addition, the magnitude of the treatment effect increases monotonically with the stars' eminence, suggesting that the intellectual vacuum arising from their deaths is commensurate with their scientific contributions.

We attempt to adjudicate between plausible mechanisms that could give rise to the extinction effect. Recent collaborations are not affected, which seems inconsistent with explanations emphasizing skill substitution within ongoing collaborative teams. We also find no differential impact on co-located coauthors, or on coauthors of stars well-connected to the NIH funding apparatus. These findings do not jibe with interpretations stressing the gatekeeping role of eminent scientists. Lastly, we show that coauthors proximate to the star in "ideas space" suffer a more pronounced decline in output, relative to coauthors who are more distant. Overall, our results are consistent with the idea that part of the scientific field to which the star contributes dies along with him, because the fount of scientific knowledge from which coauthors can draw is greatly diminished.

The lack of differential impact for co-located collaborators is especially surprising in light of the prior literature on this topic. Zucker et al. (1998) establish a robust correlation between the location of star biologists and the number of new biotechnology firms spawned in a given locale. Ham and Weinberg (2008) finds that among physicists, geographic proximity to earlier Nobelists correlates with the start of the research that eventually results in a

Nobel Prize. However, this result is consistent with a recent body of evidence 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). While it is tempting to ascribe this development to recent decreases in the cost of scientific communication, Waldinger (2008) finds no evidence of localized spillovers even in an earlier era. Using the dismissal of Jewish faculty members from German universities in the 1930s as a source of exogenous variation in peer group composition, he uncovers no effect of the dismissal on faculty members in the same department and university. In contrast, he observes suggestive effects on collaborators — a result strikingly consistent with the evidence we present. Together, our findings suggest that instead of manipulating physical or social space, future work in this area could usefully focus on identifying quasi-experiments in intellectual space. For instance, how do scientists adjust to sudden changes in scientific opportunities in their field?

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."*

References

Agrawal, Ajay K., and Avi Goldfarb, "Restructuring Research: Communication Costs and the Democratization of University Innovation," *American Economic Review*, 98 (2006), 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 (2008).

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

Azoulay, Pierre, Joshua Graff Zivin, and Gustavo Manso, "Incentives and Creativity: Evidence from the Academic Life Sciences," Working Paper, MIT Sloan School (2008).

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).

Benner, Mary, and Joel Waldfogel, "Close to You? Bias and Precision in Patent-based Measures of Technological Proximity," NBER Working Paper #13322 (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.

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

Costa, Dora L., and Matthew E. Kahn, "Cowards and Heroes: Group Loyalty in the American Civil War," *Quarterly Journal of Economics*, 118 (2003), 519-548.

Crane, Diana, *Invisible Colleges: Diffusion of Knowledge in Scientific Communities* (Chicago: University of Chicago Press, 1972).

de Solla Price, Derek J., *Little Science, Big Science* (New York: Columbia University Press, 1963).

de Solla Price, Derek J., and Donald D. Beaver, "Collaboration in an Invisible College," *American Psychologist*, 21 (1966), 1011-1018.

Fafchamps, Marcel, Sanjeev Goyal, and Marco van de Leij, "Matching and Network Effects," 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).

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," NBER Working Paper #11739 (2005).

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: Cambridge University Press, 1999), pp. 267-311.

Hirsch, Jules E., "An Index to Quantify an Individual's Scientific Research Output," *Proceedings of the National Academy of Sciences*, 102 (2005), 16569–16572.

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.

Jones, Benjamin F., "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?," NBER Working Paper #11360 (2005).

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).

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

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: MacMillan, 1890).

Mas, Alexandre, and Enrico Moretti, "Peers at Work," NBER Working Paper #12508 (2006).

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

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).

Powell, Walter W., Douglas R. White, Kenneth W. Koput, and Jason Owen-Smith, "Network Dynamics and Field Evolution: The Growth of Inter-organizational Collaboration in the Life Sciences," *American Journal of Sociology*, 110 (2005), 1132-1205.

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.

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).

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).

Table 1A: Summary Statistics for Superstars

	Mean	Std. Dev	Min.	Max.
Controls (N=161)				
Birth Age at Death	57.06	7.34	40	78
Career Age at Death	30.30	7.39	13	52
Degree Year	1964.81	8.39	1936	1986
Year of Birth	1938.04	8.01	1912	1956
# Positions	2.08	1.11	1	5
# Coauthors	46.49	35.04	2	177
NIH funding	$9,791,164	$11,479,553	$0	$100,083,848
# Papers	153.10	96.39	20	553
# Citations	8,709	7,524	738	58,875
# Citations/Paper	58.69	40.51	13.73	331.42
h index	52.22	20.18	13	130
Extinct (N=161)				
Birth Age at Death	58.60	6.64	38	67
Career Age at Death	31.76	7.33	7	45
Degree Year	1963.34	9.05	1940	1986
Year of Birth	1936.50	8.12	1916	1959
# Positions	1.79	0.93	1	6
# Coauthors	51.06	34.80	5	177
NIH funding	$9,590,272	$9,640,493	$0	$64,927,440
# Papers (total)	156.07	101.17	20	528
# Citations	9,041	6,978	282	34,625
# Citations/Paper	58.34	29.05	8.06	189.45
h index	49.70	20.08	9	111

Note: For control superstars, the (counterfactual) year of death is thatof his "nearest neighbor" among extinct superstars. See Appendix II for further details.

Table 1B: Summary Statistics for Superstars (Counts)

	MD	PhD	MD/ PhD	NAS	HHMI	MERIT	Female	US born
Control	64	82	15	24	4	50	20	126
(n=161)	39.80%	50.90%	9.30%	14.90%	2.50%	31.10%	12.40%	78.30%
Extinct	73	68	20	33	11	51	16	135
(n=161)	45.3%	42.20%	12.40%	20.50%	6.80%	31.70%	9.90%	83.90%
Total	137	150	35	57	15	101	36	261
(n=322)	42.5%	46.60%	10.9%	17.70%	4.70%	31.40%	11.20%	81.10%

Table 2A: Demographic Characteristics of Coauthors

	Female	MD	PhD	MD/PhD	NAS	NIH Grantee	Basic Dept.
Controls	994	2,493	2,204	402	87	3,198	1,632
(n=5,131)	(19.50%)	(48.90%)	(43.20%)	(7.90%)	(1.70%)	(62.70%)	(32.00%)
Treatment	1,014	2,633	2,155	549	143	3,465	1,714
(n=5,365)	(19.00%)	(49.30%)	(40.40%)	(10.30%)	(2.70%)	(64.90%)	(32.10%)
Total	2,008	5,126	4,359	951	230	6,663	3,346
(n=10,496)	(19.20%)	(49.10%)	(41.80%)	(9.10%)	(2.20%)	(63.80%)	(35.10%)

Note: These tabulations exclude control and treatment colleagues whereby s/he is found to coauthor both with a treatment and a control superstar.

Table 2B: Coauthor Age and Achievement Over the Entire Career

		Mean	Std. Dev	Min.	Max.
Control Coauthors	**(N=5,131)**				
Year of Highest Degree		1975.531	9.708	1950	1996
Career Nb. of Papers		89.292	78.457	2	796
Career NIH Funding		$3,448,310	$6,708,865	$0	$141,176,240
Treatment Coauthors	**(N=5,365)**				
Year of Highest Degree		1974.549	9.966	1950	1996
Career Nb. of Papers		96.489	86.881	2	885
Career NIH Funding		$3,967,731	$10,230,741	$0	$451,590,368
Total	**(N=10,496)**				
Year of Highest Degree		1975.029	9.966	1950	1996
Career Nb. of Papers		92.973	82.946	2	885
Career NIH Funding		$3,713,811	$8,692,747	$0	$451,590,368

Table 2C: Number of Superstars per Coauthor

	Freq.	Proportion
1	9,566	91.14%
2	808	7.70%
3	103	0.98%
4	16	0.15%
5	3	0.03%
Total	10,496	100%

Note: These tabulations exclude control and treatment colleagues whereby s/he is found to coauthor both with a treatment and a control superstar.

Table 3: Summary Statistics for Dyads

Control Dyads (N=5,564)	Mean	Std. Dev	Min.	Max.
Dyad-level characteristics at the time of superstar extinction				
Cum. Nb. of Coauthorships	2.871	5.364	1	165
Years since first coauthorship	10.189	7.356	0	41
Years since last coauthorship	8.127	7.4	-3	41
Former trainee of the star	0.093	0.29	0	1
Co-Located	0.231	0.421	0	1
Within 10 miles	0.25	0.433	0	1
Normalized Keyword Overlap	0.311	0.167	0	1
"Accidental" Coauthorship	0.084	0.278	0	1
Coauthor-level characteristics at the time of superstar extinction				
Cum. Nb. of Papers	65.264	62.021	1	703
Cum. Nb. of Papers, JIF-weighted	235.765	293.868	0.526	3,226.33
Cum. NIH Funding	$2,267,215	$4,658,176	$0	$94,993,448
Holds R01	0.54	0.498	0	1
Career Age	21.325	9.048	0	52
"Elite" [NAS, HHMI, MERIT]	0.053	0.224	0	1

Treatment Dyads (N=6,006)	Mean	Std. Dev	Min.	Max.
Dyad-level characteristics at the time of superstar extinction				
Cum. Nb. of Coauthorships	3.049	5.458	1	112
Years since first coauthorship	10.973	7.823	0	39
Years since last coauthorship	8.809	7.817	-3	39
Former trainee of the star	0.094	0.292	0	1
Co-Located	0.226	0.418	0	1
Within 10 miles	0.255	0.436	0	1
Normalized Keyword Overlap	0.288	0.158	0	1
"Accidental" Coauthorship	0.073	0.259	0	1
Coauthor-level characteristics at the time of superstar extinction				
Cum. Nb. of Papers	72.205	70.266	1	726
Cum. Nb. of Papers, JIF-weighted	269.255	324.101	0.523	4,577.732
Cum. NIH Funding	$2,608,550	$6,550,636	$0	$206,150,592
Holds R01	0.552	0.497	0	1
Career Age	22.004	9.535	0	52
"Elite" [NAS, HHMI, MERIT]	0.066	0.248	0	1

Note: These tabulations exclude control and treatment dyads whereby the colleague is found to coauthor both with a treatment and a control superstar. For control dyads, the (counterfactual) year of death is that of their superstar collaborator's "nearest neighbor" among extinct superstars. See Appendix II for further details. Years since last coauthorship can be negative because of posthumous publications. More details on the construction of the normalized keyword overlap variable are provided in Appendix III.

Table 4: Trends in Stars' Publication Output Around the Time of Death

	(1)	(2)	(3)	(4)
	161 Superstars, Age at death≤67	**69 Superstars whose deaths were sudden**	**92 Superstars whose deaths were anticipated**	**30 Superstars Age at death>75**
	161 Controls	**69 Controls**	**92 Controls**	**30 Controls**
2 years after year of death	0.326**	0.306**	0.343**	0.208**
	[6.21]	[3.82]	[5.20]	[2.80]
1 year after year of death	0.763*	0.861	0.665**	1.108
	[2.20]	[0.73]	[3.04]	[0.27]
year of death	1.203	0.983	1.382*	0.709
	[1.64]	[0.09]	[2.39]	[0.99]
1 year before year of death	1.302**	1.146	1.432**	0.747
	[2.87]	[0.86]	[3.39]	[1.08]
2 years before year of death	1.100	0.949	1.218†	1.020
	[0.99]	[0.34]	[1.65]	[0.07]
3 years before year of death	1.176†	1.210	1.117	1.074
	[1.68]	[1.34]	[0.92]	[0.23]
4 years before year of death	1.211*	1.038	1.356**	1.021
	[2.01]	[0.22]	[2.83]	[0.12]
Log Quasi-Likelihood	-58,498	-25,113	-32,796	-11,039
Nb. of Observations	9,746	4,121	5,625	1,985
Nb. of Scientists	322	138	184	60

The estimates above are taken from a conditional fixed effects Poisson specification that also include 7 indicator variables corresponding to different age brackets and a full suite of calendar year effects (estimates not reported). The estimates are displayed as incidence rate ratios, e.g., the estimate in column (1) implies a statistically significant (1-0.326)=67.4% decrease in the rate of publication two years after a superstar scientist passes away (regardless of cause of death). Robust (QML) z-statistics are reported in brackets. The dependent variable is the weighted article count for the superstar, including only those 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.

†significant at 10%; *significant at 5%; **significant at 1%

Table 5: Impact of Superstar Death on Coauthors' Publication Rates

Panel A: Treatment Dyads Only, JIF-weighted Total Publications

	All		Sudden		Anticipated	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
After Death	0.926**	0.932**	0.895**	0.917**	0.953*	0.945*
	[4.57]	[3.96]	[4.73]	[3.17]	[2.07]	[2.38]
After Death ×		0.994		0.903*		1.064†
Regular Collab.		[0.22]		[2.17]		[1.71]
After Death ×		0.931		0.949		0.916
Close Collab.		[1.51]		[0.94]		[1.21]
Log Quasi-Likelihood	-1,432,370	-1,432,257	-610,192	-609,928	-821,290	-821,028
Nb. of Obs.	216,746	216,746	91,620	91,620	125,126	125,126
Nb. of Dyads	8,220	8,220	3,509	3,509	4,711	4,711
Nb. of Superstars	161	161	69	69	92	92

Panel B: Treatment Dyads Only, JIF-weighted Publications written with others

	All		Sudden		Anticipated	
	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
After Death	0.962*	0.944**	0.927**	0.927**	0.992	0.958†
	[2.37]	[3.39]	[3.26]	[2.84]	[0.35]	[1.86]
After Death ×		1.041		0.943		1.116**
Regular Collab.		[1.30]		[1.25]		[2.92]
After Death ×		1.178**		1.211**		1.152†
Close Collab.		[3.32]		[3.33]		[1.86]
Log Quasi-Likelihood	-1,406,641	-1,406,059	-598,472	-598,005	-807,309	-806,757
Nb. of Obs.	216,746	216,746	91,620	91,620	125,126	125,126
Nb. of Dyads	8,220	8,220	3,509	3,509	4,711	4,711
Nb. of Superstars	161	161	69	69	92	92

Estimates are displayed as incidence rate ratios (exponentiated coefficients). For example. the estimates in column (4b) of Panel B imply that casual coauthors suffer a statistically significant (1-0.944)=5.6% decrease in the rate of publication written with others after their superstar coauthor passes away, but that close collaborators (10 or more joint publications) incur an additional increase of 1.211-1=21.1%, for a net increase of 21.1-5.6=15.6%. All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

†significant at 10%; *significant at 5%; **significant at 1%.

Table 5: Impact of Superstar Death on Coauthors' Publication Rates

Panel C: Treatment and Control Dyads, JIF-weighted Total Publications

	All		Sudden		Anticipated	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
After Death	0.949**	0.954*	0.928**	0.950†	0.967	0.960
	[2.72]	[2.33]	[2.63]	[1.78]	[1.30]	[1.50]
After Death ×		1.028		0.939		1.089*
Regular Collab.		[0.76]		[1.01]		[2.01]
After Death ×		0.849*		0.860†		0.840
Close Collab.		[2.31]		[1.84]		[1.60]
Log Quasi-Likelihood	-1,659,785	-1,659,405	-729,397	-729,249	-929,953	-929,550
Nb. of Obs.	294,463	294,463	128,556	128,556	165,907	165,907
Nb. of Dyads	11,570	11,570	5,103	5,103	6,467	6,467
Nb. of Superstars	322	322	138	138	184	184

Panel D: Treatment and Control Dyads, JIF-weighted Publications written with others

	All		Sudden		Anticipated	
	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
After Death	0.982	0.955*	0.959	0.948†	1.003	0.962
	[0.96]	[2.40]	[1.55]	[1.89]	[0.10]	[1.50]
After Death ×		1.100**		1.003		1.168**
Regular Collab.		[2.61]		[0.05]		[3.58]
After Death ×		1.142†		1.185†		1.111
Close Collab.		[1.78]		[1.92]		[0.93]
Log Quasi-Likelihood	-1,615,834	-1,615,423	-709,860	-709,729	-905,588	-905,121
Nb. of Obs.	294,463	294,463	128,556	128,556	165,907	165,907
Nb. of Dyads	11,570	11,570	5,103	5,103	6,467	6,467
Nb. of Superstars	322	322	138	138	184	184

All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

†significant at 10%; *significant at 5%; **significant at 1%.

Table 6: Impact of Superstar Death on Coauthors' Citation Impact [Elite Subsample]

	All Pubs		Pubs above the Median		Pubs in the Top Quartile		Pubs in the Top Ventile		Pubs in the Top Percentile	
	(1a) No Ctrls	(1b) With Ctrls	(2a) No Ctrls	(2b) With Ctrls	(3a) No Ctrls	(3b) With Ctrls	(4a) No Ctrls	(4b) With Ctrls	(5a) No Ctrls	(5b) With Ctrls
After Death	0.965*	0.935**	0.967†	0.943**	0.963*	0.943**	0.939*	0.941*	0.904*	0.898**
	[2.08]	[3.45]	[1.81]	[2.82]	[2.01]	[2.79]	[2.30]	[2.22]	[2.38]	[2.66]
Log Quasi-Lkl.	-205,857	-289,513	-181,119	-255,159	-161,846	-228,001	-107,977	-150,078	-58,093	-79,136
Nb. of Obs.	75,939	110,062	75,908	110,031	75,865	109,988	75,354	109,218	72,279	103,461
Nb. of Dyads	2,648	3,870	2,647	3,869	2,644	3,866	2,625	3,837	2,516	3,636
Nb. of Superstars	155	307	155	307	155	307	155	307	155	307

Conditional dyad fixed effects quasi-MLE 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. Because article-level citation data is only available for scientists in the elite subsample (n=8,963), we restrict the estimation sample to elite coauthors, which account for a third of the dyads in the overall sample (i.e., the estimation sample in Table 7). Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

†significant at 10%; *significant at 5%; **significant at 1%.

Table 7: Impact of Superstar Death on Coauthors' NIH Grants

Panel A: Treatment Dyads Only, Number of Research NIH Grants

	All		Sudden		Anticipated	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
After Death	0.934†	0.938†	0.924	0.949	0.942	0.929
	[1.90]	[1.76]	[1.55]	[0.97]	[1.21]	[1.51]
After Death ×		1.002		0.909		1.079
Regular Collab.		[0.04]		[1.30]		[1.27]
After Death ×		0.938		0.924		0.959
Close Collab.		[0.59]		[0.43]		[0.44]
Log Quasi-Likelihood	-76,156	-76,155	-32,777	-32,774	-43,353	-43,351
Nb. of Obs.	146,339	146,339	62,887	62,887	83,452	83,452
Nb. of Dyads	5,426	5,426	2,358	2,358	3,068	3,068
Nb. of Superstars	161	161	69	69	92	92

Panel B: Treatment and Control Dyads, Number of Research NIH Grants

	All		Sudden		Anticipated	
	(4a)	(4b)	(5a)	(5b)	(6a)	(6b)
After Death	0.997	1.005	0.978	1.016	1.008	0.994
	[0.10]	[0.13]	[0.43]	[0.27]	[0.18]	[0.13]
After Death ×		1.026		0.936		1.091
Regular Collab.		[0.49]		[0.86]		[1.30]
After Death ×		0.824†		0.733†		0.920
Close Collab.		[1.82]		[1.84]		[0.74]
Log Quasi-Likelihood	-89,153	-89,147	-39,316	-39,310	-49,820	-49,818
Nb. of Obs.	186,494	186,494	82,208	82,208	104,286	104,286
Nb. of Dyads	7,133	7,133	3,174	3,174	3,959	3,959
Nb. of Superstars	322	322	138	138	184	184

Conditional dyad fixed effects quasi-MLE estimates for the determinants of the number of NIH Research Grants among coauthors of "superstar" academic life scientists. Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

†significant at 10%; *significant at 5%; **significant at 1%.

Table 8: Impact of Superstar Status on Coauthor Publication Rates

	Superstar's Total Cites at Time of Death		**Superstar's Total Cites at Time of Death, normalized by career length**		**Superstar's Career NIH Funding at Time of Death**		**Superstar's Career NIH Funding at Time of Death, normalized by career length**	
	(1a) w/o Controls	(1b) with Controls	(2a) w/o Controls	(2b) with Controls	(3a) w/o Controls	(3b) with Controls	(4a) w/o Controls	(4b) with Controls
After Death × Star in 1st Quartile	0.992 [0.19]	1.022 [0.52]	0.996 [0.08]	1.048 [0.99]	0.898* [2.44]	0.910† [1.89]	0.929 [1.56]	0.946 [1.06]
After Death × Star in 2nd Quartile	0.941* [2.02]	0.981 [0.58]	0.931* [2.26]	0.955 [1.25]	0.940† [1.73]	0.967 [0.86]	0.944† [1.65]	0.973 [0.70]
After Death × Star in 3rd Quartile	0.901** [3.26]	0.898** [2.82]	0.926* [2.40]	0.938† [1.81]	0.929* [2.45]	0.972 [0.76]	0.917** [2.77]	0.967 [0.90]
After Death × Star in 4th Quartile	0.924** [3.15]	0.959 [1.48]	0.914** [3.76]	0.939** [2.20]	0.915** [3.11]	0.956 [1.42]	0.910** [3.25]	0.934* [2.04]
Log Quasi-Likld.	-1,432,131	-1,659,458	-1,432,205	-1,659,618	-1,298,843	-1,534,733	-1,298,860	-1,534,786
Nb. of Obs.	216,746	294,463	216,746	294,463	196,970	273,945	196,970	273,945
Nb. of Dyads	8,220	11,570	8,220	11,570	7,493	10,779	7,493	10,779
Nb. of Superstars	161	322	161	322	150	302	150	302

Conditional dyad fixed effects quasi-MLE estimates for the determinants of JIF-weighted publication rates among coauthors of "superstar" academic life scientists. Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar. We interact the treatment variable with 4 indicator variables corresponding to quartiles for four distinct metrics of achievement for the superstars at the time of their death: total citations, total citations normalized by years of career, career NIH funding, and career NIH funding normalized by years of career. In the latter two cases, we exclude 7 scientists who spend all their careers at NIH campus in Bethesda, MD, and are therefore not eligible to receive extramural NIH funding.

†significant at 10%; *significant at 5%; **significant at 1%.

Table 9: Impact of Coauthor Status at the Time of Superstar Death

	No Other Elite Coauthor		Coauthor's R01 Grantee Status		MERIT, NAS, or HHMI Coauthor	
	(1a) w/o Controls	(1b) with Controls	(2a) w/o Controls	(2b) with Controls	(3a) w/o Controls	(3b) with Controls
After Death	0.941** [3.52]	0.970 [1.56]	0.933** [2.92]	0.983 [0.55]	0.915** [4.81]	0.933** [3.39]
After Death × no other elite coauth.	0.692** [8.94]	0.708** [8.02]				
After Death × R01 Grantee			0.990 [0.41]	0.951 [1.40]		
After Death × Coauthor "Elite"					1.057† [1.79]	1.116** [2.74]
% of Treatment Dyads Affected	18.41%	22.17%	58.10%	53.62%	9.06%	5.72%
Log Quasi-Likelihood	-1,430,523	-1,658,509	-1,432,363	-1,659,686	-1,432,225	-1,659,505
Nb. of Obs.	216,746	294,463	216,746	294,463	216,746	294,463
Nb. of Dyads	8,220	11,570	8,220	115,70	8,220	11,570
Nb. of Superstars	161	322	161	322	161	322

Conditional dyad fixed effects quasi-MLE estimates for the determinants of JIF-weighted publication rates among coauthors of "superstar" academic life scientists. Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar. 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.

†significant at 10%; *significant at 5%; **significant at 1%.

Table 10: Coauthor Publication Rates and Imperfect Skill Substitution

	Recent Coauthorship		Collaboration Age		Coauthor Age at Time of Death	
	(1a) w/o Controls	(1b) with Controls	(2a) w/o Controls	(2b) with Controls	(3a) w/o Controls	(3b) with Controls
After death	0.888**	0.902**	0.892**	0.905**		
	[6.50]	[4.53]	[6.44]	[4.47]		
After Death × At least 1 coauthorship in the three years preceding star's death	1.154**	1.187**				
	[5.34]	[4.85]				
After Death × First coauth. in the 5 years before star's death			1.145**	1.197**		
			[4.25]	[4.72]		
After Death × Coauthor less than 10 years of career age					0.863*	0.858*
					[2.11]	[2.37]
After Death × Coauthor b/w 10 and 20 years of career age					0.888**	0.924**
					[4.85]	[3.12]
After Death × Coauthor b/w 20 and 30 years of career age					0.970	0.983
					[1.28]	[0.65]
After Death × Coauthor more than 30 years of career age					0.913**	0.960
					[3.21]	[1.02]
% of Treatment Dyads Affected	24.12%	25.57%	23.43%	25.08%	—	—
Log Quasi-Likelihood	-1,430,937	-1,658,647	-1,431,149	-1,658,612	-1,431,854	-1,659,543
Nb. of Obs.	216,746	294,463	216,746	294,463	216,746	294,463
Nb. of Dyads	8,220	11,570	8,220	11,570	8,220	11,570
Nb. of Superstars	161	322	161	322	161	322

Conditional dyad fixed effects quasi-MLE estimates for the determinants of JIF-weighted publications among coauthors of academic life sciences superstar academics. Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

Table 11: Coauthor Publication Rates and Access to Resources

	Star and Coauthor Co-located at Time of Death		Star and Coauthor Separated by Less than 10 Miles at Time of Death		Star's Ties to NIH Funding Process		Quartile of Betweenness Centrality	
	(1a) w/o Controls	(1b) w/o Controls	(2a) with Controls	(2b) with Controls	(3a) w/o Controls	(3b) with Controls	(4a) w/o Controls	(4b) with Controls
After death	0.920**	0.951*	0.921**	0.953*	0.911**	0.933*		
	[4.21]	[2.27]	[4.05]	[2.08]	[4.09]	[2.56]		
After Death × Co-located	1.025	0.994						
	[0.80]	[0.16]						
After Death × Within 10 Miles			1.017	0.988				
			[0.56]	[0.34]				
After Death × Star on NIH Study Section					1.079	0.981		
					[1.39]	[0.37]		
After Death × Star Tied to NIH Study Section Members					1.007	1.011		
					[0.86]	[1.18]		
After Death × Star in 1st Quartile							1.010	1.041
							[0.28]	[1.17]
After Death × Star in 2nd Quartile							0.884*	0.892*
							[2.49]	[2.19]
After Death × Star in 3rd Quartile							0.930	0.955
							[1.47]	[0.82]
After Death × Star in 4th Quartile							0.921**	0.941*
							[3.88]	[2.54]
% of Treatment Dyads Affected	23.50%	22.96%	26.30%	25.38%	—	—	—	—
Log Quasi-Likelihood	-1,432,329	-1,659,784	-1,432,350	-1,659,779	-1,432,224	-1,659,718	-1,431,959	-1,659,451
Nb. of Obs.	216,746	294,463	216,746	294,463	216,746	294,463	216,746	294,463
Nb. of Dyads	8,220	11,570	8,220	11,570	8,220	11,570	8,220	11,570
Nb. of Superstars	161	322	161	322	161	322	161	322

Conditional dyad fixed effects quasi-MLE estimates for the determinants of JIF-weighted publications among coauthors of academic life sciences superstar academics. Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

Table 12: Coauthor Publication Rates and Proximity in Intellectual Space

	Former Trainee Status		Quartile of Unique Keyword Overlap		"Accidental" Coauthor	
	(1a) w/o Controls	(1b) with Controls	(2a) w/o Controls	(2b) with Controls	(3a) w/o Controls	(3b) with Controls
After Death	0.931** [4.25]	0.959* [2.15]			0.925** [4.71]	0.949** [2.75]
After Death × Coauthor is Former Trainee	0.927† [1.81]	0.876** [3.24]				
After Death × "Accidental" Coauthor					1.015 [0.35]	1.000 [0.00]
After Death × Kwd. Overlap in 1st Quartile			0.936* [2.21]	0.964 [1.14]		
After Death × Kwd. Overlap in 2nd Quartile			0.957† [1.78]	1.017 [0.57]		
After Death × Kwd. Overlap in 3rd Quartile			0.932** [2.88]	0.950 [1.61]		
After Death × Kwd. Overlap in 4th Quartile			0.862** [3.88]	0.859** [3.83]		
% of Treatment Dyads Affected	8.47%	9.46%	—	—	7.37%	7.78%
Log Quasi-Likelihood	-1,432,257	-1,659,558	-1,431,919	-1,659,104	-1,432,365	-1,659,785
Nb. of Obs.	216,746	294,463	216,746	294,463	216,746	294,463
Nb. of Dyads	8,220	11,570	8,220	11,570	8,220	11,570
Nb. of Superstars	161	322	161	322	161	322

Conditional dyad fixed effects quasi-MLE estimates for the determinants of JIF-weighted publications among coauthors of academic life sciences superstar academics. Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

†significant at 10%; *significant at 5%; **significant at 1%.

Table 13: Sensitivity Checks

	Stars 60 years old or less at time of death		**Stars 70 years old or less at time of death**		**Stars 75 years old or more at time of death**		**Placebo Death Dates for Control Superstars**
	(1a) w/o Controls	(1b) with Controls	(2a) w/o Controls	(2b) with Controls	(3a) w/o Controls	(3b) with Controls	(4) Controls Only
After death	0.923** [3.64]	0.953† [1.68]	0.925** [4.79]	0.944** [3.26]	1.012 [0.17]	0.970 [0.56]	1.001 [0.21]
Log Quasi-Likelihood	-757,997	-830,110	-1,731,842	-1,995,984	-213,869	-309,533	-761,217
Nb. of Obs.	105,649	139,374	260,137	350,927	32,843	56,672	140,474
Nb. of Dyads	4,036	5,500	9,855	13,742	1,230	2,190	2,564
Nb. of Superstars	86	172	190	380	30	60	161

Conditional dyad fixed effects quasi-MLE estimates for the determinants of JIF-weighted publications among coauthors of academic life sciences superstar academics. Estimates are displayed as incidence rate ratios (exponentiated coefficients). All models incorporate year effects and seven age category indicator variables (career age less than 5 years is the omitted category). Absolute value of robust (QML) z-statistics in brackets, clustered at the level of the superstar.

†significant at 10%; *significant at 5%; **significant at 1%.

Figure 1: Avoiding Contamination of the Control Sample

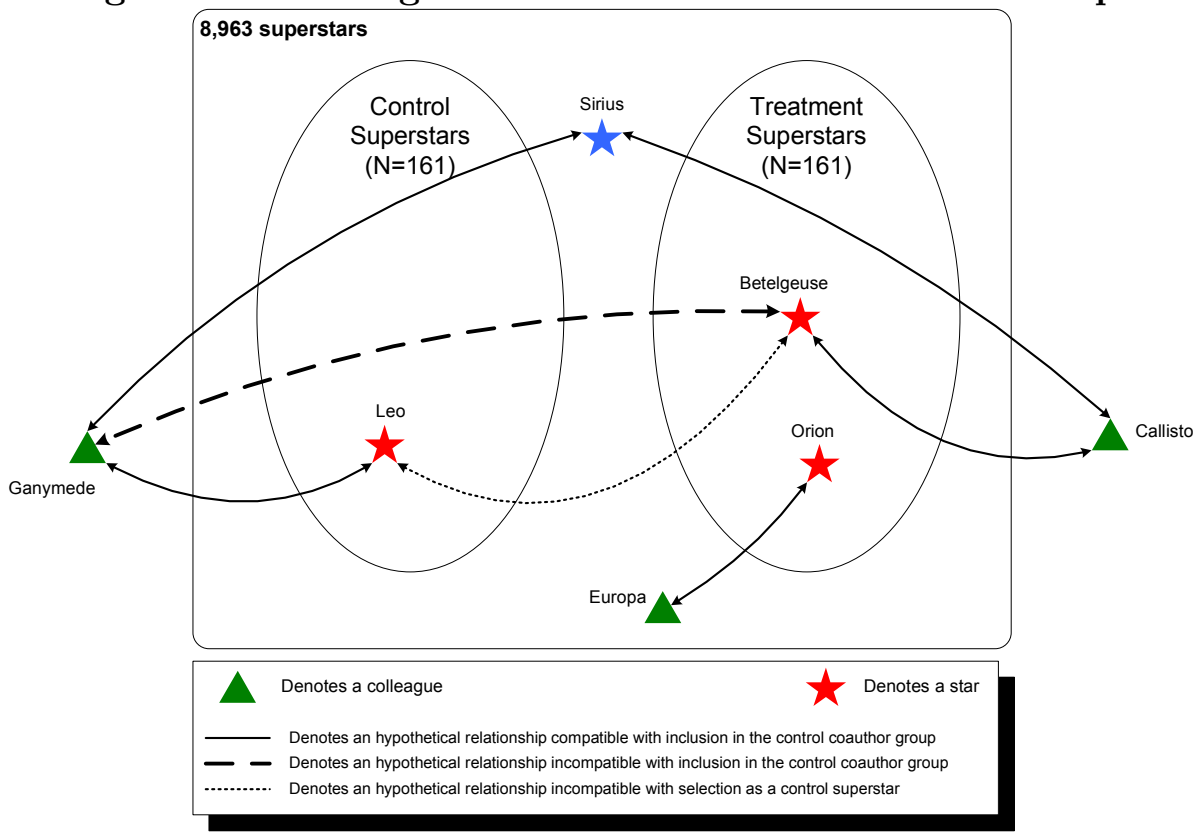


Figure 2: Number of Coauthors per Superstar

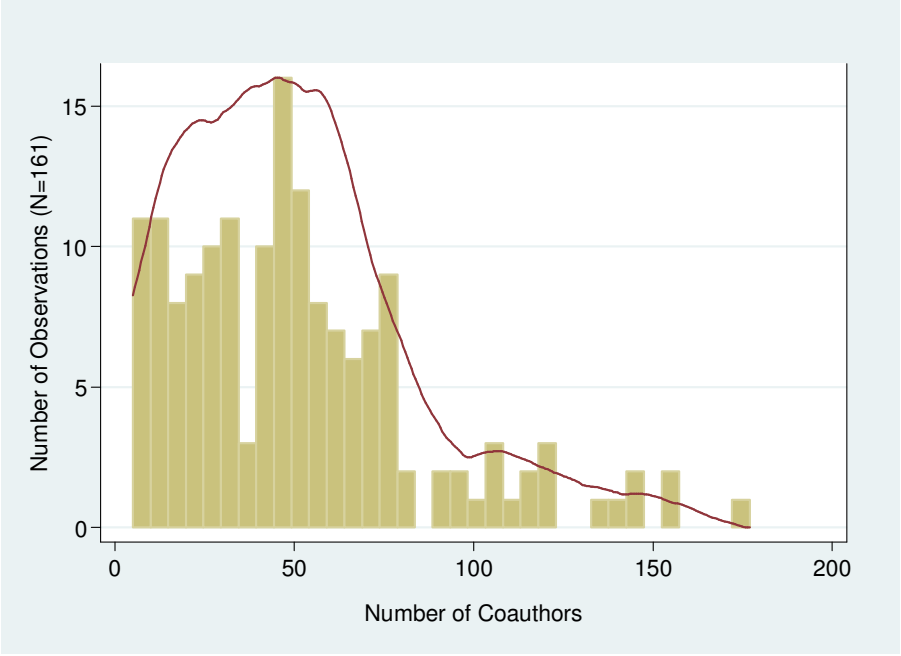


Figure 3: Intensity of Coauthorship [Dyad-level]

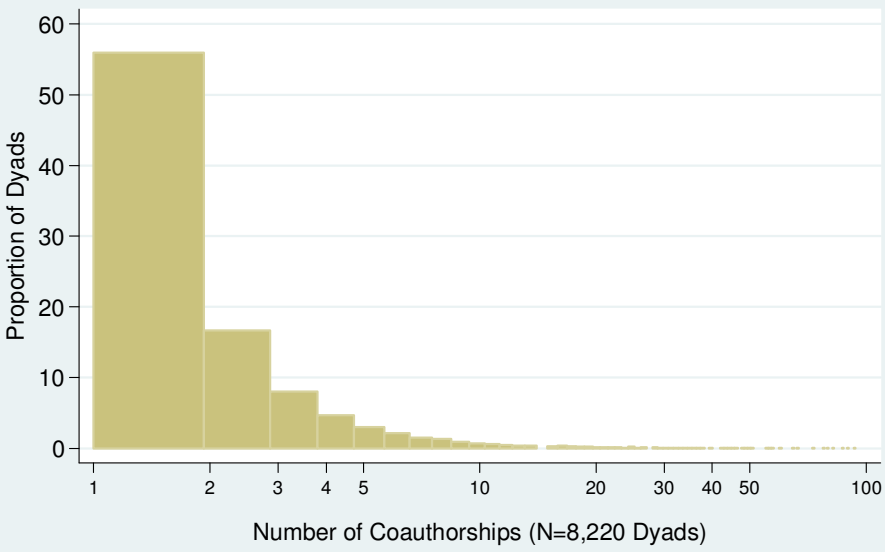


Figure 4: Collaboration Recency [Dyad-level]

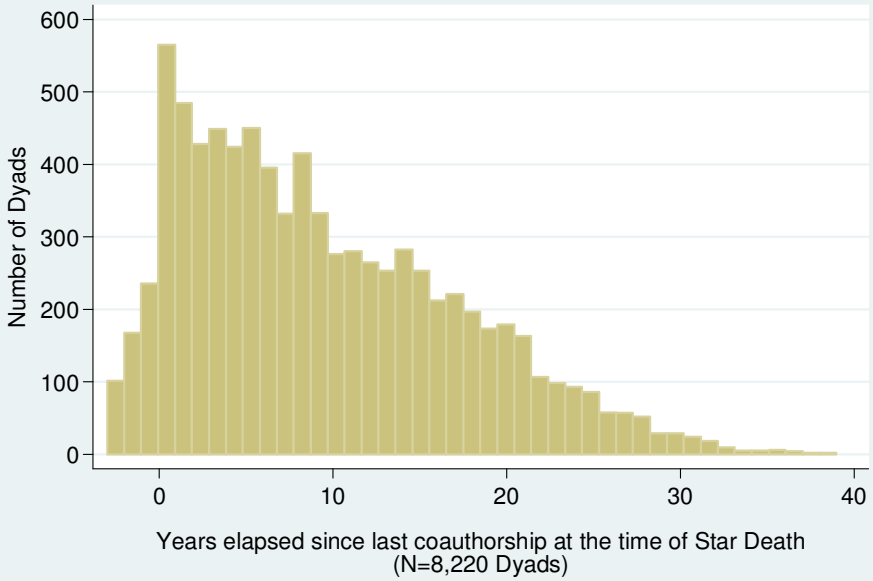


Figure 5: Proximity in Ideas Space [Dyad-level]

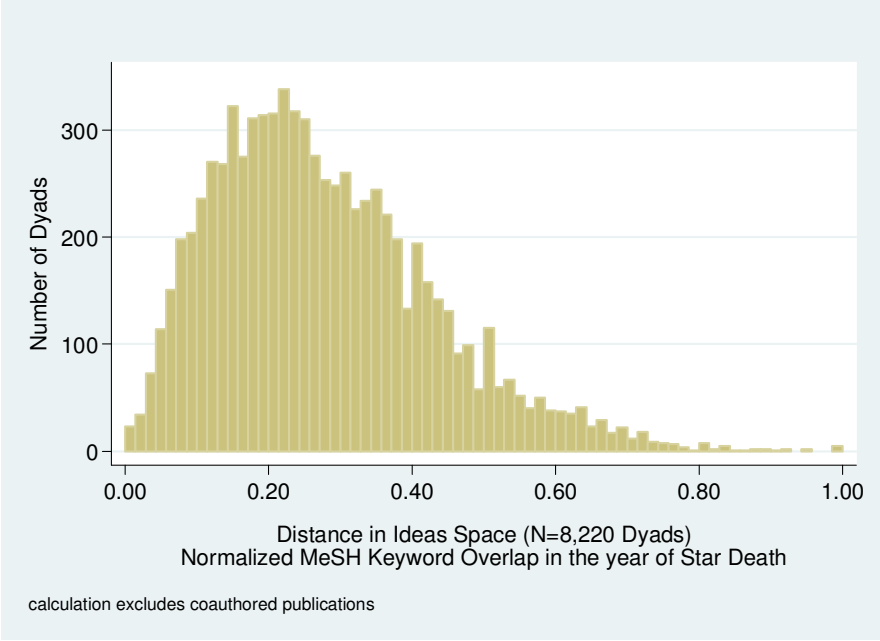
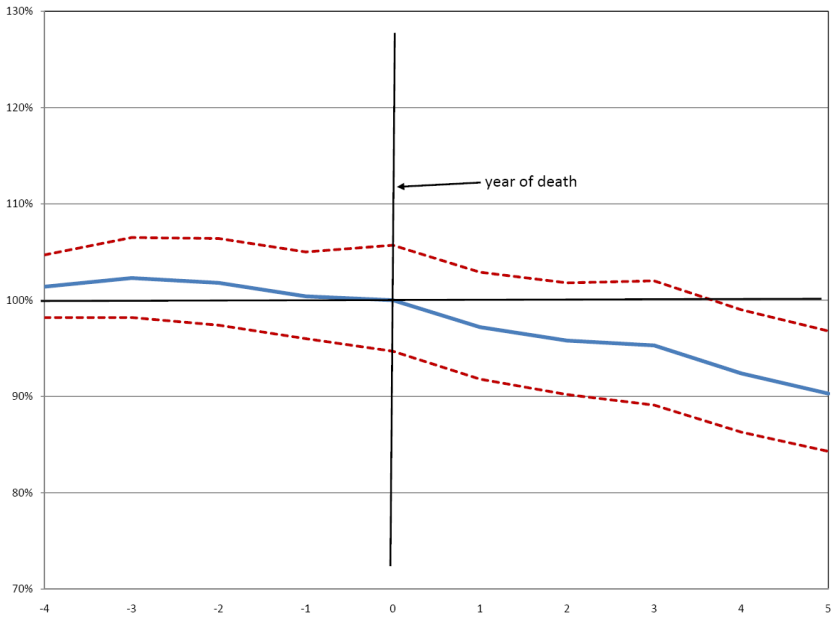


Figure 6: Time plot of coefficient estimates for the treatment effect



The solid blue lines in the above plot correspond to the coefficient estimates for the incidence rate ratios of a Poisson regression in which the weighted publication output of a colleague with other faculty than the dead superstar is regressed onto year effects, 7 indicator variables corresponding to different age brackets, and interactions of the treatment effect with 11 dummy variables corresponding to 4 years before the year of death, 3 years before the year of death,..., 5 years after the year of death, and 6 years after the year of death and above (not plotted). The 95% confidence interval (corresponding to robust standard errors, clustered around superstars) around these estimates is plotted with dashed red lines.

Appendix I: Matching Superstars and their Coauthors

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 superstar, 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 8,963 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 A1. 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 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 A1. 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.[27] Out of the 355 matches mentioned above, only 177 correspond to coauthors with unambiguous PubMED names. For the set of 161 superstars, S/CGen identifies 7,111 distinct coauthors with unambiguous PubMED names — an average of 51 coauthors per superstar (the median is 48).

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

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.[28] Specifically, we dropped from the main analysis all coauthors whose last name appear 29 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, in Table 6, we limit the estimation sample to elite coauthors (i.e., coauthors who belong to the set of 8,963 "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 (e.g., Table 5A).

Appendix II: Construction of the Control Group

The construction of the control group of *dyads* relies heavily on the preliminary selection of a control group of *superstars*. We aim to create a control population of superstars which matches closely the population of extinct superstars in terms of demographic characteristics (gender, degree, etc.), achievement (publications, citations, funding, etc.) and coauthorship patterns (number of casual, regular, and close coauthors). Since we have no way of capturing our stars' underlying health status, we have no qualms in engaging in some *ad hoc* specification search to create a control population which achieves a high degree of balance.

Control Superstars. From the set of from the set 8,963 superstars, we begin by eliminating individuals who coauthor with any of our 161 extinct superstars. We are left with $3,818 + 161 = 3,979$ superstars. Pooling all years between 1982 and 2003, we specify a logit model with death in a particular year as the outcome variable. As can be seen in Appendix Table 2, our preferred specification includes a long list of covariates; the pseudo-R^2 is .173. We then select, for each extinct superstar, the "nearest neighbor" based on the propensity score *in the year the superstar dies*. This caveat is important. The procedure will, by construction, balance observable variables that are fixed over time, such as vintage, degree, or gender. But there is no guarantee that it will balance observables such as number of coauthors or publications, since these change over time.[29] The characteristics of control and extinct superstars can be compared in Tables 1A and 1B.

Control Dyads. We simply add to the sample of dyads formed by the extinct superstars and their coauthors the set of dyads formed by control superstars and their coauthors. This yields a set of 15,705 dyads (7,485 control dyads and 8,220 treatment dyads). Alas, approximately 25% of these dyads involve scientists who collaborate both with a control and an extinct superstars (1,921 control dyads and 2,214 treatment dyads). It seems unreasonable to drop these problematic dyads only for the control group, since these "gregarious" scientists are significantly more accomplished than those who coauthor solely with extinct superstars, or solely with control superstars (see Appendix Table 3) .

As a consequence, whenever we make use of control dyads to identify the extinction effect, we eliminate from the estimation sample (a) treatment coauthors who also collaborate with control superstars; and (b)

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

[29] In practice, we found that selecting control superstars based on a weighted average of the difference in squared propensity scores and the absolute value of the difference in the stock of citations at the time of death yields a control population that balances all covariates. Formally, define $d(\alpha) = \alpha(p_i - p_j)^2 + (1-\alpha)|stkcites_i - stkcites_j|$ where p is the propensity score and *stkcites* the cumulative number of cites up until the year of death. We select as control the nearest neighbor for each extinct star with respect to the norm $d(\alpha)$. Empirically, $\alpha = .91$ yielded the most well-balanced population of control superstars.

control coauthors who also collaborate with treatment superstars. The final sample comprises observations for 11,570 dyads (5,564 control dyads and 6,006 treatment dyads). Figure 1 presents a stylized schema of the coauthorship ties between control and treatment coauthors that are allowed by our data assembly process, and contrasts them with the ties that we rule out.

Covariate Balance. Whether matching superstars on observables to predict death results in a sample of dyads that are also balanced between treatment and control observations is an empirical question. We assess the effectiveness of the procedure in Table 3 and Appendix Table 3. In Table 3, it can be observed that the average of dyad characteristics (number of coauthorships, time since first/last coauthorship, co-location, etc.) are very similar for the treatment and control groups. There are meaningful differences between these two groups when we focus on achievement. Treatment coauthors have written about 7 papers more than their control counterparts when their superstar coauthor passes away (of course, the year of death is counterfactual in the case of the control group) — a 9% difference. They have also received $341,000 more in NIH funding — a 13% difference. Appendix Table 3 shows that these differences are not reflecting the ineffectiveness of our two-step matching procedure. Rather, they reflect the fact that the "problematic coauthors" mentioned above are not a random sample of the complete set of coauthors. If we focus on the complete universe of dyads (the first panel of Appendix Table 3), the differences between control and treatment groups are less marked. For instance, treatment coauthors have garnered 5 more publications on average relative to the controls, but this is merely a 4% difference, since the average number of publications is much higher in the complete set. The funding advantage for treatment coauthors is also much less pronounced.

In conclusion, the merely approximate degree of covariate balance highlights the challenges involved in building a valid control group in this setting. Complete lack of contamination between treatment and control subjects is a salient ideal, but one that is bound to remain out of reach because of indirect coauthorship ties between scientists. Our estimates relying on two levels of difference should not be thought of as superior, in some absolute sense, to those relying on a single level of difference. Rather, they are informative in a different way: they address the concern that effects based solely on changes in output trends for treatment coauthors confound the effect of exposure to the star with collaboration-specific life-cycle effects.

Appendix III: 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.[30] 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,

[30] 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.

which run the gamut from the very general ("Humans", "RNA/Genetics") to the very specific ("Repetitive Sequences, Nucleic Acid", "CpG Islands", "DNA Transposable Elements").[31]

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 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 Appendix Table 4; it is automated by SciDist, an open-source software program we specifically designed for this purpose.[32] 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. Appendix Table 4 lists the Top 10 overlapping terms with highest and lowest combined use, respectively.[33]

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).

[31] 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).

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

[33] 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.

Appendix Table 1a: Superstar Sample, Sudden Deaths

			Cause of Death	Institutional Affiliation	Field
Raymond R. Margherio	(1940-2000)	MD	aneurysm	Wayne State University School of Medicine	clinical studies in age-related eye diseases
A. Arthur Gottlieb	(1937-1998)	MD	pulmonary embolus following surgery	Tulane University School of Medicine	role of macrophage nucleic acid in antibody production
George B. Craig, Jr.	(1930-1995)	PhD	heart attack	University of Notre Dame	genetics and reproductive biology of aedes mosquitoes
Walter F. Heiligenberg	(1938-1994)	PhD	plane crash	UCSD	neuroethological studies of electrolocation
Donald T. Witiak	(1935-1998)	PhD	stroke	University of Wisconsin	stereochemical studies of hypocholesterolemic agents
D. Martin Carter	(1936-1993)	MD/PhD	dissecting aortic aneurysm	Rockefeller University	susceptibility of pigment and cutaneous cells to DNA injury by UV
Harold A. Menkes	(1938-1987)	MD	car accident	Johns Hopkins University	occupational and environmental lung disease
Jonathan M. Mann	(1943-1998)	MD	plane crash	Harvard University School of Public Health	AIDS prevention
Gary J. Miller	(1950-2001)	MD/PhD	heart attack	University of Colorado HSC	vitamin D receptors in the growth regulation of prostate cancer cells
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
Roy D. Schmickel	(1936-1990)	MD	died tragically	University of Pennsylvania	isolation and characterization of human ribosomal DNA
Neil S. Jacobson	(1949-1999)	PhD	heart attack	University of Washington	marital therapy, domestic violence, and the treatment of depression
George Streisinger	(1927-1984)	PhD	scuba-diving accident	University of Oregon	genetic mutations and the nervous system development in lower vertebrates
Roland D. Ciaranello	(1943-1994)	MD	heart attack	Stanford University	molecular neurobiology and developmental disorders
Christopher A. Dawson	(1942-2003)	PhD	suddenly	Medical College of Wisconsin	pulmonary hemodynamics
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
Hymie L. Nossel	(1930-1983)	MD/PhD	heart attack	Columbia University	causes of thrombosis and the nature of hemostasis
James N. Davis	(1939-2003)	MD	airplane crash	SUNY HSC at Stony Brook	mechanisms underlying neuronal injury after brain ischemia
Sandy C. Marks, Jr.	(1937-2002)	DDS/PhD	heart attack	UMASS	bone cell biology
George J. Schroepfer, Jr.	(1932-1998)	MD/PhD	heart attack	Rice University	regulation of the formation and metabolism of cholesterol
Edward V. Evarts	(1926-1985)	MD	heart attack	NIH	electrophysiological activity of in vivo neurons in waking and sleeping states
Stanley R. Kay	(1946-1990)	PhD	heart attack	Albert Einstein College of Medicine	symptoms and diagnostic tests of schizophrenia
Howard S. Tager	(1945-1994)	PhD	heart attack	University of Chicago	structure, action, regulation and degradation of insulin and glucagon
Lewis W. Wannamaker	(1923-1983)	MD	heart attack	University of Minnesota Medical School	clinical and epidemiologic aspects of streptococcal infections
Emil T. Kaiser	(1938-1988)	PhD	complications from kidney transplant	Rockefeller University	mechanism of carboxypeptidase action
Mu-En Lee	(1954-2000)	MD/PhD	complications from routine surgery	Harvard Medical School/MGH	characterization of vascular smooth muscle LIM protein
Thomas P. Dousa	(1937-2000)	MD/PhD	heart attack	Mayo Clinic	cellular action of vasopressin in the kidney
Robert M. Macnab	(1940-2003)	PhD	accidental fall	Yale University	sequence analysis and function of bacterial flagellar motor
Mary Lou Clements	(1946-1998)	MD	airplane crash	Johns Hopkins University	development of AIDS vaccines
Abraham M. Lilienfeld	(1920-1984)	MD	heart attack	Johns Hopkins University School of Public Health	epidemiological methods for the study of chronic diseases
Julio V. Santiago	(1942-1997)	MD	heart attack	Washington University in St. Louis	social factors, lifestyle practices, and medication in the onset of type II diabetes
John J. Jeffrey, Jr.	(1937-2001)	PhD	stroke	Albany Medical College	mechanism of action and the physiologic regulation of mammalian collagenases
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
John J. Wasmuth	(1946-1995)	PhD	heart attack	University of California — Irvine	human-hamster somatic cell hybrids/localization of Hnyington's disease gene
Dolph O. Adams	(1939-1996)	MD/PhD	unexpected	Duke University	Development and regulation of macrophage activation
Fredric S. Fay	(1943-1997)	PhD	heart attack	UMASS	generation and regulation of force in smooth muscle
D. Michael Gill	(1940-1990)	PhD	heart attack	Tufts University	biochemistry of cholera toxin and other pathogenic toxins
Gerald P. Rodnan	(1927-1983)	MD	heart attack	University of Pittsburgh	renal transport if uric acid and protein
Donald C. Shreffler	(1933-1994)	PhD	heart attack	Washington University in St. Louis	organization and functions of H-2 gene complex
Thomas F. Burks, II	(1938-2001)	PhD	heart attack	University of Texas HSC at Houston	central and peripheral neuropeptide pharmacology
Roger R. Williams	(1944-1998)	MD	airplane crash	University of Utah	genetics and epidemiology of coronary artery diseases
James R. Neely	(1936-1988)	PhD	heart attack	Penn State University	effects of diabetes and oxygen deficiency in regulation of metabolism in the heart
Matthew L. Thomas	(1953-1999)	PhD	died while travelling	Washington University in St. Louis	function and regulation of leukocyte surface glycoproteins
Simon J. Pilkis	(1942-1995)	MD/PhD	heart attack	University of Minnesota	carbohydrate metabolism and diabetes
William H. Oldendorf	(1925-1992)	MD	complications from heart disease	UCLA	x-ray shadow radiography and cerebral angiography
Norbert Freinkel	(1926-1989)	MD	heart attack	Northwestern University	metabolic regulation in normal and diabetic pregnancies
Takis S. Papas	(1935-1999)	PhD	unexpected and sudden	Medical University of South Carolina	characterization of ETS genes and retroviral onc genes
Susumu Hagiwara	(1922-1989)	PhD	bacterial infection	UCLA	evolutionary and developmental properties of calcium channels in cell membranes
Philip J. Fialkow	(1933-1996)	MD	trekking accident in Nepal	University of Washington	origins of myeloid leukemia tumors
Richard E. Heikkila	(1942-1991)	PhD	murder	UMDNJ Robert Wood Johnson Medical School	oxidation-reduction reactions and the dopamine receptor system
John B. Penney, Jr.	(1947-1999)	MD	heart attack	Harvard Medical School/MGH	receptor mechanisms in movement disorder pathophysiology
Gerald P. Murphy	(1934-2000)	MD	heart attack	Roswell Park Cancer Institute/SUNY Buffalo	detection, immunotherapy, and prognostic indicators of prostate cancer
Ronald G. Thurman	(1941-2001)	PhD	massive heart attack	University of North Carolina	hepatic metabolism, alcoholic liver injury and toxicology
John P. Merrill	(1917-1984)	MD	drowned	Harvard Medical School/Brigham & Women's Hospita	role of the immune system in kidney transplantation
DeWitt S. Goodman	(1930-1991)	MD	pulmonary embolism	Columbia University	lipid metabolism and its role in the development of heart and artery disease

Gerald D. Aurbach	(1927-1991)	MD	hit in a head by a stone	NIH	bone metabolism and calcium homeostasis
Peter M. Steinert	(1945-2003)	PhD	heart attack	NIH	structures and interactions of the proteins characteristic of epithelial cells
John H. Walsh	(1938-2000)	MD	heart attack	UCLA	gastrointestinal hormones, gastric acid production and peptic ulcer disease
Paul B. Sigler	(1934-2000)	MD/PhD	heart attack	Yale University	structural analysis of biological macromolecules
Alan P. Wolffe	(1959-2001)	PhD	car accident	NIH	role of DNA methylation in regulating gene expression
Victor J. Ferrans	(1937-2001)	MD/PhD	complications from diabetes	NIH	myocardial and vascular pathobiology
Demetrios Papahadjopoulos	(1934-1998)	PhD	adverse drug reaction/multi-organ failure	UCSF	phospholipid-protein interactions, lipid vesicles, and membrane function
William L. McGuire	(1937-1992)	MD	scuba-diving accident	University of Texas HSC at San Antonio	mechanisms of hormonal control and growth/regression of mammary carcinoma
Jeffrey M. Isner	(1947-2001)	MD	heart attack	Tufts University	therapeutic angiogenesis in vascular medicine, cardiovascular laser phototherapy
Patricia S. Goldman-Rakic	(1937-2003)	PhD	struck by a car	Yale University	development and plasticity of the primate frontal lobe
Don C. Wiley	(1944-2001)	PhD	accidental fall	Harvard University	viral membrane and glycoprotein structure
Henry G. Kunkel	(1916-1983)	MD	complications after vascular surgery	Rockefeller University	identification of MHC Class II molecules
Zanvil A. Cohn	(1926-1993)	MD	aortic dissection	Rockefeller University	macrophage in cell biology and resistance to infectious disease

Appendix Table 1b: Superstar Sample, Anticipated Deaths

			Cause of Death	Institutional Affiliation	Field
Jack E. White	(1921-1988)	MD	cancer	Howard University School of Medicine	epidemiology and treatment of cancer among african-americans
Gregory Mooser	(1942-2003)	DDS/PhD	complications from alzheimer's disease	University of Southern California	characterization of glucosyltranserase enzymes secreted by oral bacteria
Teruzo Konishi	(1920-1984)	MD/PhD	cancer	NIEHS	physiological and biophysical functions of the inner ear
Samuel W. Perry, 3rd	(1941-1994)	MD	pancreatic cancer	Weill Medical College — Cornell University	psychological course of prolonged infection among AIDS patients
Leo J. Neuringer	(1928-1993)	PhD	cancer	MIT	NMR studies of normal and transformed cell membranes
Elizabeth M. Smith	(1939-1997)	PhD	cancer	Washington University School of Medicine	psychiatric problems among disaster survivors
Elizabeth A. Bates	(1974-2003)	PhD	pancreatic cancer	UCSD	cross-linguistic studies of language development, processing and breakdown in aphasia
John Gibbon	(1934-2001)	PhD	cancer	Columbia University	CNS functions underlying the interval time sense in animals and humans
Robert F. Spencer	(1949-2001)	PhD	gastric carcinoma	Medical College of Virginia	neuroanatomy of the oculomotor system
Laird S. Cermak	(1942-1999)	PhD	leukemia	Boston University	psychological studies of memory and cognitive deficits related to chronic alcoholism
Larry C. Clark	(1948-2000)	PhD	prostate cancer	University of Arizona	nutritional prevention of cancer
Keith Green	(1940-2001)	PhD	died after lengthy illness	Medical College of Georgia	ion and water movement in ocular tissues, ocular response to drugs
Lawrence H. Piette	(1932-1992)	PhD	cancer	Utah State University	electron spin resonance spectroscopy
Eleanor M. Saffran	(1938-2002)	PhD	amyotrophic lateral sclerosis	Temple University School of Medicine	cognitive deficits in brain-damaged patients
Joseph Stokes, 3rd	(1924-1989)	MD	cancer	Boston University School of Medicine	epidemiological studies of coronary heart disease
Jane Pitt	(1938-2003)	MD	chronic lymphocytic leukemia	Columbia University College of Physicians and Surgeo	perinatal transmission of HIV and retroviral infections
Tai-Shun Lin	(1939-1994)	PhD	non hodgkin's lymphoma	Yale University School of Medicine	synthesis/development of nucleoside analogs as antiviral and anticancer compounds
Gareth M. Green	(1931-1998)	MD/PhD	cancer	Harvard University School of Public Health	role of alveolar macrophages in pulmonary defense mechanisms
Barbara H. Bowman	(1930-1996)	PhD	cancer	University of Texas HSC at San Antonio	genetic control of the structure of human proteins
Joseph B. Warshaw	(1936-2003)	MD	multiple myeloma	University of Vermont College of Medicine	developmental neurobiology of respiratory control
Robert J. Fass	(1939-2002)	MD	lung cancer	Ohio State University	In vitro methods to test antimicrobial susceptibility of infectious agents
Priscilla A. Campbell	(1940-1998)	PhD	cervical cancer	University of Colorado HSC/Nat. Jewish center	cell biology of the immune response to bacteria
Michael J. Goldstein	(1930-1997)	PhD	cancer	UCLA	contributing factors to the onset of schizophrenia
Frank Lilly	(1930-1995)	PhD	prostate cancer	Albert Einstein College of Medicine	role of hereditary factors in governing susceptibility to cancer-causing agents
William L. Chick	(1938-1998)	MD	diabetes complications	UMASS	studies of islet and beta cells in pancreatic transplantation
Ernest G. Peralta	(1959-1999)	PhD	brain cancer	Harvard University	signal transduction mechanisms of muscarinic receptors
Charlotte Friend	(1921-1987)	PhD	lymphoma	Mount Sinai School of Medicine	tissue studies of murine virus-induced leukemia
Jiri Palek	(1934-1998)	MD	2 year illness	Tufts University	membrane properties of abnormal red cells
Helene S. Smith	(1941-1997)	PhD	breast cancer	UCSF	malignant progression of the human breast/predictors of breast cancer prognosis
Bruce S. Schoenberg	(1942-1987)	MD	cancer	NIH	prevention and control of neurological disorders
Marian W. Fischman	(1939-2001)	PhD	colon cancer	Columbia University	behavioral pharmacology of cocaine
William H. Tooley	(1925-1992)	MD	long illness	UCSF School of Medicine	prevention and treatment of respiratory distress in neonates
Mette Strand	(1937-1997)	PhD	cancer	Johns Hopkins University	parasite immunochemistry and vaccine development
Joachim G. Liehr	(1942-2003)	PhD	pancreatic cancer	University of Texas Medical Branch at Galveston	mechanism of estrogen-induced carcinogenesis
C. Richard Taylor	(1939-1995)	PhD	heart failure	Harvard University	Energetics of animal locomotion
David G. Marsh	(1940-1998)	PhD	glioblastoma	Johns Hopkins University	genetics of allergy and asthma
Michael Solursh	(1942-1994)	PhD	AIDS	University of Iowa School of Medicine	extracellular matrix and cell migration
Roy H. Steinberg	(1935-1997)	MD/PhD	multiple myeloma	UCSF	pigment epithelium interactions with neural retina
Harvey D. Preisler	(1941-2002)	MD	lymphoma	Rush Medical College	clinical and biological studies of myeloid leukemias
George Némethy	(1934-1994)	PhD	brain cancer	Mount Sinai School of Medicine	methods to analyze and predict the structures of protein molecules

Irving Kupfermann	(1938-2002)	PhD	Creutzfeldt-Jacob's disease	Columbia University	Behavioral and neural analysis of learning in aplaysia
B. Frank Polk	(1942-1988)	MD	brain cancer	Johns Hopkins University	epidemiology of HIV infection
Murray Rabinowitz	(1927-1983)	MD	muscular dystrophy	University of Chicago	mitochondrial assembly and replication
Lois K. Miller	(1945-1999)	PhD	melanoma	University of Georgia	genetics and molecular biology of baculoviruses
Gerald T. Babcock	(1946-2000)	PhD	cancer	Michigan State University	bioenergetic mechanisms in multicenter enzymes
Aaron Janoff	(1930-1988)	PhD	long illness	SUNY HSC at Stony Brook	pathology of smoking and emphysema
Peter A. Kollman	(1944-2001)	PhD	cancer	UCSF	free energy perturbation calculations and their application to macromolecules
G. Harrison Echols, Jr.	(1933-1993)	PhD	lung cancer	University of California — Berkeley	Genetic and chemical studies of phage lambda development
Kiichi Sagawa	(1926-1989)	MD/PhD	cancer	Johns Hopkins University	modelling the mechanics of cardiac chamber contraction
Albert Dorfman	(1916-1982)	MD/PhD	kidney failure	University of Chicago	biochemistry of connective tissues
Edwin H. Beachey	(1934-1989)	MD	cancer	University of Tennessee at Memphis	chemistry and immunology of streptococcal m proteins
John C. Liebeskind	(1935-1997)	PhD	cancer	UCLA	behavioral and electrophysiological studies of pain
Janis V. Giorgi	(1947-2000)	PhD	uterine cancer	UCLA	cellular immunology of resistance to HIV
Bernard N. Fields	(1938-1995)	MD	pancreatic cancer	Harvard Medical School/Brigham & Women's Hospita	genetic and molecular basis of viral injury to the nervous system
Eva J. Neer	(1937-2000)	MD	breast cancer	Harvard Medical School/Brigham & Women's Hospita	regulation and cellular levels of G protein subunits
Richard P. Bunge	(1932-1996)	MD	esophageal cancer	University of Miami	schwann cell biology and human spinal cord injury
Thoralf M. Sundt, Jr.	(1930-1992)	MD	bone marrow cancer	Mayo Clinic	surgical techniques for intracranial aneurysms
George Khoury	(1943-1987)	MD	lymphoma	NIH	genetics of simian virus 40, human papovavirus and HIV
John S. O'Brien	(1934-2001)	MD	postpolio complications	UCSD	discovery of the gene responsible for Tay-Sachs disease
Merton Bernfield	(1938-2002)	MD	Parkinson's Disease	Harvard Medical School/Children's Hospital	nature and interactions of cell surface proteoglycans during morphogenesis
Richard K. Gershon	(1932-1983)	MD	lung cancer	Yale University	immunologic responses to tumor grafts
Donald J. Cohen	(1940-2001)	MD	ocular melanoma	Yale University	Tourette's syndrome and autism in children
Theodore S. Zimmerman	(1937-1988)	MD	lung cancer	Scripps Research Institute	platelet/plasma protein interaction in blood coagulation
Nelson Butters	(1937-1995)	PhD	Lou Gehrig's disease	UCSD	cognitive deficits related to chronic alcoholism
Edward C. Franklin	(1928-1982)	MD	brain cancer	New York University	structure and properties of rheumatoid antibodies
Thomas W. Smith	(1936-1997)	MD	mesothelioma	Harvard Medical School/Brigham & Women's Hospita	Mechanism and reversal studies of digitalis
Norton B. Gilula	(1944-2000)	PhD	lymphoma	Scripps Research Institute	cell junction biosynthesis and biogenesis/cell-cell communication
Paul C. MacDonald	(1930-1997)	MD	cancer	University of Texas Southwestern Medical Center at D	origin and interconversion of gonadal and adrenal streoid hormones
Edwin L. Bierman	(1930-1995)	MD	bone cancer	University of Washington	Metabolism of particulate fat in diabetes and atherosclerosis
Ora M. Rosen	(1935-1990)	MD	breast cancer	Sloan Kettering Institute for Cancer Research	Cloning and characterization of gene for human insulin receptor
Joel D. Meyers	(1944-1991)	MD	colon cancer	University of Washington/FHCRC	infections caused by suppression of the immune system in organ transplant and AIDS
Sidney H. Ingbar	(1925-1988)	MD	lung cancer	Harvard Medical School/Beth Israel Medical Center	physiology of the thyroid gland and its clinical diseases
Melvin L. Marcus	(1940-1989)	MD	colon cancer	UMASS	cardiology, heart disease, coronary vascular adaptations to myocardial hypertrophy
George G. Glenner	(1927-1995)	MD	systemic senile amyloidosis	UCSD	molecular structure of the amyloid protein
J. Christian Gillin	(1938-2003)	MD	esophageal cancer	UCSD	serotenergic mechanisms in sleep and depression
John R. Williamson	(1934-2000)	PhD	cancer	University of Pennsylvania School of Medicine	molecular mechanisms of hormonal signal transduction
Howard M. Temin	(1934-1994)	PhD	lung cancer	University of Wisconsin	molecular biology and genetics of tumor viruses
Ira Herskowitz	(1946-2003)	PhD	pancreatic cancer	UCSF	genetics of yeast mating type
Harold C. Neu	(1934-1998)	MD	glioblastoma	Columbia University	surface enzymes in bacteria
Charles D. Heidelberger	(1920-1983)	PhD	carcinoma of nasal sinus	University of Southern California	effects of fluorinated pyrimidines on tumors
Gerald L. Klerman	(1928-1992)	MD	diabetes	Weill Medical College — Cornell University	studies of depression, schizophrenia and panic and other anxiety disorders
Sydney E. Salmon	(1936-1999)	MD	pancreatic cancer	University of Arizona	quantitative method for evaluating changes in myeloma tumor mass
Markku Linnoila	(1947-1998)	MD/PhD	cancer	NIH	studies on the biological bases of impulsivity and aggression
Wallace P. Rowe	(1926-1983)	MD	colon cancer	NIH	genetic basis of disease in murine leukemia viruses
Richard J. Wyatt	(1939-2002)	MD	lung cancer	NIH	biochemistry of schizophrenia
Sheldon M. Wolff	(1930-1994)	MD	complications from a renal malignancy	Tufts University School of Medicine	treatment of fevers from infectious diseases like wegener's granulomatosis
Charles G. Moertel	(1927-1994)	MD	Hodgkin's Disease	Mayo Clinic	clinical treatments of gastrointestinal cancer
Allan C. Wilson	(1934-1991)	PhD	leukemia	University of California — Berkeley	use of molecular approaches to understand evolutionary change
Edgar Haber	(1932-1997)	MD	multiple myeloma	Harvard University School of Public Health	biological regulation of the renin-angiotensin system
Henry S. Kaplan	(1918-1984)	MD	lung cancer	Stanford University School of Medicine	radiation-induced leukemia in the C57BL mouse
Charles A. Janeway, Jr.	(1943-2003)	MD	B-cell lymphoma	Yale University	innate immunity and T lymphocyte biology
Harold Weintraub	(1945-1995)	MD/PhD	brain cancer	University of Washington/FHCRC	characterization and function of MyoD gene

Appendix Table 2: Predicting Death in the Superstar Sample [Logit]

	dep. var.=1 if star dies in year t, 0 otherwise
Cumulative nb. of publications	1.003*
	[2.44]
Nb. of "casual" coauthors	1.017**
	[5.02]
Nb. of "regular" coauthors	0.984
	[1.22]
Nb. of "close" coauthors	0.917**
	[3.08]
Female	1.026
	[0.10]
Member of the NAS	1.065
	[0.33]
HHMI	2.197†
	[1.76]
Ph.D	0.843
	[1.11]
MD/Ph.D	1.677*
	[1.99]
Other Health Doctorate/PhD	2.081
	[1.33]
Top 10 School	1.546**
	[3.14]
NIH Intramural Scientist	1.008
	[0.03]
U.S. Born	0.972
	[0.18]
Log Quasi-Likelihood	-1,316
Nb. of Observations	97,602
Nb. of Superstars	3,979

Robust z-statistics in brackets, clustered by superstar
The specification also incluses career age indicator variables, degree vintage indicator variables, and year effects.
†significant at 10%; *significant at 5%; **significant at 1%

Appendix Table 3: Assessing Covariate Balance

All Dyads	Mean	Std. Dev	Min.	Max.
Control Dyads (N=7,485)				
Cum. Nb. of Coauthorships	2.788	5.630	1	201
Cum. Nb. of Papers	87.972	96.005	1	1,559
Cum. Nb. of Papers, JIF-weighted	353.053	497.297	0.572	5,636.42
Cum. NIH Funding	$3,240,821	$6,727,974	$0	$102,335,856
Treatment Dyads (N=8,220)				
Cum. Nb. of Coauthorships	3.082	5.756	1	112
Cum. Nb. of Papers	92.310	95.933	1	930
Cum. Nb. of Papers, JIF-weighted	379.330	502.994	0.577	5,841.22
Cum. NIH Funding	$3,347,435	$7,340,808	$0	$206,150,592
Excluding "Problematic" Dyads	**Mean**	**Std. Dev**	**Min.**	**Max.**
Control Dyads (N=5,564)				
Cum. Nb. of Coauthorships	2.871	5.364	1	165
Cum. Nb. of Papers	65.264	62.021	1	703
Cum. Nb. of Papers, JIF-weighted	235.765	293.868	0.526	3,226.33
Cum. NIH Funding	$2,267,215	$4,658,176	$0	$94,993,448
Treatment Dyads (N=6,006)				
Cum. Nb. of Coauthorships	3.049	5.458	1	112
Cum. Nb. of Papers	72.205	70.266	1	726
Cum. Nb. of Papers, JIF-weighted	269.255	324.101	0.523	4,577.73
Cum. NIH Funding	$2,608,550	$6,550,636	$0	$206,150,592

Appendix Table 4: 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

Appendix Table 5
Impact of Superstar Death on Coauthors' Publication Rates [OLS]
Treatment Dyads Only, JIF-weighted Total Publications

	All		Sudden		Anticipated	
	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
After Death	-1.383**	-1.265**	-2.141**	-1.651**	-0.795	-0.940†
	[0.364]	[0.377]	[0.480]	[0.560]	[0.506]	[0.506]
After Death ×		-0.155		-1.998*		1.097
Regular Collab.		[0.580]		[0.916]		[0.688]
After Death ×		-1.406		-1.027		-1.721
Close Collab.		[0.932]		[0.976]		[1.489]
Dep. Var Mean	19.406		21.059		18.175	
R^2	0.039	0.039	0.042	0.043	0.037	0.037
Nb. of Obs.	216,746	216,746	91,620	91,620	125,126	125,126
Nb. of Dyads	8,220	8,220	3,509	3,509	4,711	4,711
Nb. of Superstars	161	161	69	69	92	92

Replication of the results in Table 5A, using OLS to estimate the effects of superstar extinction. All models incorporate year effects and 50 age category indicator variables. The results in column 1a imply that coauthors suffer a 1.38 yearly decline in JIF-weighted publication output following the death of their superstar collaborator. This represents a 7.11% decrease relative to the mean of the dependent variable at the time of death. The figure matches up closely with the 7.40% decline from Table 5A, column 1a. A similar exercise can be performed for columns 2a and 3a, and the departures betwen the magnitudes implied by OLS and QML Poisson are once again very slight. Absolute value of robust standard errors in brackets, clustered at the level of the superstar.

†significant at 10%; *significant at 5%; **significant at 1%.

Figure A1: Coauthor Matching for a Sample Publication

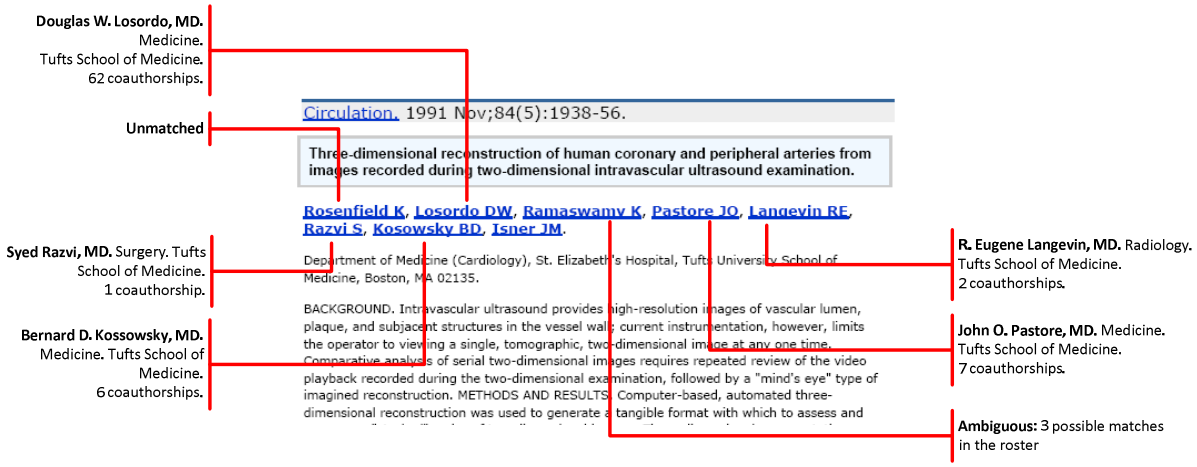


Figure A2: Vintage-specific Empirical Distributions for the Distribution of Citations at the Article-level