DIVERSE EXPERTISE, PEER EFFECTS AND RESEARCH PRODUCTIVITY
Does diversity in idea space matter?

WEI SI *and* QI WANG

Stockholm University *and* KTH-Royal Institute of Technology

Preliminary Draft: Please do not circulate or cite

May 15, 2016

Abstract

This paper empirically explores whether cognitive diversity between collaborators affects peer effects in knowledge production. In order to measure the cognitive diversity between knowledge producers precisely, we introduce and calculate a novel index, namely the *cognitive distance* between two researchers, based on their publication distributions and citation relations across academic journals in which they have publications. Using individual-level panel data from the Thomson Reuters Web of Science (WoS) database of academic papers published from 1980 to 2013, we estimate the changes in productivity of the coauthors of 63 active life scientists who passed away unexpectedly and prematurely between 1995 and 2009 and examine whether the effect of this adverse shock to coauthors of a deceased scientist differs in terms of the *cognitive distance* between the coauthor and the deceased scientist. The results show that, following the death of an active and prominent life scientist, coauthors with close cognitive distance from the deceased scientist are more likely to experience a lasting decrease in productivity. While the cognitive distance between a coauthor and a deceased scientist attenuates the negative shock, the relationship appears to be non-linear for different measures of productivity. The results indicate that both knowledge spillovers and skill complementarity play a role in collaborations; the former has impacts on both quantity and quality of the research productivity, while the latter mainly affects the research quantity. Knowledge spillovers are more likely to occur between two researchers who are cognitively close in their idea space, while coauthors who are cognitively distant may be more likely to benefit from skill complementarity. The findings suggest that, after losing a preeminent collaborator, it is the loss of an irreplaceable source of ideas that has a more adverse impact on a scientist's productivity than the potentially imperfect skill substitution that follows such a loss.

Keywords Peer Effect, Knowledge Spillover, Cognitive Diversity, Research Productivity

JEL classification: J24, O31, D83.

*Wei Si: PhD student at the Department of Economics at Stockholm University, email: wei.si@ne.su.se. Qi Wang: PhD student at the Department of Industrial Economics and Management at KTH-Royal Institute of Technology, email: qi.wang@indek.kth.se. We are grateful to Peter Fredriksson, Andreas Madestam, Jan Sebastian Nimczik, and seminar and conference participants at Stockholm University, Stockholm School of Economics, and Warwick Economics PhD Conference 2016 for their insightful comments. We especially thank Ludo Waltman and the Centre for Science and Technology Studies (CWTS) of Leiden University for providing the data used in this study. We acknowledge the IAAE travel grant for the financial support to attend the 3rd IAAE Conference.

1 Introduction

It is widely recognized that the development of science and technology plays a fundamental role in economic growth. Endogenous growth theory points out that human capital enables increasing returns to scale and thus sustainable economic growth thanks to substantial knowledge spillovers and positive externalities (Lucas, 1988; Romer, 1990). However, the process and mechanism of knowledge production are less thoroughly understood. Exploring the "black box" of knowledge generation and understanding the mechanism behind it are vital for promoting innovation, allocating human capital resources efficiently, and facilitating sustainable economic growth. As knowledge is systematically created by knowledge producers in modern society and nowadays increasingly through the teamwork of scientists (Wuchty et al., 2007; Jones, 2009), understanding the mechanism of collaboration and spillovers among knowledge producers is essential for unpacking the black box of knowledge creation.

Empirical evidence for interpreting the mechanism of knowledge production and interaction among knowledge producers is insufficient and inconclusive. One of the challenges to conducting a causal inquiry is a consequence of the selection and endogeneity issues that arise because scientific production is usually a goal-oriented process and collaboration rarely takes place at random. Recent empirical work on the peer effect and productivity attempts to confront this challenge by exploiting exogenous supply shocks in different quasi-natural experiment settings, but results have been inconsistent. On the one hand, the study by Azoulay et al. (2010) of the unexpected and premature deaths of eminent life scientists on their coauthors finds that the impact on the productivity of the latter is negative and lasting. Waldinger (2010) makes use of the historical event of the dismissal of professors by the Nazi government and finds that the research accomplishments of PhD students of dismissed supervisors suffered both in the short- and long-term. On the other hand, in a similar setting, Waldinger (2012) finds no peer effect on the research output of colleagues in the same departments from which scientists have been dismissed. Moreover, Borjas and Doran (2012) show that the immigration of former Soviet mathematicians to the US had adverse impacts on the output of incumbent US mathematicians who worked on similar research questions. The divergent findings contributed by previous research suggest that further studies are needed in order to distinguish various factors that may influence the peer effect and the productivity of knowledge producers in different ways. Borjas and Doran (2015b) provide such an analysis by disentangling and defining three types of peer effect, namely geographic, intellectual, and collaborative.

Following the lead of previous research, this paper focuses on the impact of the cognitive diversity between collaborators on knowledge production and explores whether this factor affects the magnitude of spillovers among collaborators as well as their productivity. The key measure here is the cognitive diversity, in other words, the diversity in idea space of collaborators, which emphasizes differences in mental process that may affect the process of knowledge production through teamwork. The topics of cognitive diversity among or within teams have caught the attention of researchers in a variety of disciplines, including management, psychology, and education, and *cognitive diversity* has been defined in myriad ways in different research contexts (Mello and Rentsch, 2015). Economists, however, have made surprisingly little inquiry into the effect of cognitive diversity on team performance and productivity. In this paper, we explore the causal relation between cognitive diversity and peer effects using evidence from the academic collaborations of researchers in life sciences. Cognitive diversity in this research context refers to differences in researchers' professional attributes, including expertise, knowledge, skills, and research interests and strategies, that underlie their previous research.

As collaboration is not a randomly-formed process, we follow the same identification strategy as Azoulay et al. (2010) that exploits unexpected and premature deaths of researchers as a quasi-experimental variation to tackle the endogeneity and selection issues[1]. More specifically, we identify 63 active and prominent life scientists who died suddenly and prematurely between 1995 and 2009 and estimate the changes in productivity of their coauthors to determine if the changes differ in the cognitive diversity between a deceased researcher and his or her coauthors. To measure and delineate the cognitive diversity between two researchers, we introduce a new index called *cognitive distance*, which is calculated using information of each scientist's publication distributions and citation relations across academic journals in life sciences. Using individual-level panel data from the Thomson Reuters Web of Science (WoS) database of academic papers in life sciences published from 1980 to 2013, we construct a panel data set of 9605 coauthor-deceased scientist pairs, and measure coauthors' research productivity using both the annual publication rates and citation impacts before and after the deaths of the prominent scientists.

Our theoretical framework, building on standard knowledge production models (Borjas and Doran, 2015b) and evidence of the relationship between cognitive distance and absorptive ca-

[1]Using unexpected and premature death of individuals as an exogenous source of variation has been applied in other recent studies, e.g., Aizenman and Kletzer (2011), Oettl (2012), Azoulay et al. (2015) and Mohnen (2016).

pacity from firms in R&D alliances (Nooteboom et al., 2007), suggests that cognitive diversity between collaborators can affect the changes in productivity after a labor supply shock. On the one hand, if collaboration is mainly for efficient production, then a smaller cognitive distance between coauthors tends to generate larger productivity impacts by enabling larger knowledge spillovers and absorptive capacity between coauthors. On the other hand, a large cognitive distance between coauthors also affects productivity by facilitating skill complementarity, since the elasticity of complementarity between coauthors in collaboration increases with larger cognitive distance. Overall, our theoretical framework predicts that a coauthorship network with smaller average cognitive distance between coauthors is more likely to experience larger productivity impacts caused by a supply shock than a coauthorship network with larger average cognitive distance if the spillover effects dominate the complementarity effects, and vice versa.

Consistent with the prediction from the theoretical framework, our main results from the empirical analysis show that, following the death of an active and prominent life scientist, coauthors with close cognitive distance from a deceased scientist experience most severe and lasting decreases in research productivity, compared with coauthors with more distant cognitive distance, especially in measures that place greater stress on research quality. However, for measures that mainly reveal the quantity of research, the relationship between cognitive distance and productivity changes appears to be non-linear. Both cognitively close and cognitively distant coauthors experience statistically significant decreases after the loss of an active collaborator, compared with coauthors at the middle level of cognitive distance. These findings indicate that both knowledge spillovers and skill complementarity may play a role in collaboration, yet the former has impacts on both quantity and quality of research output while the latter mainly affects quantity. Knowledge spillovers seem to be circumscribed within certain levels of cognitive diversity among researchers. Overall, after losing a preeminent collaborator, the loss of an irreplaceable source of ideas causes a more adverse impact on a survivor's productivity than the potentially imperfect skill substitution.

The paper contributes to the existing literature by providing empirical evidence of the causal relation between cognitive diversity and peer effects in scientific research. As we focus on the academic collaborations, our results may have policy implications for issues concerning public R&D funds allocation, university personnel, high-skill immigration, and innovation promotion, and so forth.

The rest of the paper is organized as follows. The next section presents the motivation for studying and the hypothesis regarding the potential impact of cognitive distance on peer effect and productivity in scientific collaborations, and goes on to describe the construction of the cognitive distance index. Section 3 presents a theoretical framework based on standard knowledge production models. Section 4 illustrates our empirical strategy and sets up the empirical models. In Section 5, we describe the data and the construction of the samples of the deceased scientists and their coauthors. Section 6 reports the empirical results, and the final section offers a summary, discussion, and conclusions.

2 Cognitive Diversity in Collaboration

2.1 Related Literature on Diversity, Motivation, and Hypothesis

There has been increasing interest in studying the relationship between diversity variously measured and the level of intellectual productivity. Yegros Yegros et al. (2015) use an inverted U-shape to describe relationship between interdisciplinary research, which is assessed using disciplinary diversity in the references of a publication, and citation impact. Freeman and Huang (2015) show that diversity among authors of a paper in terms of ethnicity, location, and references is positively associated with the paper being published in higher-impact journals and receiving more citations than papers that are less diverse in these respects. Both studies reveal interesting correlations between diversity and the quality of publications. Furthermore, as alluded to above, a series of studies by Borjas and Doran (2012, 2015a,b) makes use of the immigration of high-skilled mathematicians to the US after the collapse of the Soviet Union and provides causal evidence that the exogenous supply increase of mathematicians from the Soviet Union generated negative impacts on their US counterparts and positive impacts on remained Soviet counterparts with whose research theirs overlapped because of competition effects. Borjas and Doran (2015b) also find that net positive spillovers are likely to occur among coauthors from a very high-quality researcher to others. Our study differs from previous work by providing causal evidence for an exogenous decrease in the supply of researchers, and by proposing a more generalized yet precise procedure to measure cognitive diversity between researchers.

As stated, in order to explore whether cognitive diversity affects the magnitude of spillovers between collaborator, we utilize evidence from academic collaborations in life sciences. As Wuchty et al. (2007) have demonstrated, research is increasingly done through collaborations among scientists instead of solo work; and teamwork is positively associated with high impact research and greater individual academic productivity (Ductor, 2015). Since there is a long tradition of collaboration in the academic life sciences, focusing on coauthorship in this field facilitates analysis owing to the availability of rich empirical evidence. Jones (2009) points out that individuals face an increasing burden of knowledge, and thus choose narrower expertise, which forces them to work in teams. From this perspective, a team of knowledge producers usually contains a diverse pool of expertise. It is therefore natural to ask whether a causal relationship exists between the degree of diversity in expertise and the productivity of individual knowledge producers.

Intuitively, cognitive diversity among collaborators may affect productivity in many different ways. On the one hand, the relative closeness of the knowledge bases of collaborators may facilitate more direct and effective knowledge spillovers and absorptive capacity among them in cases where spillovers or mutual learning are more likely to take place within a certain range of cognitive levels. Thus Azoulay et al. (2010) provide suggestive evidence for the view that spillovers are more likely to be circumscribed in ideas space[2]. On the other hand, coauthors with a high level of cognitive distance are more likely to come from different sub-fields; and a relatively broad range of knowledge and skills may benefit a research team because of the skill complementarity that may facilitate the accomplishment of large and comprehensive projects. Bridging and connecting diverse knowledge may be especially effective in promoting innovations when creativity requires fresh ideas from various perspectives. Nooteboom et al. (2007) present evidence from the innovation performance of firms to show that an absorption effect decreases while a novelty effect increases with larger cognitive distance between firms cooperating in technology-based alliances. Likewise, the cognitive diversity among collaborators could be a crucial factor in the process of knowledge creation, with optimum levels of diversity contributing to valuable research outcomes.

2.2 Constructing Index of *Cognitive Distance*

In order to estimate the potential effect of the cognitive diversity among collaborators, we propose an index called *cognitive distance* to measure and delineate the extent of the cognitive diversity between two researchers. In this empirical setting, we calculate the cognitive distance

[2]Our analysis puts forward the work of Azoulay et al. (2010) on this point by proposing a more rigorous measure of cognitive diversity among researchers and examining more intensively the potential impacts of different cognitive levels.

between a deceased scientist and each of his or her coauthors.

The index combines the publication distributions of researchers across journals with the similarity of journals, which is measured by the citation relations between any two journals in which researchers have publications. A large number of citations between articles published in two academic journals suggests that papers from these two journals are proximate in research topics and methodologies, and *vice versa*. This "research proximity" is captured to reveal the cognitive diversity in the context of scientific research. More specifically, we calculate the index of cognitive distance in three steps. First, we create a citation matrix $[c_{ab}]$ that contains direct citation relations between any two journals that were active between 1980 and 2013 and are in the WoS database[3]. For each pair of a citing journal (Journal a) and a cited journal (Journal b), c_{ab} denotes the number of citations by publications in Journal a of publications in Journal b from 1980 to 2013.

The second step is to set up a distance matrix $[d_{ab}]$ that measures the cognitive distance between any two journals in our database. We begin by constructing a similarity matrix $[s_{ab}]$ of journals based on the citation matrix created in the first step. The cosine similarity measure[4] (Salton and McGill, 1983) is used to calculate the similarity of each pair of journals in the database. For instance, the similarity of Journal a and Journal b is given by

$$s_{ab} = \frac{\sum_k c_{ak} c_{bk}}{\sqrt{(\sum_k c_{ak}^2)(\sum_k c_{bk}^2)}} \tag{1}$$

where k denotes any one of the journals in our database. Thus, the distance between Journal a and Journal b is defined as

$$d_{ab} = 1 - s_{ab} \tag{2}$$

Each element d_{ab} of the distance matrix $[d_{ab}]$ varies between 0 and 1. Similar ways of measuring the distance between two journals have been used in such studies as those of Leydesdorff and Rafols (2011); Rafols et al. (2012). In addition, Fafchamps et al. (2010) have applied the cosine similarity measure in order to calculate an index of research overlap among economists based on the JEL codes of their papers.

[3]Due to the accessibility of our database, we only have complete information on publications published since 1980.

[4]The cosine similarity measure is a standard measure of similarity, which is widely used in computer sciences, e.g., for developing search engines, and has been applied in many other sciences as well.

Finally, the cognitive distance between a researcher i and his or her deceased collaborator j is measured as a weighted average of journal distance using each researcher's publication distributions across journals before the death of researcher j. Let n_i, n_j, n_i^a, n_j^b, and d_{ab} denote the total number of papers of researcher i, the total number of papers of researcher j, the number of papers of i published in Journal a, the number of papers of j published in Journal b, with the distance between Journal a and Journal b calculated in the second step; under these circumstances, the index of cognitive distance between researchers i and j is defined as

$$Distance_{ij} = \frac{\Sigma\Sigma n_i^a n_j^b d_{ab}}{n_i n_j} \tag{3}$$

The cognitive distance between two researchers ranges from zero, if each pair of journals in which i and j have publications has zero distance, to one, if i and j published in totally different journals that have no citation relations. For each pair of a deceased scientist and his or her coauthor, the cognitive distance is calculated based on their publication history before the year in which the scientist passes away, and the index is time-invariant. Figure 1 displays the distribution of the index of cognitive distance between each pair of a deceased author and his or her coauthor in the sample that has been calculated using the procedures described above.

We employ the citation relations of journals to calculate the cognitive distance between researchers as an indirect measure in preference to such direct measures as using the citation relations of researchers to depict the similarity or differences of their research. This preference is primarily due to the half-life of life sciences; that is, research papers have lifespans, thus the number of citations a paper received will decrease eventually after it has published for a certain period (Davis, 2013). Hence, if two researchers are active in quite different time periods, they may have loose citation relations, given the rapid rate of change in life sciences, in which case it is hardly possible to conclude that the two researchers differ in the focus of their work. As a result, direct use of the citation relations between the articles of two researchers in order to calculate their cognitive distance has the potential to yield incomplete and potentially biased information regarding their research proximity[5]. By contrast, the cognitive distance between journals in which researchers have publications can offer a more comprehensive measure of their cognitive diversity, since this indirect indicator takes into account the citation history between academic journals.

[5]The same problem will occur if we calculate the index of cognitive distance between two researchers based on the overlap of references in their articles (Wang and Sandström, 2015).

In addition, compared with methods used in measuring research proximity or research overlap in previous studies, the index of cognitive distance proposed here has several merits. First, rather than specific to particular disciplines, it is more generalized and could be applied into other sciences. On the contrary, indicators of authors' research overlap constructed in Borjas and Doran (2012) and Fafchamps et al. (2010) are specific to mathematics and economics, respectively. Second, compared with methods that infer research proximity from the overlap in subject terms, topics, or keywords as in (Azoulay et al., 2010, 2015; Wang and Sandström, 2015), which focus on specific contents of articles, our index depicts a broader and more dynamic picture by making use of the history of citation relations across journals in which researchers have publications, and thus is a more appropriate measure of researchers' cognitive diversity in this research context.

3 Theoretical Framework

We follow the theoretical framework of Borjas and Doran (2015b) that emphasizes the role of human capital externalities in knowledge production models. Suppose the aggregate production function of knowledge in academic life sciences, Y, depends on the stock of ideas, I, the stock of capital (research inputs such as laboratories, apparatus resources, and faculty slots, etc.), K, and the stock of life scientists, L. Assume constant returns to K and L, and the production function is given as Cobb-Douglas,

$$Y = I^{\phi} K^{1-\alpha} L^{\alpha}, \tag{4}$$

where ϕ is the externalities elasticity, which measures the percentage change in knowledge in life sciences in response to a one percent change in the stock of ideas. It is also assumed that the stock of ideas is proportional to the number of life scientists.

Further, we define the total labor supply of life scientists, L, in terms of the contribution of the supply in different coauthorship networks, L_n, that is,

$$L = (I_1^{\eta} L_1^{\rho} + \cdots + I_N^{\eta} L_N^{\rho})^{\frac{1}{\rho}}, \tag{5}$$

where I_n is the stock of ideas in the coauthorship network n and L_n is the size of the life sciences workforce in the coauthorship network n. η is the local externalities elasticity, which measures the magnitude of the coauthorship network-specific spillovers. The elasticity of substitution between life scientists in different coauthorship networks is $\sigma_n = \frac{1}{1-\rho}$.

Furthermore, we define the labor supply of life scientists in each coauthorship network, L_n, in terms of the contribution of each life scientist, L_i, in that coauthorship network, that is,

$$L_n = (I_i^\theta L_i^\lambda + \cdots + I_j^\theta L_j^\lambda)^{\frac{1}{\lambda}}, \tag{6}$$

where I_i is the stock of ideas that scientist i provides and L_i gives the labor input of scientist i. θ is the collaboration-specific externalities elasticity that measures the magnitude of human capital spillovers from the scientist i to his or her coauthors. The elasticity of substitution between coauthors within each coauthorship network is $\sigma_c = \frac{1}{1-\lambda}$, where λ determines the degree of substitutability among coauthors within each coauthorship network, and $\frac{1}{\sigma_c}$ is the elasticity of complementarity between coauthors within each coauthorship network.

To investigate whether the cognitive diversity among coauthors would affect peer effects in collaboration, let d be the average cognitive distance between coauthors in a coauthorship network. By definition, it is likely that the magnitude of elasticity of substitution between coauthors in each coauthorship network decreases with larger cognitive distance, i.e.,

$$\frac{\partial \sigma_c}{\partial d} < 0.$$

Since a smaller cognitive distance implies that coauthors are more cognitively similar, the magnitude of elasticity of substitution should be larger; while a larger cognitive distance suggests that coauthors are more distinct in knowledge and skill bases, thus the magnitude of elasticity of substitution should be smaller. It implies that a high level of cognitive diversity may contribute to skill complementarity in collaboration.

Moreover, as discussed in the hypothesis of the potential effect of cognitive diversity, knowledge spillovers and absorptive capacity are likely to decrease with cognitive distance, while novelty effect tends to increase with cognitive distance. We follow a similar structure as Nooteboom et al. (2007) and describe the relationship between knowledge spillovers (or absorptive capacity) and cognitive distance as a downward sloping line, that is,

$$K = \kappa_1 - \kappa_2 d, \tag{7}$$

where $\kappa_1, \kappa_2 > 0$, and the relationship between novelty effect and cognitive distance as a upward sloping line, that is,

$$N = \nu_1 + \nu_2 d, \tag{8}$$

where $\nu_1, \nu_2 > 0$. Suppose the collaboration-specific externalities elasticity is a product of knowledge spillover and novelty effect, thus

$$\theta = KN = (\kappa_1 - \kappa_2 d)(\nu_1 + \nu_2 d) = \kappa_1 \nu_1 + (\kappa_1 \nu_2 - \kappa_2 \nu_1)d - \kappa_2 \nu_2 d^2. \tag{9}$$

Eq.(9) implies that

$$\text{If } d < \frac{\kappa_1 \nu_2 - \kappa_2 \nu_1}{2\kappa_2 \nu_2}, \text{ then } \frac{\partial \theta}{\partial d} > 0; \text{ If } d > \frac{\kappa_1 \nu_2 - \kappa_2 \nu_1}{2\kappa_2 \nu_2}, \text{ then } \frac{\partial \theta}{\partial d} < 0.$$

There appears to be an inverted-U-shaped relationship between the collaboration-specific externality and the average cognitive distance between coauthors in a coauthorship network. The shape of the curve is determined by the values of κ_1, κ_2, ν_1, and ν_2. For instance, if collaboration is only for efficient knowledge production, then the consideration of novelty would be redundant and the novelty value of cognitive distance would be zero, i.e., $\nu_2 = 0$; in that case, the collaboration-specific externalities elasticity only declines with the cognitive distance, i.e.,

$$\frac{\partial \theta}{\partial d} < 0.$$

Altogether, we have set up a three-level CES nesting model in Eq.(4), (5), and (6). To examine the change in marginal product for a life scientist i from a coauthorship network n following a labor supply change of scientist i, we define the percentage change in the total supply of efficiency units of life scientists as $l = d \log L$, the percentage change in the efficiency units of life scientists in coauthorship network n as $l_n = d \log L_n$, and the percentage change of labor supply of the scientist i in efficiency units as $l_i = d \log L_i$. Then, we can get the change in the marginal product of life scientist i as[6]

$$d \log MP_i = s_k d \log K + (\phi - s_k + \frac{1}{\sigma_n})l + (\eta - \frac{1}{\sigma_n} + \frac{1}{\sigma_c})l_n + (\theta - \frac{1}{\sigma_c})l_i, \tag{10}$$

[6]The proof of Eq.(10) is given in Appendix C.

where $s_k = 1 - \alpha$ is the capital's share of output in the aggregate production function in Eq.(4). Moreover, by definition, capital resources K are fixed in the short-run, i.e., $d\log K = 0$, and can be fully adjusted in the long-run. Assume the market of academic life sciences are competitive, then the long-run equilibrium gives $d\log K = d\log L = l$. Hence, the short- and long-run productivity impacts following a labor supply change of scientist i are

$$d\log MP_i = \begin{cases} (\phi - s_k + \frac{1}{\sigma_n})l + (\eta - \frac{1}{\sigma_n} + \frac{1}{\sigma_c})l_n + (\theta - \frac{1}{\sigma_c})l_i, & \text{if short-run} \\ (\phi + \frac{1}{\sigma_n})l + (\eta - \frac{1}{\sigma_n} + \frac{1}{\sigma_c})l_n + (\theta - \frac{1}{\sigma_c})l_i, & \text{if long-run.} \end{cases} \tag{11}$$

Therefore, the cognitive distance (d) between coauthors could potentially affect the magnitude of the short- and long-run productivity impacts by affecting the levels of θ and σ_c. As shown in Eq.(11), the level of cognitive distance (d) could influence the impact of l_i by changing the magnitude of spillovers of ideas (θ) between coauthors and relaxing the scarcity constraints ($-\frac{1}{\sigma_c}$) that caused by the diminishing returns. The level of cognitive distance (d) could also affect the impact of l_n by changing the effect of skill complementarity ($\frac{1}{\sigma_c}$) in a coauthorship network.

Consider a simple case when $\frac{\partial \sigma_c}{\partial d} < 0$ and $\frac{\partial \theta}{\partial d} < 0$ (if $\nu_2 = 0$), assume there is a negative supply shock (i.e., $l_i < 0, l_n < 0, l < 0$), then the larger the d, the larger the $\frac{1}{\sigma_c}$ and the smaller the θ, and the smaller the $(\theta - \frac{1}{\sigma_c})$, while the smaller the d, the smaller the $\frac{1}{\sigma_c}$ and the larger the θ, and the larger the $(\theta - \frac{1}{\sigma_c})$, thus a smaller d will induce a larger negative impact of l_i. However, d also affects the impact of l_n through the effect of skill complementarity, and a larger d will cause a larger negative impact of l_n. Overall, it implies that a coauthorship network with smaller average cognitive distance between coauthors tends to experience larger productivity impacts caused by a supply shock than a coauthorship network with larger average cognitive distance if the spillover effects dominate the complementarity effects, and vice versa.

4 Empirical Strategy

4.1 Identification Strategy

Since coauthorship is not generated randomly, one of the empirical challenges in this study is to identify exogenous changes in the coauthoring ties. Our identification strategy depends on a quasi-experimental setting in which the unexpected and premature death of an active life scientist provides an exogenous shock to the productivity of his or her coauthors, who in turn have different levels of cognitive distance with respect to the deceased scientist. For each focal scientist whose unforeseen death serves as a source of exogenous variation, we estimate the changes in research productivity of other scientists who have been coauthors with the focal scientist after the loss of this coauthor in order to determine how the effect may differ in terms of the cognitive distance between each coauthor and the deceased scientist. Therefore, we employ a difference-in-differences strategy, where the "treatment" for a coauthor of a deceased scientist is his or her cognitive distance from the latter. In other words, for each deceased scientist, we identify the "treatment effect" on his or her coauthors by comparing the changes in the productivity of those who have different levels of cognitive distance from this scientist before and after his or her death.

4.2 Empirical Models

The empirical model is as follows,

$$Y_{ijt} = \exp[\beta_0 + \beta_1 Death_{jt} + \beta_2 Distance_{ij} \times Death_{jt} + f(Career\_Age_{it}) + \gamma_t + \delta_{ij} + \epsilon_{ijt}] \quad (12)$$

where the dependent variable Y_{ijt} is the research output of a scientist i in year t who coauthored one or more papers with a deceased scientist j. The dependent variables are measured using both the publication rates and the citation impacts. The first outcome variable is the number of journal articles that Scientist i published in each publication year. To adjust the research quality better, we further weight each article by its journal impact factor (JIF)[7] in order to determine the number of JIF-weighted articles. However, the journal impact factor to some extent averages the impact of individual articles included in the same journal, so the number of JIF-weighted publications may represent a crude measure of the research quality for individual researchers. For this reason, we also include the number of citations of each researcher in order to assess a researcher's productivity. However, the absolute number of citations for each paper will bias the estimation because of a truncation problem regarding citation data; that is, papers that published earlier are likely to receive more citations than those published later. We therefore instead count for each researcher the number of papers that fall within the top tail of the distribution of citations among all life sciences-related papers published in the same year. To do this, we first compute the

[7]The impact factor of an academic journal is a measure of the average number of citations received by articles published in this journal in a certain period. We use the two-year JIF of each journal to calculate the number of JIF-weighted publications variable.

article-level distribution of total (until 2013) citations for articles in the broadly defined categories of life sciences that were published in each year from 1980 to 2013 based on the WoS subject categories. For instance, 43 citations are required in order to rank above the 75th (top 25th) total citation percentile for papers published in 2000, and 33 citations for papers published in 2005, based on the information in our database. We next count for each researcher, the number of papers that he or she has published in each given year that falls within the 50th and 75th citation percentile among all papers published in the same year as another set of dependent variables that evaluate a researcher's citation impact.

The variables of interest in our empirical model in Eq.(12) are $Death_{jt}$, an indicator that equals 1 the year after Scientist j passes away, and $Distance_{ij}$, the index of cognitive distance between Scientists i and j as measured in Section 2. The coefficient β_1 captures the net change in productivity of Scientist i after a coauthor j dies regardless of their cognitive distance, and β_2 captures the change in productivity of Scientist i after a coauthor j dies that interacts with the level of the cognitive distance between Scientists i and j. $f(Career\_Age_{it})$ is a function that flexibly controls for the career age[8] of Scientist i, so that factors that change over a scientist's career life cycle, such as academic incentives, which affect research productivity will be taken into account. γ_t controls for publication year fixed effects. δ_{ij} controls for pairwise fixed effects that absorb all unobservable time-invariant characteristics that correspond to the coauthorship dyad. ϵ_{ijt} is an error term.

Considering the potential non-linear relationship between cognitive distance and peer effects, we further propose two additional models as follows,

$$\begin{aligned} Y_{ijt} = \exp[&\beta_0 + \beta_1 Death_{jt} + \beta_2 Distance_{ij} \times Death_{jt} + \beta_3 Distance_{ij}^2 \times Death_{jt} \\ &+ f(Career\_Age_{it}) + \gamma_t + \delta_{ij} + \epsilon_{ijt}] \end{aligned} \tag{13}$$

$$\begin{aligned} Y_{ijt} = \exp[&\beta_0 + \beta_1 Death_{jt} + \beta_2 Close_{ij} \times Death_{jt} + \beta_3 Distant_{ij} \times Death_{jt} \\ &+ f(Career\_Age_{it}) + \gamma_t + \delta_{ij} + \epsilon_{ijt}] \end{aligned} \tag{14}$$

To begin with, the model in Eq.(13) adds an interaction term between the indicator $Death_{jt}$ and a quadratic form of the cognitive distance between Scientists i and j in order to determine whether there is a curvilinear relationship between cognitive distance and the magnitude of spillovers.

[8]Since we do not have the information of the exact year when a scientist obtains his or her highest degree, we calculate the career age of Scientist i as the differences between the publication year t and the year when Scientist i publishes his or her first article. The career age cohort dummies are created as three-year intervals.

Furthermore, we divide the coauthors of deceased scientists into three groups: close distance coauthors, middle distance coauthors, and distant distance coauthors, based on the percentile rank of the cognitive distance between a scientist and his or her late coauthor. Figure 2 plots changes in different measures of scientists' research output after they lose their coauthors against the percentile rankings of their cognitive distance over the full sample. Plots of each measure of research outcomes, i.e., the number of publications, the number of JIF-weighted publications, the number of publications that rank above the 50th percentile of citations, and the number of publications that rank above the 75th percentile of citations are presented. Figure 2 provides a first clue that the impact of a researcher's unexpected death on his or her coauthors may differ in the pairwise cognitive distance. Based on the trend of changes in productivity revealed roughly by the plots, we define the close distance group as coauthors whose cognitive distance from the deceased scientist falls into the interval of percentile rank (0th, 50th), the middle distance group as coauthors whose cognitive distance falls into the interval of percentile rank (50th, 85th), and the distant distance group as coauthors whose cognitive distance falls into the interval of percentile rank (85th, 100th)[9]. We therefore propose the model in Eq.(14) with the key coefficients β_2 and β_3 that capture the changes in productivity of Scientist i after his or her coauthor j dies, depending on whether i is a close distance coauthor or distant distance coauthor of j, relative to coauthors in the middle distance group with the same deceased collaborator. The model in Eq.(14) would provide a more straightforward interpretation of any potential non-linear relationship between cognitive distance and peer effects.

Considering the highly skewed and non-negative nature of the dependent variables, and following the standard practice in research on R&D and scientific productivity, we estimate the above models using the fixed effects Poisson quasi-maximum likelihood estimator (QMLE) developed by Hausman et al. (1984), which allows for consistent parameter estimates of even non-count dependent variables (Silva and Tenreyro, 2006). Poisson QMLE standard errors are consistent without assuming a Poisson distribution of the underlying data, and are robust in the face of arbitrary serial correlation (Wooldridge, 1999). We cluster the standard error at the deceased scientist level. Additionally, we also estimate the models using OLS estimation, which yields results similar to those obtained using the Poisson QMLE specifications[10].

[9]The cut-off points are chosen based on the trend charts in Figure 2. The results, however, do not sensitive to the exact cut-off points. For instance, if we move the cut-off point 50th to 45th or 55th, or the cut-off point 85th to 80th or 90th, the main results in the following are still preserved.

[10]The results of OLS estimation are presented in Tables A.1 and A.2 in Appendix A.

5 Data and Sample Construction

5.1 Data

We collected data for the empirical analysis from the Thomson Reuters Web of Science (WoS) database. Since the study is focused on life scientists, we collected publication information in the broadly defined field of life sciences, namely 30 life science-related WoS subject categories, such as biology, biomedical sciences, microbiology, and medicine[11]. We retrieved the publications from 1980 to 2013, and only included articles published in academic journals; working papers, meeting abstracts, editorial materials, and other types of documents from the WoS database are excluded to ensure a collection of original and unrepeated research output of life scientists.

5.2 Sample of Deceased Scientists

One of the fundamental aims of the empirical analysis is to identify accurately a sample of active life scientists who passed away unexpectedly and prematurely. We limit the year of death to between 1995 and 2009 in order to be able to observe the changes in productivity of deceased scientists' coauthors within reasonable time windows before and after the unexpected deaths. Moreover, the intent is to identify those scientists who had accomplished distinguished research achievements by the time of their deaths, since, as Azoulay et al. (2010) observe, the impact on coauthors from the death of a mediocre researcher may not be as detectable as that caused by the death of a "superstar". Hence, we begin the search by tracing the trend in researchers' annual number of publications. It is reasonable to assume that, after the sudden death of a preeminent researcher, the annual number of publications will decrease sharply in the following years, rapidly reaching zero. In addition, by starting from the annual number of publications, it is easier to identify the superstars based on their research output. To be more specific, we selected researchers who have had more than 50 publications in total, and identified those whose trend in the annual number of publications experienced a sharp and non-rebounding decline over a two-year period, and who have more than 20 publications in the beginning of this descent period and less than three afterward. It is possible that some articles published after the sudden death of a researcher represent unfinished studies. Figure B.1 illustrates an example of the pattern of a researcher's annual number of publications as we have traced them. Based on these criteria,

[11] See Table B.1 in Appendix B for a full list of disciplines in life sciences included in the sample of analysis.

we identify 102 researchers from the filed of life sciences. However, there are obviously other reasons that could lead to a similar pattern in the trend in the annual number of publications of researchers, for instance, retirement or leaving academia to become a professional physician, which is not uncommon for researchers with an M.D. Therefore, we double-checked the sample of 102 researchers by searching their information from such sources as obituaries posted by their institutions or in journals for which they served as editors, personnel information from their departments, and their personal websites in order to distinguish those who experience sudden and premature death from those who ceased research for other reasons. We thus have excluded researchers who retired, do not work in academia, or passed away at an advanced age or not suddenly, as well as ones for whom we do not have definite information. In the end, we identified 63 active and preeminent life scientists who died unexpectedly and prematurely and for whom we determined the exact year of death between 1995 and 2009 as our sample of deceased scientists.

Table 1 provides summary statistics for the sample of deceased scientists, including the year of death, the year of first and last publication, career age, which is defined as the difference between the year of death and the year of first publication, the cumulative number of publications, the cumulative number of citations by the year of death, and the number of citations by career age, as well as the number of coauthors with whom each scientist published.

5.3 Sample of Coauthors

As the research focus of this study, the coauthors of deceased scientists are defined in the following ways. First, they are researchers who coauthored with any of the 63 deceased scientists at least once on academic papers published since 1980. That is, we focus on the one-degree coauthorship network of the 63 deceased scientists. Second, we only consider researchers who not only have coauthored with the deceased scientists but also have publications with others or on their own. In other words, we exclude coauthors whose only publications are with one of the deceased scientists. This restriction is intended to exclude current graduate students of the decreased scientists from the sample of coauthors, since the focus here is on faculty peer effects rather than mentorship impacts. These parameters yielded 9343 coauthors of the 63 deceased scientists. To estimate the changes in the research output of coauthors of the deceased scientists, the unit of analysis is a coauthor-deceased scientist-year triad. At the cross-sectional level, the unit is the coauthor-deceased scientist dyad. As there are a few researchers who have more than one coauthor from the sample of deceased scientists, the sample of coauthors contains 9605 coauthor-deceased scientist dyads. The estimation sample is a pool of observations in the period between the publication years 1980 and 2013 for the coauthor-deceased scientist dyads, which is an unbalanced panel containing 157,336 observations of the coauthor-deceased scientist-year triad.

Table 2 presents descriptive statistics for coauthors with covariates of interest measured in the year of death of the deceased scientist, for the entire sample as well as each group of coauthors. The average cognitive distance is 0.61 for the full sample of coauthors, and 0.48, 0.71, and 0.84 for the group of closed distance, middle distance, and distant distance coauthors, respectively. Measures of research outcomes include the number of raw and JIF-weighted publications and the number of papers that rank above the 50th and 75th citation percentile in each publication year; cumulative measures of these outcomes are listed as well. The career age of a coauthor here is defined as the difference between the year of death of the deceased scientist and the coauthor's first year of publication. A senior coauthor is a researcher whose career age is greater than that of his or her deceased coauthor. We count, for each coauthor, the number of cumulative citations until the year of his or her coauthor's death, and the number of citations by career age. Based on the number of cumulative citations of the coauthors in the sample, we define an elite coauthor as one whose cumulative citations rank above the top 10 percentile of the entire sample of coauthors. In addition, we also count the number of collaborations between a coauthor and a later deceased scientist between 1980 and 2013. A histogram of the number of collaborations shows that the distribution is highly skewed, as can be seen in Figure A.1. Based on the coauthorship intensity between two researchers, a regular coauthor is defined as one who collaborated more than five times with the deceased scientist. Moreover, we define a recent coauthor based on whether he or she has published with the deceased scientist within the three years prior to the latter's death. Finally, we determined whether a researcher and his or her deceased coauthor have worked in the same institution. As shown in Table 2, the covariates of interest for coauthors in the close distance, middle distance, and distant distance groups are fairly comparable.

A comparison of the average total number of publications, cumulative citations, and citations by career age of coauthors in the year of a deceased scientist's death with that of the deceased scientist, as shown in Table 1, confirms that the deceased scientists are indeed relatively more preeminent in view of their research outcomes.

Figure A.2 presents the trends in the average number of each research outcome by year for the close distance, middle distance, and distant distance groups without any adjustment for career age and publication year effects. The nonparametric plots provide a rough picture of the trends in coauthors' productivity within the three groups in terms of different levels of cognitive distance. It seems that, before the death of the deceased scientist, the productivity of coauthors in the three groups follows a parallel trend, while the trends after the death of the focal scientist appear to be different for different groups of coauthors. These plots reveal another preliminary clue that coauthors with different cognitive distance may be affected differently when they lose a preeminent coauthor.

6 Results

6.1 Event Study

To examine more precisely the trends in the annual research productivity of coauthors in different cognitive distance groups before and after they lose a coauthor, we perform an event study and plot the graphs in Figure 3. The event study graph shows the annual predicted means trend of each measure of research output of coauthors from the close distance, middle distance, and distant distance groups in each year before and after the unexpected death of a coauthor. We plot the graphs by taking predicted values of each research outcome using the Poisson QMLE models that control for career age and calendar year effects. The graphs show that, before the death of the deceased scientist, there is little difference in trends in each measure of research output among the close distance, middle distance, and distant distance groups. However, after the loss of the focal scientist, the trends in productivity of coauthors in each group become different, especially from the sixth year on. Considering that there are likely to be ongoing projects running when an active researcher passes away suddenly, it is reasonable that the changes in the productivity of coauthors gradually become obvious. The event study graphs thus suggest that the unexpected death of a preeminent scientist has impacts on coauthors' productivity, and that these impacts differ depending on the cognitive distance between a coauthor and the deceased scientist.

6.2 Main Results

To explore the causal relation between cognitive distance and peer effects, we propose the empirical models in represented Eq.(12), (13), and (14). Table 3 provides estimates of the effect of the unexpected death of prominent life scientists on their coauthors' publication rates. Columns (1) to (4) present coauthors' productivity changes in the number of publications, and columns (5) to (8) present the results for changes in the number of JIF-weighted publications.

To begin with, columns (1) and (5) show coefficient estimates of β_1 and β_2 for the model in Eq.(12). The results indicate that the cognitive distance between a coauthorship dyad attenuates the negative impact of losing a prominent coauthor on a researcher's productivity. However, the estimates are statistically significant only for measuring the impact on a researcher's number of JIF-weighted publications.

In order to determine whether there is a non-linear relationship between cognitive distance and the peer effects, the coefficient estimates of β_1, β_2, and β_3 for the model in Eq.(13) are presented in column (2) and (6). The results suggest that the impact of cognitive distance on the magnitude of peer effects seems to approximate an inverted-U-shaped relationship, but that this impact is only statistically significant for the impact on the number of publications.

Before moving to the model in Eq.(14), we estimate the changes in publication rates for coauthors in the close distance group compared with coauthors with greater cognitive distance in columns (3) and (7). The results show that cognitively close coauthors suffer the steepest decreases in the yearly number of JIF-weighted publications compared with others, but no statistically significant differences when there is no quality adjustment.

The model in Eq.(14) explores the non-linear relationship further and the coefficient estimates of β_1, β_2, and β_3 are shown in columns (4) and (8). For the measure of the yearly number of publications in column (4), the deaths of preeminent coauthors are associated with a 8.15%[12] decrease in the performance of researchers in the close distance group relative to the middle distance group. For researchers in the distant distance group, there is a 10.95% decrease relative to the middle distance group. There are no statistically significant changes in the yearly number of publications for coauthors in the middle distance group. The pattern is in accordance with an inverted-U-shaped relationship, as revealed in column (2). For the measure of the yearly number of JIF-weighted publications in column (8), coauthors in the close distance group suffer

[12]The results are calculated as $1-\exp[-0.085] = 8.15\%$, and so on.

a 17.80% decrease relative to the middle distance group. For coauthors in the middle and distant distance group, the changes in performance are not statistically significant. The results indicate that only cognitively close coauthors experience a severe decline in the number of JIF-weighted publications, which is in line with results in column (5) and (7). Overall, the findings from the models in Eq.(12), (13), and (14) are consistent with one another.

Table 4 provides estimates of the impact of the unexpected death of prominent life scientists on their coauthors' citation distribution. Columns (1) to (4) present the results of changes in the number of papers that rank above the 50th citation percentile among all papers published in the same year, and columns (5) to (8) are results for that of the 75th citation percentile. Estimates from columns (1) and (5) indicate that the magnitude of negative impact on coauthors' performance decreases in the cognitive distance between the coauthorship dyads. Results in column (2) suggest that cognitive distance may have an inverted-U-shaped impact on coauthors' numbers of articles that above the 50th citation percentile. Consistently, results in column (3) and (4) also show that relative to others, coauthors in the close distance group suffer the steepest decrease in this measure of output, and coauthors in the distant distance groups also experience a more severe decline than peers in the middle distance group. Finally, as shown in columns (7) and (8), for the measure of the number of papers above the 75th citation percentile, coauthors in the close distance group experience largest declines relative to others.

Comparing the results for the publication rates in Table 3 with those for the citation impacts in Table 4, we find that the effects on coauthors' numbers of papers and the number of articles above the 50th citation percentile share a similar non-linear pattern; that is, those in the close distance and distant distance groups experience larger declines in output than those in the middle distance group. For the impacts on coauthors' numbers of JIF-weighted papers and the number of articles above the 75th citation percentile, the declines in output are only statistically significant for coauthors in the close distance group. It is reasonable to consider outcomes for the number of publications and the number of articles that rank above the 50th citation percentile are measures that primarily reflect the quantity of research output, while outcomes for the number of JIF-weighted publications and the number of articles that rank above the 75th citation percentile place greater stress on the quality level.

Furthermore, we examined the dynamics of the effects on coauthors in the close distance and distant distance groups for each measure of research output relative to the middle distance group, namely the dynamics of the effects revealed in columns (4) and (8) in Tables 3 and 4, by estimating a specification in which the indicators for each coauthor group interact with a set of dummies corresponding to every particular year before and after the deaths of the prominent scientists[13] and plotting the coefficient estimates and the 95% confidence interval around them, as shown in Figure 4. The middle distance group is treated as the reference group. As illustrated in the plots, before a coauthor's unexpected death, the productivity of researchers in the close distance and distant distance groups relative to those in the middle distance group moves in parallel trends. After the loss of a prominent coauthor, the negative impact on researchers' productivity is lasting and irrecoverable for cognitively close ones in all measures of research output and for cognitively distant ones in the two measures of the quantity of research. The patterns of relative impact on the close distance and distant distance groups are consistent with that of the average effect revealed in Tables 3 and 4 for each measure of research outcomes.

As suggested in the theoretical framework, closeness in cognitive distance between coauthors may contribute to stronger knowledge spillovers and absorptive capacity, which lead to larger human capital spillovers, while a high level of cognitive distance increases the elasticity of complementarity between coauthors, thus collaborations between researchers with a large cognitive distance may be motivated by the complementarity of different skills. Under these circumstances, the negative impacts on the productivity of coauthors in the close distance group, as revealed by the main results, may result from a loss of influential ideas that directly affect the coauthors' research horizons, while the lost of a coauthor with distant cognitive distance would affect a researcher's productivity due to imperfect skill substitution. Our results suggest that, after losing a preeminent collaborator, the negative impact on productivity caused by the loss of an irreplaceable source of ideas overshadows that which is caused by the potentially imperfect skill substitution among the surviving researchers. The loss of ideas may have impacts on both the quantity and quality of research productivity, while the loss of skills mainly affects the research quantity.

[13]The interactions of group indicators with the year of death are omitted.

6.3 Robustness Checks

6.3.1 Subsample Estimations

To test for the robustness of the main results, we tried the models in Eq.(12) and Eq.(14) on subsamples. Results of the impact on coauthors' publication rates and citation impacts on different subsamples are presented in Table 5 to 8. In Table 5 and Table 6, columns (1) to (2) and (5) to (6) exclude publications that coauthored with the deceased scientists, namely only consider the changes in publication rates and citation impacts for publications of coauthors with other researchers or on their own before and after the deaths of the focal scientist, in order to examine whether the results are only caused by the loss of contributions from the deceased collaborators. Columns (3) to (4) and (7) to (8) exclude those who coauthored with more than one deceased scientist in order to determine whether the results are driven by those who experience multiple losses. In Table 7 and Table 8, columns (1) to (2) and (5) to (6) remove the coauthors of the most highly-cited deceased scientists, whose cumulative citations rank above the 90th citation percentile among all deceased scientists. Columns (3) to (4) and (7) to (8) remove the coauthors of the most "popular" deceased scientists, whose total number of coauthors rank above the 90th citation percentile among all deceased scientists. Considering that the most highly-cited scientists or scientists with a great number of coauthors are likely be the most influential and active scientists in their fields, their deaths may generate relatively greater impacts on their coauthors.

The results in Table 5 to 8 show that the main results are robust with respect to each subsample estimation, and that the patterns of changes in different measures of research outcomes are preserved. The robustness checks demonstrate that the main findings of the impact on coauthors' productivity are neither simply driven by the loss of contributions from the deceased scientists, nor changes in the productivity of those researchers who experience multiple losses of prominent coauthors, nor by changes in the productivity of coauthors who lose the most influential researchers, the ones who have the highest number of cumulative citations or the ones who have the largest number of coauthors.

6.3.2 Leave-out Index of Cognitive Distance

Furthermore, since the construction of the cognitive distance index takes into account the publication distributions of researchers, a concern may arise that researchers who have published many papers together may have a low value mechanically even if they are from very different research backgrounds. To tackle this potential threat, we re-construct a "leave-out" index of cognitive distance, which removes the common publications of a researcher and his or her deceased coauthor when calculating the cognitive distance index with the method proposed in Section 2 for each coauthorship dyad, and re-estimate the main effect using the leave-out index of cognitive distance. The results are presented in Table 9 and Table 10[14]. As seen in Tables 9 and 10, estimations using the leave-out index of cognitive distance generate quite consistent results with the main effects estimated using the original index of cognitive distance in Tables 3 and 4, and the patterns of changes in both the publications rates and the citation impacts are preserved.

6.4 Heterogeneous Effects and Underlying Mechanisms

To determine whether there is any heterogeneous effect owing to the characteristics of coauthorship dyads, and to explore further the underlying mechanisms that give rise to the findings above, we add interactions between each group indicator and a variety of dummies that capture different characteristics of coauthorship dyads, respectively, to the model in Eq.(14). Table 11 presents results from the inclusion of interactions with dummies for recent coauthoring, regular coauthoring, working in the same institution, and cases in which the coauthor of a deceased scientist is him- or herself an elite researcher, for the outcomes of publication rates. Table 12 shows the corresponding outcomes of citation impacts. The findings illustrated in the two tables are consistent with each other.

First, among researchers who are recent coauthors, i.e., have published with the deceased scientist within three years before his or her death, those cognitively close coauthors experience a more severe decline in some measures of research productivity, but no significant impacts for coauthors with middle or distant cognitive distance. This finding is consistent with the story of knowledge spillovers, since recent coauthors with close cognitive distance are more likely to share common research ideas with the deceased scientists by the time of the latter's unexpected deaths. Besides, since recent coauthors are likely to engage in ongoing projects with the deceased scientists regardless of the cognitive distance between them, this result also implies that the potentially negative impact on ongoing projects is not an only or major factor that accounts for the declines

[14]Estimations using the leave-out index slightly decreases the original number of observations for analysis, because there is a few researchers in the sample who only have publications with the deceased coauthor before the death of the latter and thus no leave-out index can be calculated for them.

in productivity revealed in the main effects analysis.

Second, we determine whether the death of a regular coauthor has a significant impact. We define a regular coauthor as one who has published with the deceased scientist more than five times before the latter passed away. The results show that losing a regular coauthor does not cause especially negative impacts on researchers' productivity. The finding that researchers do not particularly suffer from losing a regular coauthor excludes the possibility that emotional factors play a significant role in the productivity declines of researchers whose coauthors pass away, since it is natural to assume that regular coauthors are also good friends.

Physical proximity is also an important issue with regard to peer effects and research collaboration, as many previous studies have pointed out, so we investigated whether the results would be different for researchers who had worked in the same institution at the time of the deceased scientist's death. The results show that, under these circumstances, coauthors in the close distance and middle distance groups experience no decline and increase, respectively, in productivity than others in the same distance group. The findings may be explicable in terms of a decrease in the competition effect. As shown in the theoretical framework, capital resources are constrained in the short run. Competition for internal resources (e.g., laboratory space, apparatus, and staff) is more likely to occur among researchers who are doing similar research. Hence, for coauthors in the close distance group who work in the same institute as the deceased scientist, it is possible that a decreased competition for internal research resources balances out the negative impact after the loss of a prominent coauthor. This finding is also consistent with results in Azoulay et al. (2010). However, cognitively distant coauthors in the same institution experience a significant decrease in productivity. This finding supports our hypothesis that researchers with large cognitive distance suffer from the death of a prominent collaborator because of imperfect skill substitution, since substitutions of skills are likely to be geographically circumscribed.

Moreover, the magnitude of impact could differ from the perspective of the abilities of the coauthors of the deceased scientist. We define an "elite" coauthor as a researcher whose cumulative number of citations ranks above the 90th percentile by the year of his or her coauthor's death. The results show that, for elite scientists so defined, the loss a prominent coauthor does not particularly affect productivity, which are somewhat to be expected. For since knowledge spillovers are likely to flow from the extraordinary to the mediocre within a certain scope, there may be little effect on elite coauthors even when they have a close cognitive distance.

Furthermore, as shown in Table A.3 and Table A.4, we also find that there is no effect on the productivity of coauthors who are more senior (have a greater career age) than the deceased scientist when the latter passed away, which is reasonable since peer effects are more common to flow from the senior and seasoned researchers to the junior ones than the other way around. Moreover, as in the life sciences, the position of a coauthor on the authorship roster of an article usually indicates his or her role in a project. Typically, the author who appears in the first position is the main executor and author of the project and the paper, while the one whose name appears in the last position is the director, supervisor, or laboratory owner. It is therefore possible to determine, for each coauthorship dyad, whether the deceased scientist is the first or last author of every paper he or she has published with this coauthor. The results show that losing a coauthor who has always been the first or the last author does not particularly affect a researcher's productivity.

7 Conclusion

This paper empirically delves into the black box of knowledge production. In order to understand the channels of peer effects among knowledge producers, we focus on the effect of cognitive diversity on researchers in this study among other factors that may affect research productivity. More specifically, we investigate whether cognitive diversity between coauthors in life sciences affects the peer effects and productivity, and the possible mechanisms behind this phenomenon. In order to measure the level of cognitive diversity between coauthors precisely, we introduce and calculate a novel index, namely the *cognitive distance* between two researchers, based on their publication distributions and citation relations across scientific journals in which they have publications.

The long tradition of collaboration in the life sciences has provided this study with abundant empirical evidence. In order to confront the endogeneity and selection issues in scientific collaborations, we made use of the unexpected and premature death of active and prominent researchers as a quasi-experimental variation, and have identified 63 active and prominent life scientists who passed away unexpectedly and prematurely between 1995 and 2009. We estimate the changes in productivity of the coauthors of these deceased scientists after their sudden deaths, and determine whether the effect of the adverse shock on their coauthors' productivity differs in the intensity level of cognitive diversity within each coauthorship dyad. The informative individual-level panel data from the WoS database, which contains academic papers published from 1980 to 2013, facilitated the empirical analysis. The productivity of researchers is measured both by their publication rates and citation impacts.

Our theoretical framework suggests that knowledge spillovers and absorptive effects are likely to be stronger between researchers with smaller cognitive distance, while coauthors with larger cognitive distance may benefit from skill complementarity. We find in the empirical analysis that, following the death of an active and prominent life scientist, coauthors with close cognitive distance from the deceased scientist are more likely to experience a lasting decrease in productivity. While the cognitive distance between a coauthor and a deceased scientist generally attenuates the negative shock, the relationship appears to be non-linear for different measures of productivity. For measures of the quantity of research, the impact of cognitive distance on the magnitude of spillovers seems to approximate an inverted-U-shaped relationship. That is, the research output of coauthors at the close or distant level of cognitive distance statistically significant decreases after the loss of an active collaborator, compared with coauthors of middle cognitive distance in relation to the deceased scientist. However, for measures that place more stress on the quality of research, only cognitively close coauthors experience a severe decrease in research productivity.

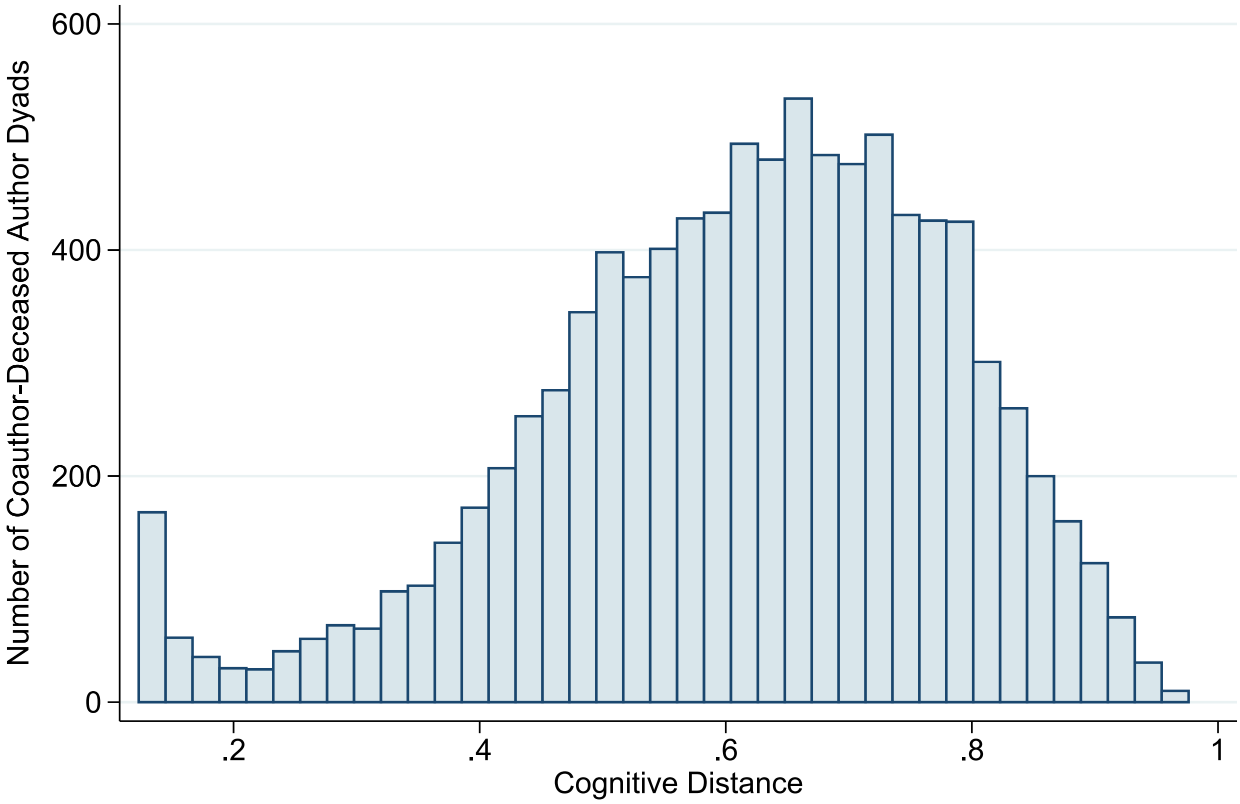
Consistent with the theoretical prediction, the empirical results indicate that both knowledge spillovers and skill complementarity affect productivity in collaborations. The findings suggest that, after losing a preeminent collaborator, the loss of an irreplaceable source of ideas causes a more adverse impact on a survivor's productivity than the potentially imperfect skill substitution. The former may have impacts on both quantity and quality of the research productivity, while the latter mainly affects the research quantity. A further exploration of the mechanism confirms the implications of the main results. Moreover, we find evidence of the competition effect for researchers who are close both in research and physical space. It must be observed that, owing to the limited data, we do not take into account other potentially significant factors, such as the sources of research funding, that may also affect the productivity of researchers.

The findings presented in this paper are in line with conclusions drawn from previous research on peer effects and knowledge production (e.g. Azoulay et al., 2010; Borjas and Doran, 2015b), though unlike those studies the focus here is on the impact of cognitive diversity between collaborators on the process of knowledge production. This study contributes to the existing literature by providing empirical evidence of the causal relation between cognitive diversity and peer effects in scientific production. Moreover, we have created a procedure designed to measure the cognitive distance between two scientists that is not discipline-specific and can thus be applied to measuring this parameter in other sciences.

For although we have in this study explored the evidence from the life sciences and have concentrated on the productivity of academic research, the implications of the findings are applicable to other knowledge-intensive environments, and thus to the creation of insightful policy. For instance, based on the findings presented here, it might be effective, in the effort to increase productivity and to promote innovation in technology-intensive industries, to form teams whose members share a fair measure of cognitive diversity. The causal effect of cognitive diversity with respect to productivity could also be useful for industry, or for the process of transition from research outcome to industrial application.

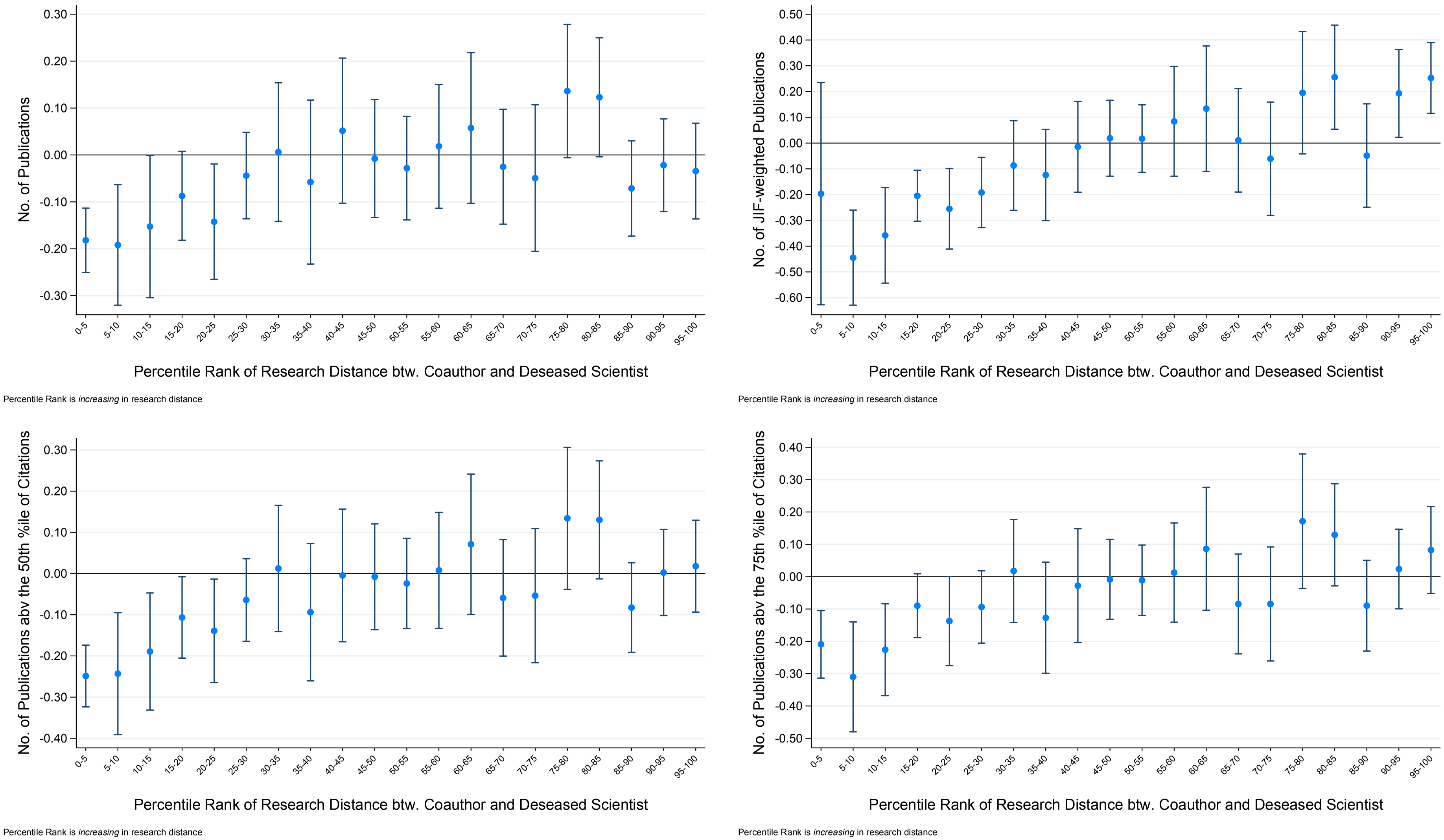
Nevertheless, this study illuminates only a tiny spot within the black box of knowledge creation. Further research on the mechanisms of knowledge production and innovation can be expected to generate further interesting and fruitful insights.

Figure 1: Dyad Cognitive Distance between a Coauthor and a Deceased Scientist



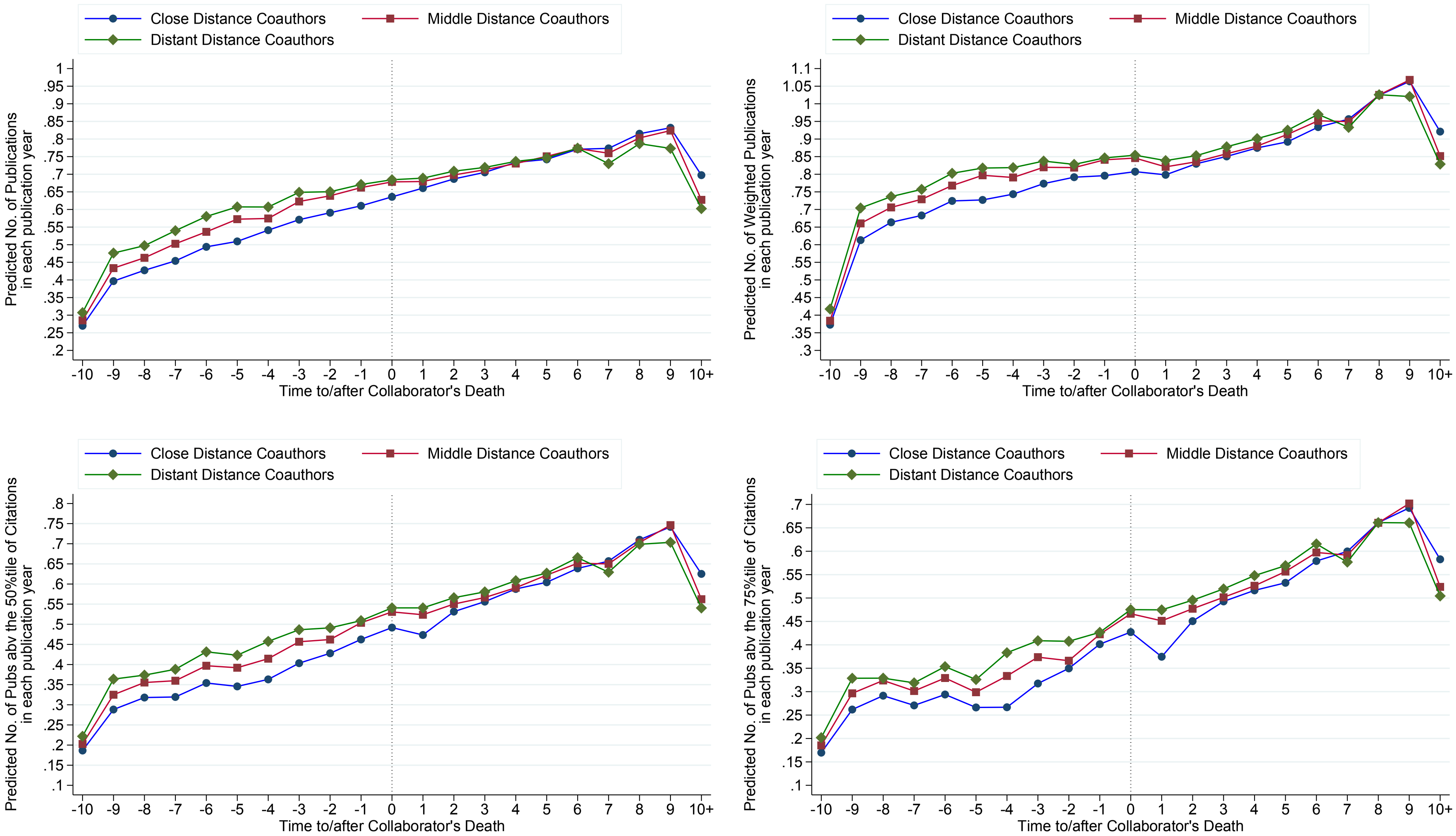
Note: Figure 1 presents a histogram of calculated cognitive distance between each coauthor-deceased scientist dyad in the sample.

Figure 2: Impact of Unexpected Death of Scientists on the Research Output of Coauthors



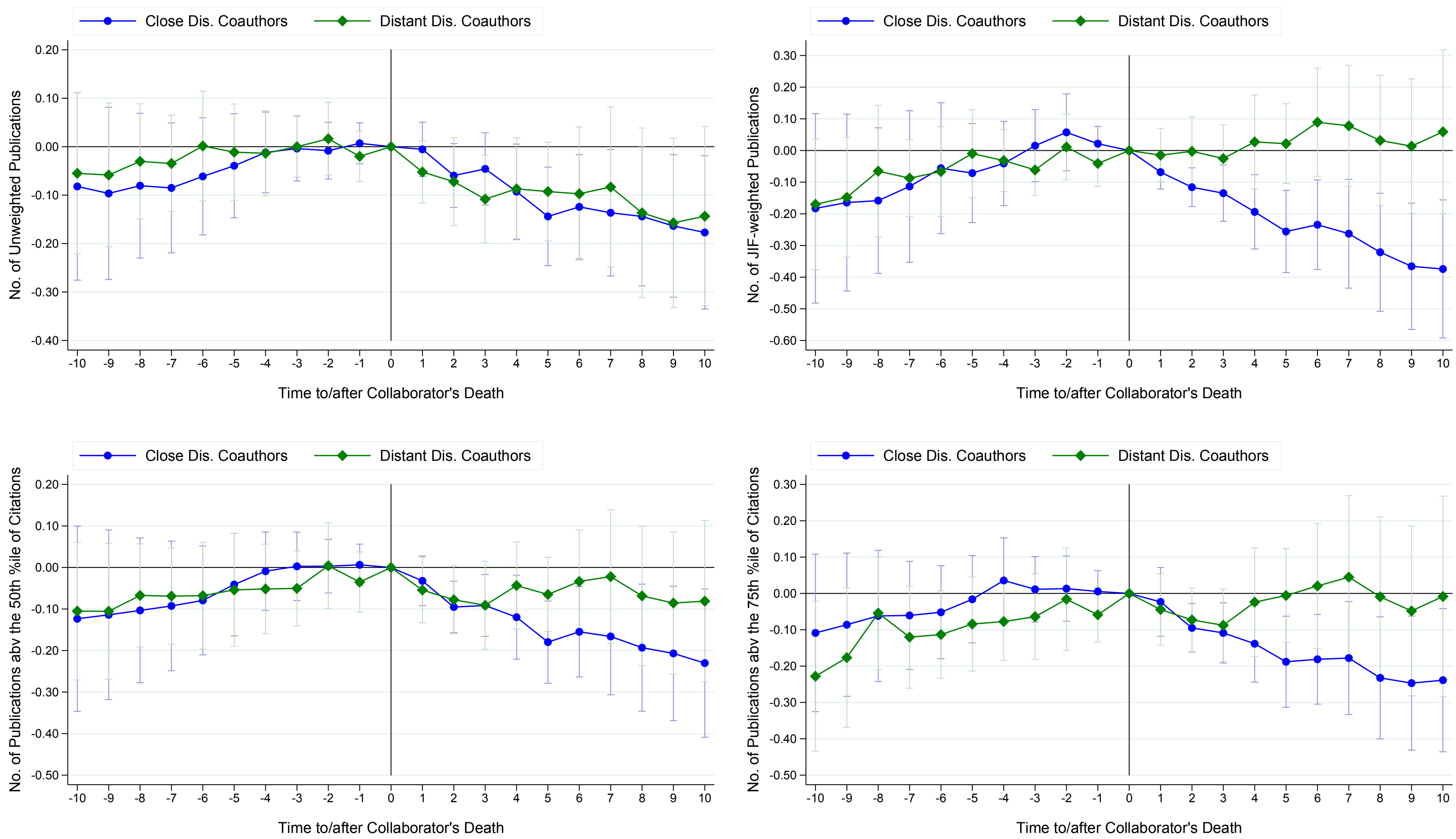
Note: Figure 2 plots the research output of coauthors after the unexpected death of a prominent scientist against the percentile rankings of cognitive distance between coauthors and deceased scientists in the sample. The blue dots correspond to coefficient estimates from conditional fixed effect QML Poisson model with controls for year fixed effects, career age fixed effects, and interaction terms between indicators of different percentile ranks of the cognitive distance between a coauthor and his/her deceased collaborator and a dummy of "death" of the collaborator. The 95% confidence interval (corresponding to robust standard errors, clustered at the level of the deceased scientist) are illustrated by the dark-blue vertical bars and their caps.

Figure 3: Predicted Means Trend of the Research Output of Coauthors with Different Cognitive Distance from Deceased Scientists



Note: Figure 3 presents event study graphs showing the predicted means trend of research output of coauthors in each year before and after the unexpected death of his/her collaborator. The predicted means trend of each measure of research output is estimated from conditional fixed effect QML Poisson model controlling for year fixed effects and career age indicators. The dotted vertical line corresponds to the death year of the deceased scientist.

Figure 4: Dynamics of the Impact of Unexpected Death of Scientists on the Research Output of Coauthors



Note: Figure 4 presents dynamics of the impact of unexpected death of scientists on the research output of their coauthors of close cognitive distance and distant cognitive distance in each year before and after the death year, relative to coauthors of middle cognitive distance. The horizontal axis measures the number of years since the unexpected death of the scientist. The plots connected by each solid line correspond to coefficient estimates of conditional fixed effects quasi-maximum likelihood Poisson regressions in which each dependent variable of a coauthor's research output is regressed onto year fixed effect, career age effects, and interaction terms between 20 indicators corresponding to a particular year relative to an unexpected death year and each indicator of a coauthor being in close distance group (blue line) and distant distance group (green line), respectively. Coauthors of cognitive distance from the deceased scientist at the middle level are treated as the reference group. The interaction terms of coauthor group indicators and the year of death are omitted. The 95% confidence interval (corresponding to robust standard errors, clustered at the level of the deceased scientist) are illustrated by the vertical bars and their caps (light-blue for close distance group and light-green for distant distance group).

Table 1: Summary Statistics for Deceased Scientists

	Mean	Median	Std. Dev.	Min	Max	Obs.
Year of death	2000	2000	2.842	1995	2008	63
Year of first publication	1981	1980	2.845	1971	1990	63
Year of last publication	2004	2004	4.058	1995	2013	63
Career age	19	19	3.634	10	31	63
Cum. no. of publications	122	99	91.957	5	570	63
No. of cum. citations at the time of death	5497	3759	5490.025	59	25692	63
No. of citations by career age	280	197	269.238	5	1352	63
No. of coauthors	152	117	125.336	5	507	63

Notes: The sample of deceased scientists contains 63 life scientists who passed away unexpectedly and prematurely between 1995 and 2009. First and last year of publication refer to the publication years of a scientist's first and last academic paper. Career age is defined as the year difference between a scientist's year of death and the year of his/her first publication. Cumulative number of publications counts the total number of publications of academic papers of a deceased scientist published since 1980. Number of cumulative citations is the total number of citations of the papers of the deceased scientist by the year of death. Number of citations by career age is the number of cumulative citations of a deceased scientist at the time of death divided by his/her career age. Number of coauthors is defined as the total number of researchers with whom a deceased scientist has coauthored in his/her published academic papers.

Table 2: Summary Statistics for Coauthors in the Year of Deceased Scientists' Death

	All Coauthors		Close Distance Coauthors		Middle Distance Coauthors		Distant Distance Coauthors	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Cognitve distance	0.61	0.17	0.48	0.13	0.71	0.04	0.84	0.04
No. of publications in each publication year	3.20	3.20	2.97	3.07	3.45	3.35	3.42	3.22
Cum. no. of publications	52.85	118.77	44.86	111.27	60.99	128.57	61.03	118.35
No. of JIF-weighted publications in each publication year	10.78	15.12	11.78	16.63	10.39	14.14	8.35	11.25
Cum. no. of JIF-weighted publications	177.95	536.12	177.25	559.68	189.16	560.11	155.63	380.92
No. of publications above 50th citation %ile in each pub year	2.27	2.48	2.18	2.41	2.44	2.63	2.18	2.36
Cum. no. of publications above 50th citation %ile	37.76	90.34	32.97	85.14	43.79	99.42	40.10	84.86
No. of publications above 75th citation %ile in each pub year	1.47	1.78	1.46	1.76	1.56	1.88	1.31	1.63
Cum. no. of publications above 75th citation %ile	24.42	62.38	21.89	58.86	28.23	69.35	24.28	56.75
Career age	13.61	6.66	12.57	6.61	14.31	6.41	15.46	6.76
Senior coauthor	0.08	0.27	0.07	0.26	0.08	0.28	0.10	0.30
No. of cumulative citations	1915.26	3989.63	1778.55	4102.89	2100.93	3997.88	1953.43	3558.94
No. of citations by career age	116.35	214.63	111.55	228.36	125.56	205.88	111.73	184.92
Elite coauthor	0.10	0.30	0.09	0.29	0.11	0.32	0.10	0.30
No. of collaborations	3.18	5.91	3.45	6.21	3.30	6.49	2.06	2.49
Regular coauthor	0.13	0.34	0.15	0.35	0.13	0.34	0.06	0.24
Recent coauthor	0.32	0.47	0.35	0.48	0.31	0.46	0.28	0.45
Same institute	0.11	0.31	0.10	0.30	0.11	0.31	0.12	0.33
Observations	9605		4851		3266		1488	

Notes: The sample of coauthors contains scientists who have coauthored with the 63 deceased life scientists in academic papers published during 1980 and 2013. All variables are measured at the time of deceased scientists' death. Citation percentiles are calculated based on total citations of articles published in each given year in life sciences. Career age is defined as the year difference between an author's year of first publication and his/her deceased coauthor's year of death. Senior coauthors are those whose career age larger than the deceased scientist when the latter passes away. Number of cumulative citations is the total number of citations of the papers of the scientist by the year of his/her deceased coauthor's death. Number of citations by career age is the number of cumulative citations of a scientist divided by his/her career age. Elite coauthors are defined as those whose number of cumulative citations ranks above the top 10 percentile of that of the full sample of coauthors. Number of collaborations is the number of coauthored articles of a coauthor- deceased scientist pair published during 1980 and 2013. Regular coauthors are defined as those who have more than five times of collaborations with her deceased coauthor. Recent coauthors are defined as those who have at least one collaboration published within the three years before his/her coauthor's death. Same institution indicates that a coauthor and a decease scientist have worked in the same institution.

Table 3: Impact of Unexpected Death of Scientists on Coauthors' Publication Rates

Dependent Variable:	No. of Publications							
	Unweighted				JIF-weighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.096	−0.549	−0.002	0.033	−0.480	−0.648	0.036	0.048
	(0.115)	(0.268)**	(0.027)	(0.027)	(0.178)***	(0.566)	(0.046)	(0.053)
Distance × Death	0.113	1.697			0.680	1.275		
	(0.160)	(0.775)**			(0.260)***	(1.756)		
Distance^2 × Death		−1.308				−0.497		
		(0.568)**				(1.332)		
Close Distance Coauthor × Death			−0.050	−0.085			−0.185	−0.196
			(0.041)	(0.040)**			(0.066)***	(0.070)***
Distant Distance Coauthor × Death				−0.116				−0.034
				(0.036)***				(0.061)
Log pseudo-likelihood	-335749.9	-335649.6	-335808.9	-335740.7	-865867.6	-865827.2	-867873.4	-867660.9
No. of Observations	157336	157336	157336	157336	157336	157336	157336	157336

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the raw number of publications authored by a coauthor (of a deceased scientist) in the year of observation. The dependent variable of models in column (5) to (8) is the number of JIF-weighted publications authored by a coauthor (of a deceased scientist) in the year of observation. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (3) and (7), coauthors whose cognitive distance from the deceased scientist ranks above the 50th percentile rank of the coauthors sample (i.e. both the middle distance and distant distance groups) are the reference group. In column (4) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 4: Impact of Unexpected Death of Scientists on Coauthors' Citation Distribution

Dependent Variable:	Distribution of Citations							
	No. of Pubs abv 50th Percentile of Citations				No. of Pubs abv 75th Percentile of Citations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.227	−0.657	0.012	0.035	−0.300	−0.638	0.018	0.032
	(0.118)*	(0.270)**	(0.028)	(0.030)	(0.126)**	(0.297)**	(0.029)	(0.033)
Distance × Death	0.302	1.822			0.410	1.611		
	(0.167)*	(0.773)**			(0.173)**	(0.853)*		
Distance^2 × Death		−1.264				−1.005		
		(0.563)**				(0.625)		
Close Distance Coauthor × Death			−0.103	−0.125			−0.125	−0.138
			(0.039)***	(0.040)***			(0.042)***	(0.044)***
Distant Distance Coauthor × Death				−0.082				−0.049
				(0.046)*				(0.049)
Log pseudo-likelihood	-298282.3	-298217.3	-298350.5	-298315.3	-245431.5	-245406	-245499.8	-245484
No. of Observations	156918	156918	156918	156918	155156	155156	155156	155156

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 50th total citation percentile of papers published that year. The dependent variable of models in column (5) to (8) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 75th total citation percentile of papers published that year. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (3) and (7), coauthors whose cognitive distance from the deceased scientist ranks above the 50th percentile rank of the coauthors sample (i.e. both the middle distance and distant distance groups) are the reference group. In column (4) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 5: Robustness Checks (Subsamples I) – Coauthors' Publication Rates

	No. of Publications							
Dependent Variable:	Unweighted				JIF-weighted			
Sample:	Pubs with others		No multiple-death		Pubs with others		No multiple-death	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.041	0.075	−0.032	0.026	−0.383	0.094	−0.378	0.046
	(0.110)	(0.030)**	(0.093)	(0.030)	(0.165)**	(0.056)*	(0.150)**	(0.056)
Distance × Death	0.082		0.033		0.603		0.549	
	(0.154)		(0.127)		(0.245)**		(0.221)**	
Close Distance Coauthor × Death		−0.088		−0.044		−0.181		−0.158
		(0.043)**		(0.038)		(0.070)***		(0.066)**
Distant Distance Coauthor × Death		−0.143		−0.104		−0.058		−0.032
		(0.035)***		(0.034)***		(0.058)		(0.047)
Log pseudo-likelihood	-319092.8	-319022.2	-291686.3	-291685.7	-818968.7	-820650.1	-720450.8	-721342.5
No. of Observations	147922	147922	145576	145576	147922	147922	145576	145576

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the raw number of publications authored by a coauthor (of a deceased scientist) in the year of observation. The dependent variable of models in column (5) to (8) is the number of JIF-weighted publications authored by a coauthor (of a deceased scientist) in the year of observation. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (2), (4), (6) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. Column (1), (2), (5) and (6) contain the subsample that includes only the publications of coauthors that published with other researchers before and after the death of the deceased scientists. Column (3), (4), (7) and (8) contain the subsample that removes the coauthors who collaborate with more than one scientist who dies. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 6: Robustness Checks (Subsamples I) – Coauthors' Citation Distribution

	Distribution of Citations							
Dependent Variable:	No. of Pubs abv 50th Percentile of Citations				No. of Pubs abv 85th Percentile of Citations			
Sample:	Pubs with others		No multiple-death		Pubs with others		No multiple-death	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.161	0.085	−0.171	0.030	−0.221	0.092	−0.247	0.020
	(0.111)	(0.036)**	(0.095)*	(0.035)	(0.121)*	(0.042)**	(0.102)**	(0.040)
Distance × Death	0.271		0.235		0.374		0.334	
	(0.163)*		(0.135)*		(0.177)**		(0.141)**	
Close Distance Coauthor × Death		−0.122		−0.091		−0.134		−0.106
		(0.044)***		(0.038)**		(0.049)***		(0.042)**
Distant Distance Coauthor × Death		−0.109		−0.061		−0.084		−0.037
		(0.046)**		(0.046)		(0.048)*		(0.054)
Log pseudo-likelihood	-282596.2	-282585.3	-257722.7	-257733.6	-231381.8	-231405.8	-210197	-210203.1
No. of Observations	147505	147505	145158	145158	145642	145642	143424	143424

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 50th total citation percentile of papers published that year. The dependent variable of models in column (5) to (8) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 75th total citation percentile of papers published that year. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (2), (4), (6) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. Column (1), (2), (5) and (6) contain the subsample that includes only the publications of coauthors that published with other researchers before and after the death of the deceased scientists. Column (3), (4), (7) and (8) contain the subsample that removes the coauthors who collaborate with more than one scientist who dies. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 7: Robustness Checks (Subsamples II) – Coauthors' Publication Rates

	No. of Publications							
Dependent Variable:	Unweighted				JIF-weighted			
Sample:	No highly-cited elites		No popular stars		No highly-cited elites		No popular stars	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.117	0.021	−0.145	0.037	−0.509	0.041	−0.436	0.052
	(0.140)	(0.030)	(0.100)	(0.032)	(0.211)**	(0.056)	(0.128)***	(0.065)
Distance × Death	0.127		0.199		0.716		0.649	
	(0.194)		(0.133)		(0.304)**		(0.193)***	
Close Distance Coauthor × Death		−0.086		−0.096		−0.187		−0.171
		(0.044)**		(0.044)**		(0.071)***		(0.075)**
Distant Distance Coauthor × Death		−0.110		−0.089		−0.020		0.000
		(0.040)***		(0.041)**		(0.072)		(0.074)
Log pseudo-likelihood	-276186.3	-276187.7	-236942.5	-236642.7	-711791.2	-713661.2	-622349.1	-622787.2
No. of Observations	128586	128586	110291	110291	128586	128586	110291	110291

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the raw number of publications authored by a coauthor (of a deceased scientist) in the year of observation. The dependent variable of models in column (5) to (8) is the number of JIF-weighted publications authored by a coauthor (of a deceased scientist) in the year of observation. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (2), (4), (6) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. Column (1), (2), (5) and (6) exclude coauthors of most elite scientists (7 out of 63) whose number of cumulative citations above the top 10th percentile. Column (3), (4), (7) and (8) exclude coauthors of most popular scientists (7 out of 63) whose number of coauthors above the top 10th percentile. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 8: Robustness Checks (Subsamples II) – Coauthors' Citation Distribution

	Distribution of Citations							
Dependent Variable:	No. of Pubs abv 50th Percentile of Citations				No. of Pubs abv 85th Percentile of Citations			
Sample:	No highly-cited elites		No popular stars		No highly-cited elites		No popular stars	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.249	0.023	−0.264	0.031	−0.329	0.025	−0.325	0.017
	(0.146)*	(0.032)	(0.084)***	(0.035)	(0.155)**	(0.033)	(0.095)***	(0.039)
Distance × Death	0.322		0.372		0.448		0.466	
	(0.204)		(0.116)***		(0.209)**		(0.134)***	
Close Distance Coauthor × Death		−0.122		−0.119		−0.135		−0.111
		(0.043)***		(0.045)***		(0.047)***		(0.049)**
Distant Distance Coauthor × Death		−0.072		−0.043		−0.030		0.014
		(0.051)		(0.049)		(0.054)		(0.051)
Log pseudo-likelihood	-244436.9	-244481	-210074.3	-209885.9	-200212.4	-200272.1	-172541.9	-172454
No. of Observations	128197	128197	109947	109947	126696	126696	108524	108524

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 50th total citation percentile of papers published that year. The dependent variable of models in column (5) to (8) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 75th total citation percentile of papers published that year. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (2), (4), (6) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. Column (1), (2), (5) and (6) exclude coauthors of most elite scientists (7 out of 63) whose number of cumulative citations above the top 10th percentile. Column (3), (4), (7) and (8) exclude coauthors of most popular scientists (7 out of 63) whose number of coauthors above the top 10th percentile. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 9: Robustness Checks (Leave-out Index) – Coauthors' Publication Rates

Dependent Variable:	No. of Publications							
	Unweighted				JIF-weighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.122	−0.436	0.000	0.036	−0.476	−0.345	0.032	0.041
	(0.115)	(0.252)*	(0.026)	(0.026)	(0.181)***	(0.524)	(0.038)	(0.040)
Distance × Death	0.152	1.230			0.669	0.217		
	(0.159)	(0.807)			(0.261)**	(1.596)		
Distance^2 × Death		−0.879				0.373		
		(0.643)				(1.190)		
Close Distance Coauthor × Death			−0.056	−0.091			−0.174	−0.183
			(0.039)	(0.037)**			(0.058)***	(0.058)***
Distant Distance Coauthor × Death				−0.127				−0.031
				(0.043)***				(0.054)
Log pseudo-likelihood	-333342.1	-333302.9	-333271.5	-333211.3	-861489	-861469.1	-862144.2	-861720.8
No. of Observations	155294	155294	155294	155294	155294	155294	155294	155294

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$- deceased scientist$_j$ -year$_t$ level. The leave-out index of cognitive distance excludes common publications between a coauthor and a deceased scientists when calculating their cognitive distance. The dependent variable of models in column (1) to (4) is the raw number of publications authored by a coauthor (of a deceased scientist) in the year of observation. The dependent variable of models in column (5) to (8) is the number of JIF-weighted publications authored by a coauthor (of a deceased scientist) in the year of observation. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (3) and (7), coauthors whose cognitive distance from the deceased scientist ranks above the 50th percentile rank of the coauthors sample (i.e. both the middle distance and distant distance groups) are the reference group. In column (4) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 10: Robustness Checks (Leave-out Index) – Coauthors' Citation Distribution

Dependent Variable:	Distribution of Citations							
	No. of Pubs abv 50th Percentile of Citations				No. of Pubs abv 75th Percentile of Citations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.238	−0.420	0.011	0.036	−0.287	−0.315	0.014	0.025
	(0.116)**	(0.267)	(0.026)	(0.027)	(0.123)**	(0.306)	(0.026)	(0.025)
Distance × Death	0.319	0.951			0.390	0.489		
	(0.163)*	(0.893)			(0.166)**	(1.000)		
Distance^2 × Death		−0.520				−0.082		
		(0.730)				(0.804)		
Close × Death			−0.099	−0.122			−0.112	−0.122
			(0.038)***	(0.037)***			(0.041)***	(0.039)***
Distant × Death				−0.093				−0.044
				(0.048)*				(0.045)
Log pseudo-likelihood	-296243.1	-296233.5	-296203	-296170.9	-243747.3	-243747.1	-243726.2	-243709.2
No. of Observations	154895	154895	154895	154895	153207	153207	153207	153207

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$- deceased scientist$_j$ -year$_t$ level. The leave-out index of cognitive distance excludes common publications between a coauthor and a deceased scientists when calculating their cognitive distance. The dependent variable of models in column (1) to (4) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that lie above the 50th lifetime citation percentile of papers published that year. The dependent variable of models in column (5) to (8) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that lie above the 75th lifetime citation percentile of papers published that year. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (3) and (7), coauthors whose cognitive distance from the deceased scientist ranks above the 50th percentile rank of the coauthors sample (i.e. both the middle distance and distant distance groups) are the reference group. In column (4) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 11: Heterogeneous Effect and Mechanisms—Coauthor's Publication Rates

Dependent Variable:	No. of Publications							
	Unweighted				JIF-weighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	0.027	0.028	0.020	0.029	0.044	0.045	0.029	0.058
	(0.026)	(0.027)	(0.026)	(0.035)	(0.055)	(0.056)	(0.051)	(0.061)
Close × Death	−0.053	−0.097	−0.087	−0.027	−0.156	−0.207	−0.191	−0.145
	(0.044)	(0.038)***	(0.042)**	(0.044)	(0.072)**	(0.070)***	(0.072)***	(0.071)**
Distant × Death	−0.099	−0.121	−0.087	−0.128	−0.032	−0.036	0.013	−0.026
	(0.036)***	(0.038)***	(0.035)**	(0.043)***	(0.061)	(0.054)	(0.054)	(0.061)
Death × Recent Coauthor	0.023				0.001			
	(0.033)				(0.047)			
Close × Death × Recent Coauthor	−0.095				−0.095			
	(0.054)*				(0.058)			
Distant × Death × Recent Coauthor	−0.082				−0.018			
	(0.087)				(0.106)			
Death × Regular Coauthor		0.026				0.016		
		(0.045)				(0.051)		
Close × Death × Regular Coauthor		0.060				0.071		
		(0.081)				(0.074)		
Distant × Death × Regular Coauthor		0.080				0.066		
		(0.067)				(0.157)		
Death × Same Institute			0.148				0.218	
			(0.056)***				(0.067)***	
Close × Death × Same Institute			0.019				−0.073	
			(0.088)				(0.112)	
Distant × Death × Same Institute			−0.286				−0.454	
			(0.155)*				(0.212)**	
Death × Elite Coauthor				−0.014				−0.063
				(0.052)				(0.062)
Close × Death × Elite Coauthor				−0.098				−0.043
				(0.078)				(0.077)
Distant × Death × Elite Coauthor				0.099				0.027
				(0.114)				(0.133)
Log pseudo-likelihood	-335643.6	-335461.8	-335335.2	-334512.4	-867068.5	-865899.1	-866146.6	-861009.7
No. of Observations	157336	157336	157336	157336	157336	157336	157336	157336

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the raw number of publications authored by a coauthor (of a deceased scientist) in the year of observation. The dependent variable of models in column (5) to (8) is the number of JIF-weighted publications authored by a coauthor (of a deceased scientist) in the year of observation. "Recent coauthor" is defined as a coauthor who has at least one collaboration published within the three years before the death of his/her coauthor. "Regular coauthor" is a coauthor who has more than five times of collaboration with the deceased scientist. "Same institute" indicates that a coauthor and the deceased scientist have worked in the same institution. "Elite coauthor" refers to the coauthors whose number of cumulative citations are above the top 10th percentile of that of the coauthors sample by the year of his/her collaborator's death. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. Coauthors in the middle level of cognitive distance from the deceased scientist are the reference group. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table 12: Heterogeneous Effect and Mechanisms—Coauthor's Citation Distribution

	Distribution of Citations							
Dependent Variable:	No. of Pubs abv 50th Percentile of Citations				No. of Pubs abv 75th Percentile of Citations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
main								
Death	0.025	0.040	0.022	0.044	0.023	0.037	0.017	0.031
	(0.031)	(0.033)	(0.029)	(0.040)	(0.033)	(0.039)	(0.032)	(0.045)
Close × Death	−0.085	−0.149	−0.132	−0.084	−0.090	−0.163	−0.150	−0.097
	(0.043)**	(0.040)***	(0.042)***	(0.045)*	(0.046)**	(0.049)***	(0.045)***	(0.047)**
Distant × Death	−0.046	−0.093	−0.057	−0.083	−0.009	−0.065	−0.022	−0.065
	(0.048)	(0.049)*	(0.042)	(0.059)	(0.047)	(0.052)	(0.048)	(0.067)
Death × Recent Coauthor	0.033				0.027			
	(0.043)				(0.049)			
Close × Death × Recent Coauthor	−0.113				−0.132			
	(0.058)*				(0.064)**			
Distant × Death × Recent Coauthor	−0.145				−0.152			
	(0.099)				(0.124)			
Death × Regular Coauthor		−0.011				−0.012		
		(0.045)				(0.047)		
Close × Death × Regular Coauthor		0.109				0.114		
		(0.075)				(0.080)		
Distant × Death × Regular Coauthor		0.082				0.070		
		(0.074)				(0.097)		
Death × Same Institute			0.144				0.168	
			(0.067)**				(0.069)**	
Close × Death × Same Institute			0.063				0.103	
			(0.085)				(0.089)	
Distant × Death × Same Institute			−0.240				−0.275	
			(0.192)				(0.175)	
Death × Elite Coauthor				−0.047				−0.021
				(0.054)				(0.059)
Close × Death × Elite Coauthor				−0.041				−0.036
				(0.068)				(0.074)
Distant × Death × Elite Coauthor				0.058				0.080
				(0.118)				(0.123)
Log pseudo-likelihood	-298213	-298064.3	-298004.1	-297215.4	-245407.4	-245290.8	-245235.6	-244662.9
No. of Observations	156918	156918	156918	156918	155156	155156	155156	155156

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 50th total citation percentile for papers published that year. The dependent variable of models in column (5) to (8) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 75th total citation percentile for papers published that year. "Recent coauthor" is defined as a coauthor who has at least one collaboration published within the three years before the death of his/her coauthor. "Regular coauthor" is a coauthor who has more than five times of collaboration with the deceased scientist. "Same institute" indicates that a coauthor and the deceased scientist have worked in the same institution. "Elite coauthor" refers to the coauthors whose number of cumulative citations are above the top 10th percentile of that of the coauthors sample by the year of his/her collaborator's death. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. Coauthors in the middle level of cognitive distance from the deceased scientist are the reference group. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

References

Aizenman, J. and Kletzer, K. (2011). The life cycle of scholars and papers in economics — the 'citation death tax'. *Applied Economics*, 43(27):4135–4148.

Azoulay, P., Fons-Rosen, C., and Zivin, J. S. G. (2015). Does Science Advance One Funeral at a Time? *National Bureau of Economic Research*, Working Paper, No.w21788.

Azoulay, P., Zivin, J. S. G., and Wang, J. (2010). Superstar Extinction. *The Quarterly Journal of Economics*, 125(2):549–589.

Borjas, G. J. and Doran, K. B. (2012). The Collapse of the Soviet Union and the Productivity of American Mathematicians. *The Quarterly Journal of Economics*, 127(3):1143–1203.

Borjas, G. J. and Doran, K. B. (2015a). Cognitive Mobility: Labor Market Responses to Supply Shocks in the Space of Ideas. *Journal of Labor Economics*, 33(S1):S109–S145.

Borjas, G. J. and Doran, K. B. (2015b). Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors. *Review of Economics and Statistics*, 97(5):1104–1117.

Davis, P. M. (2013). Journal Usage Half-Life. http://publishers.org/sites/default/files/uploads/PSP/journalusagehalflife.pdf. [Online; accessed 26-January-2016].

Ductor, L. (2015). Does Co-authorship Lead to Higher Academic Productivity? *Oxford Bulletin of Economics and Statistics*, 77(3):385–407.

Fafchamps, M., Leij, M. J., and Goyal, S. (2010). Matching and Network Effects. *Journal of the European Economic Association*, 8(1):203–231.

Freeman, R. B. and Huang, W. (2015). Collaborating with People Like Me: Ethnic Coauthorship within the United States. *Journal of Labor Economics*, 33(S1 Part 2):S289–S318.

Hausman, J., Hall, B. H., Griliches, Z., et al. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica*, 52(4):909–38.

Jones, B. F. (2009). The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? *The Review of Economic Studies*, 76(1):283–317.

Levin, S. G. and Stephan, P. E. (1991). Research Productivity Over the Life Cycle: Evidence for Academic Scientists. *The American Economic Review*, pages 114–132.

Leydesdorff, L. and Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, Centrality, and Citations. *Journal of Informetrics*, 5(1):87–100.

Lucas, R. E. (1988). On the Mechanics of Economic Development. *Journal of monetary economics*, 22(1):3–42.

Mello, A. L. and Rentsch, J. R. (2015). Cognitive Diversity in Teams A Multidisciplinary Review. *Small Group Research*, 46(6):623–658.

Mohnen, M. (2016). Stars and Brokers: Peer Effects among Medical Scientists. *Working Paper.*, University College London.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., and Van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research policy*, 36(7):1016–1034.

Oettl, A. (2012). Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science*, 58(6):1122–1140.

Page, S. E. (2008). *The Difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P., and Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management. *Research Policy*, 41(7):1262–1282.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, 98(5):S71–S102.

Salton, G. and McGill, M. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill computer science series. McGraw-Hill.

Silva, J. S. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and statistics*, 88(4):641–658.

Waldinger, F. (2010). Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany. *Journal of Political Economy*, 118(4):787–831.

Waldinger, F. (2012). Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. *The Review of Economic Studies*, 79(2):838–861.

Wang, Q. and Sandström, U. (2015). Defining the role of cognitive distance in the peer review process with an explorative study of a grant scheme in infection biology. *Research Evaluation*, pages 271–281.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models. *Journal of Econometrics*, 90(1):77–97.

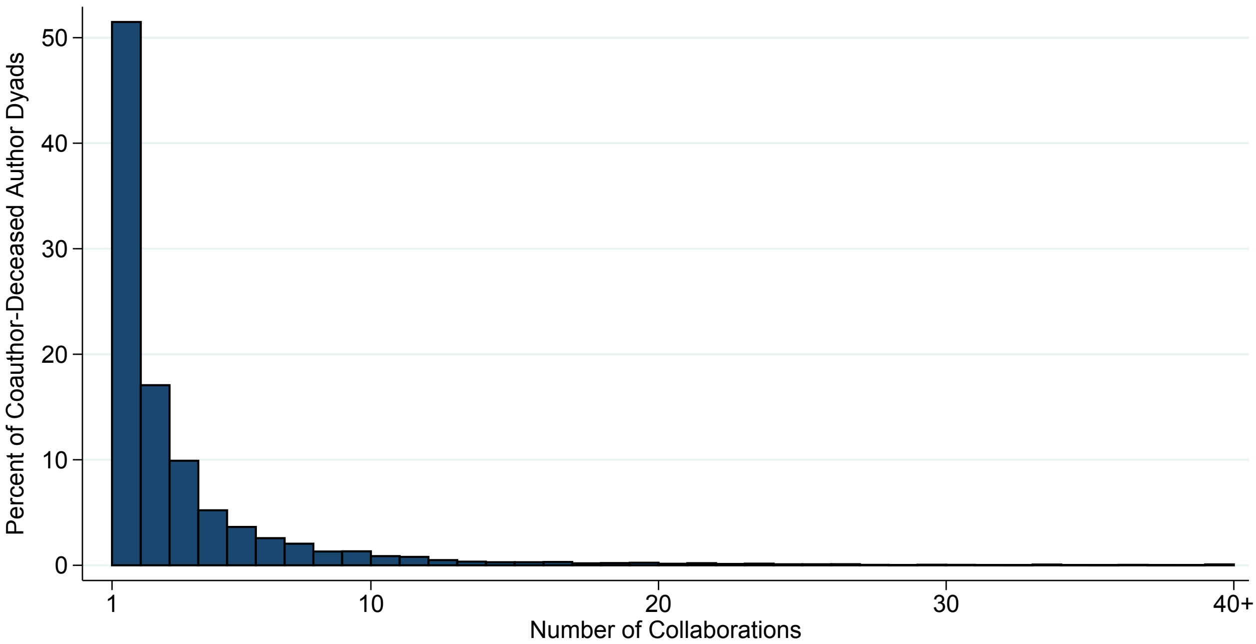
Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, 2 edition.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, 316(5827):1036–1039.

Yegros Yegros, A., D'Este Cukierman, P., Rafols, I., et al. (2015). Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. *PLoS ONE*, 10(8):e0135095.

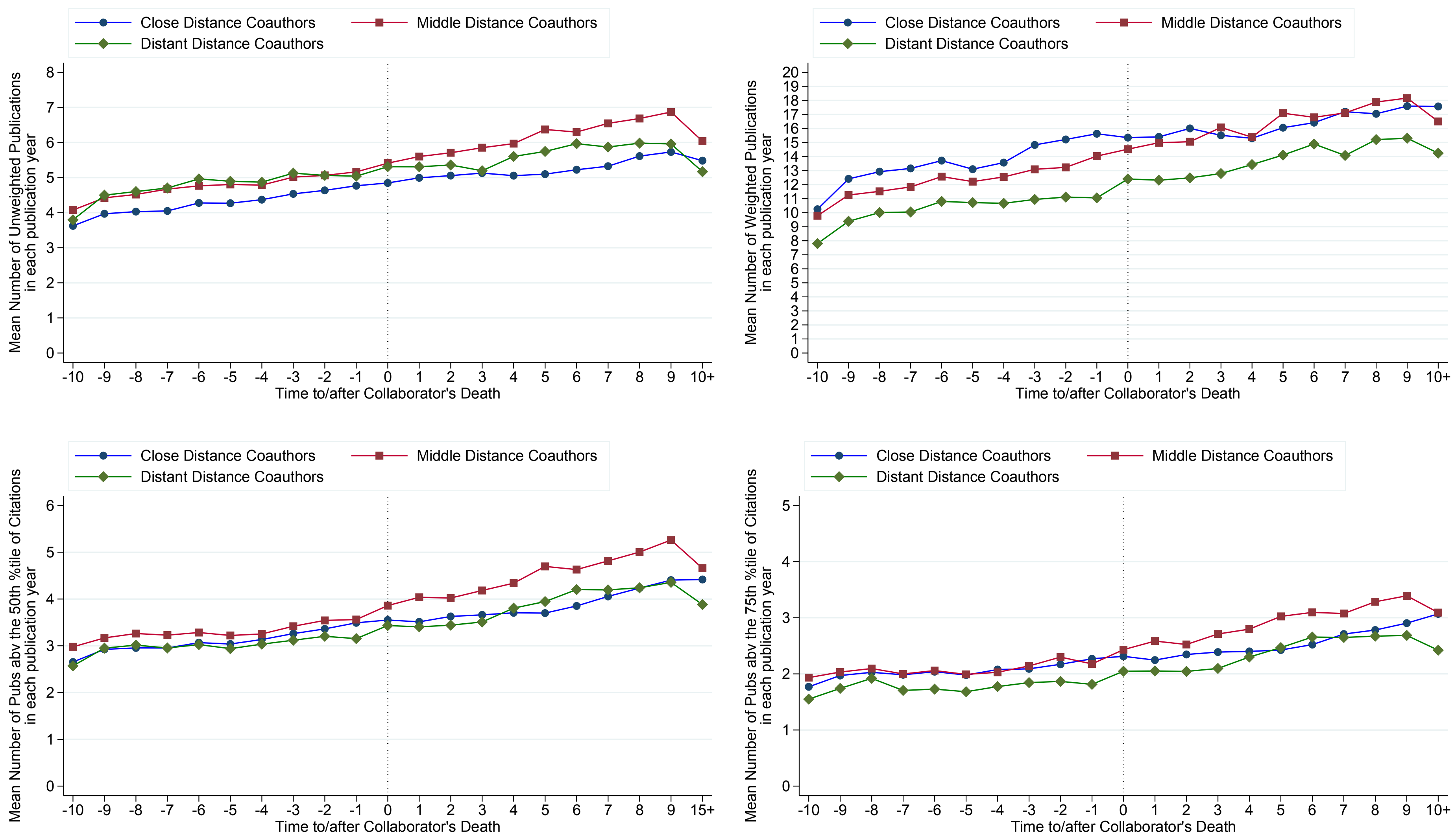
Appendix A

Figure A.1: Distribution of Coauthorship Intensity



Note: Histogram of the number of collaborations for the 9605 coauthor- deceased author dyads in the sample.

Figure A.2: Research Output Trends for Coauthors with Different Cognitive Distance from Deceased Scientists



Note: Figure A.2 presents sample mean by year of the number of papers, the number of JIF-weighted papers, and the number of papers above the 50% and 75% percentile of citations, for coauthors of different groups, respectively, before and after the unexpected death of her collaborator. The dotted vertical line corresponds to the death year of the deceased scientist.

Table A.1: Impact of Unexpected Death of Scientists on Coauthors' Publication Rates (OLS)

	No. of Publications							
Dependent Variable:	Unweighted				JIF-weighted			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.055	−0.392	−0.031	−0.010	−0.290	−0.166	−0.017	−0.014
	(0.087)	(0.184)**	(0.027)	(0.027)	(0.200)	(0.523)	(0.040)	(0.042)
Distance × Death	0.031	1.288			0.348	−0.114		
	(0.129)	(0.495)**			(0.299)	(1.577)		
Distance^2 × Death		−1.085				0.399		
		(0.351)***				(1.158)		
Close Distance Coauthor × Death			−0.016	−0.035			−0.118	−0.119
			(0.033)	(0.031)			(0.054)**	(0.056)**
Distant Distance Coauthor × Death				−0.073				−0.006
				(0.032)**				(0.046)
R^2	0.09	0.09	0.10	0.09	0.10	0.10	0.10	0.10
No. of Observations	157336	157336	157336	157336	157336	157336	157336	157336

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from coauthor-deceased author pairwise fixed effect specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the raw number of publications authored by a coauthor (of a deceased scientist) in the year of observation. The dependent variable of models in column (5) to (8) is the number of JIF-weighted publications authored by a coauthor (of a deceased scientist) in the year of observation. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (3) and (7), coauthors whose cognitive distance from the deceased scientist ranks above the 50th percentile rank of the coauthors sample (i.e. both the middle distance and distant distance groups) are the reference group. In column (4) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table A.2: Impact of Unexpected Death of Scientists on Coauthors' Citation Distribution (OLS)

	Distribution of Citations							
Dependent Variable:	No. of Pubs abv 50th Percentile of Citations				No. of Pubs abv 75th Percentile of Citations			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Death	−0.131	−0.412	−0.008	0.003	−0.122	−0.298	−0.004	0.006
	(0.084)	(0.174)**	(0.025)	(0.026)	(0.066)*	(0.154)*	(0.017)	(0.021)
Distance × Death	0.154	1.202			0.148	0.820		
	(0.123)	(0.457)**			(0.092)	(0.420)*		
Distance^2 × Death		−0.904				−0.585		
		(0.316)***				(0.297)*		
Close Distance Coauthor × Death			−0.058	−0.067			−0.052	−0.062
			(0.030)*	(0.030)**			(0.024)**	(0.025)**
Distant Distance Coauthor × Death				−0.044				−0.036
				(0.033)				(0.027)
R^2	0.07	0.07	0.07	0.07	0.05	0.05	0.05	0.05
No. of Observations	156918	156918	156918	156918	155156	155156	155156	155156

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from coauthor-deceased author pairwise fixed effect specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (4) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 50th total citation percentile for papers published that year. The dependent variable of models in column (5) to (8) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 75th total citation percentile for papers published that year. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. In column (3) and (7), coauthors whose cognitive distance from the deceased scientist ranks above the 50th percentile rank of the coauthors sample (i.e. both the middle distance and distant distance groups) are the reference group. In column (4) and (8), the reference group contains coauthors in the middle level of cognitive distance from the deceased scientist. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Table A.3: Heterogeneous Effect and Mechanisms—Coauthor's Publication Rates

	No. of Publications					
Dependent Variable:	Unweighted			JIF-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)
Death	0.038	0.030	0.042	0.057	0.046	0.053
	(0.030)	(0.027)	(0.027)	(0.061)	(0.054)	(0.056)
Close × Death	−0.070	−0.082	−0.097	−0.187	−0.197	−0.197
	(0.042)*	(0.041)**	(0.045)**	(0.078)**	(0.072)***	(0.079)**
Distant × Death	−0.151	−0.129	−0.121	−0.075	−0.043	−0.038
	(0.038)***	(0.035)***	(0.038)***	(0.067)	(0.063)	(0.065)
Death × Senior Coauthor	0.009			−0.022		
	(0.085)			(0.106)		
Close × Death × Senior Coauthor	−0.139			−0.075		
	(0.134)			(0.139)		
Distant × Death × Senior Coauthor	0.034			−0.002		
	(0.173)			(0.246)		
Death × First position		0.125			0.119	
		(0.097)			(0.117)	
Close × Death × First position		−0.121			0.085	
		(0.122)			(0.143)	
Distant × Death × First position		−0.132			−0.215	
		(0.157)			(0.181)	
Death × Last position			−0.054			−0.039
			(0.047)			(0.082)
Close × Death × Last position			0.078			0.014
			(0.089)			(0.103)
Distant × Death × Last position			0.009			0.008
			(0.067)			(0.093)
Log pseudo-likelihood	-335436.3	-335528.8	-335540	-865565.4	-866797	-866420.2
No. of Observations	157336	157336	157336	157336	157336	157336

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (3) is the raw number of publications authored by a coauthor (of a deceased scientist) in the year of observation. The dependent variable of models in column (4) to (6) is the number of JIF-weighted publications authored by a coauthor (of a deceased scientist) in the year of observation. "Senior Coauthor" is defined as a coauthor whose career age is larger than the deceased scientist when the latter passes away. "First position" means that in each paper of the coauthor that published together with the deceased scientist, the deceased scientist appears in first position on the authorship roster. "Last Position" means that in each paper of the coauthor that published together with the deceased scientist, the deceased scientist appears in last position on the authorship roster. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. Coauthors in the middle level of cognitive distance from the deceased scientist are the reference group. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

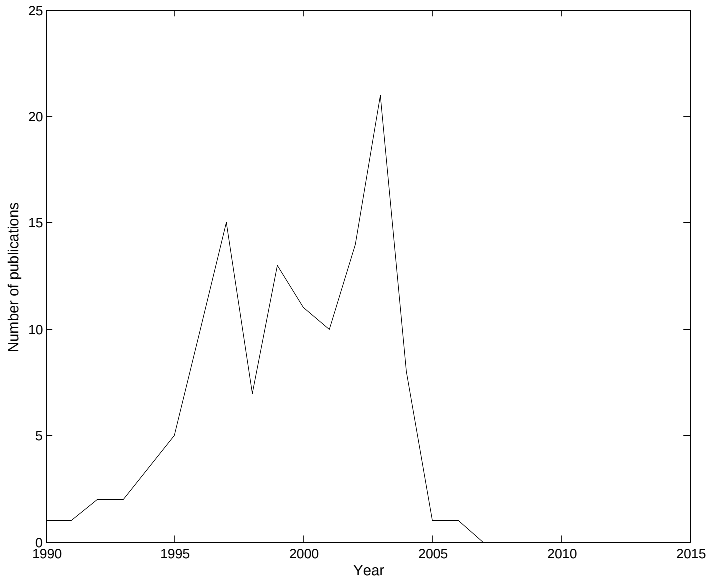
Table A.4: Heterogeneous Effect and Mechanisms—Coauthor's Citation Distribution

	Distribution of Citations					
Dependent Variable:	No. of Pubs abv 50th Percentile of Citations			No. of Pubs abv 75th Percentile of Citations		
	(1)	(2)	(3)	(4)	(5)	(6)
Death	0.043	0.032	0.044	0.046	0.030	0.042
	(0.035)	(0.031)	(0.031)	(0.039)	(0.034)	(0.034)
Close × Death	−0.121	−0.124	−0.137	−0.142	−0.140	−0.150
	(0.044)***	(0.042)***	(0.043)***	(0.050)***	(0.046)***	(0.046)***
Distant × Death	−0.120	−0.093	−0.095	−0.084	−0.057	−0.071
	(0.055)**	(0.047)**	(0.046)**	(0.059)	(0.052)	(0.051)
Death × Senior Coauthor	−0.000			−0.022		
	(0.095)			(0.091)		
Close × Death × Senior Coauthor	−0.085			−0.065		
	(0.143)			(0.149)		
Distant × Death × Senior Coauthor	0.036			−0.030		
	(0.208)			(0.200)		
Death × First position		0.179			0.128	
		(0.108)*			(0.169)	
Close × Death × First position		−0.111			0.067	
		(0.141)			(0.191)	
Distant × Death × First position		−0.229			−0.394	
		(0.186)			(0.264)	
Death × Last position			−0.054			−0.074
			(0.065)			(0.082)
Close × Death × Last position			0.072			0.084
			(0.081)			(0.094)
Distant × Death × Last position			0.058			0.120
			(0.086)			(0.108)
Log pseudo-likelihood	-298061.8	-298146.8	-298110.2	-245249.3	-245348.7	-245288.4
No. of Observations	156918	156918	156918	155156	155156	155156

Notes: Robust standard errors in parentheses are clustered at the level of the deceased scientist. Estimates are from conditional quasi-maximum likelihood Poisson specifications. Observations are at the coauthor$_i$-deceased scientist$_j$-year$_t$ level. The dependent variable of models in column (1) to (3) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 50th total citation percentile for papers published that year. The dependent variable of models in column (4) to (6) is the number of papers authored by a coauthor (of a deceased scientist) in a given year that rank above the 75th total citation percentile for papers published that year. "Senior Coauthor" is defined as a coauthor whose career age is larger than the deceased scientist when the latter passes away. "First position" means that in each paper of the coauthor that published together with the deceased scientist, the deceased scientist appears in first position on the authorship roster. "Last Position" means that in each paper of the coauthor that published together with the deceased scientist, the deceased scientist appears in last position on the authorship roster. All columns control for publication year dummies, indicator variables for career age (year of publication minus year of the first publication, career age zero is the omitted category), and the interaction terms between career age indicators and each covariate of interest. Coauthors in the middle level of cognitive distance from the deceased scientist are the reference group. *significant at 10 percent, **significant at 5 percent, ***significant at 1 percent.

Appendix B

Figure B.1: The Trend of Annual Number of Publications of a Deceased Scientist



Note: Figure B.1 displays the trend of annual number of publications of a deceased scientist in our sample.

Table B.1: Multidisciplinary Scope in Life Sciences

BIODIVERSITY CONSERVATION
PSYCHOLOGY, BIOLOGICAL
BIOCHEMICAL RESEARCH METHODS
BIOCHEMISTRY & MOLECULAR BIOLOGY
BIOLOGY
BIOPHYSICS
BIOTECHNOLOGY & APPLIED MICROBIOLOGY
CELL BIOLOGY
CHEMISTRY, MEDICINAL
EMERGENCY MEDICINE
DENTISTRY/ORAL SURGERY & MEDICINE
EVOLUTIONARY BIOLOGY
DEVELOPMENTAL BIOLOGY
ENGINEERING, BIOMEDICAL
MEDICINE, LEGAL
MARINE & FRESHWATER BIOLOGY
MEDICAL INFORMATICS
MEDICAL LABORATORY TECHNOLOGY
MEDICINE, GENERAL & INTERNAL
MEDICINE, RESEARCH & EXPERIMENTAL
MATERIALS SCIENCE, BIOMATERIALS
MICROBIOLOGY
RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING
REPRODUCTIVE BIOLOGY
SOCIAL SCIENCES, BIOMEDICAL
TROPICAL MEDICINE
CRITICAL CARE MEDICINE
MATHEMATICAL & COMPUTATIONAL BIOLOGY
INTEGRATIVE & COMPLEMENTARY MEDICINE
MEDICAL ETHICS

Notes: Table B.1 lists the 30 disciplines included in the paper as the broadly defined life sciences.

Appendix C The productivity impacts following a labor supply change of scientist i

Given a three-level nesting CES production function and assume the stock of ideas is proportional to the stock of labor, for simplicity, $I = L$, $I_n = L_n$, and $I_i = L_i$, we have

$$
\begin{aligned}
Y &= I^{\phi} K^{1-\alpha} L^{\alpha} \\
Y &= I^{\phi} K^{1-\alpha} [I_1^{\eta} L_1^{\rho} + \cdots + I_n^{\eta} (I_i^{\theta} L_i^{\lambda} + \cdots + I_j^{\theta} L_j^{\lambda})^{\frac{\rho}{\lambda}} + \cdots + I_N^{\eta} L_N^{\rho}]^{\frac{\alpha}{\rho}} \\
Y &= K^{1-\alpha} [I_1^{\eta} L_1^{\rho} + \cdots + I_n^{\eta} (I_i^{\theta} L_i^{\lambda} + \cdots + I_j^{\theta} L_j^{\lambda})^{\frac{\rho}{\lambda}} + \cdots + I_N^{\eta} L_N^{\rho}]^{\frac{\alpha+\phi}{\rho}}.
\end{aligned}
$$

The marginal product of life scientist i is

$$
\begin{aligned}
MP_i &= \frac{\partial Y}{\partial L_i} \\
MP_i &= K^{1-\alpha}(\alpha+\phi)[I_1^{\eta} L_1^{\rho} + \cdots + I_n^{\eta} (I_i^{\theta} L_i^{\lambda} + \cdots + I_j^{\theta} L_j^{\lambda})^{\frac{\rho}{\lambda}} + \cdots + I_N^{\eta} L_N^{\rho}]^{\frac{\alpha+\phi-\rho}{\rho}} I_n^{\eta} (I_i^{\theta} L_i^{\lambda} + \cdots + I_j^{\theta} L_j^{\lambda})^{\frac{\rho-\lambda}{\lambda}} I_i^{\theta} L_i^{\lambda-1} \\
MP_i &= (\alpha+\phi) K^{1-\alpha} L^{\alpha+\phi-\rho} L_n^{\eta+\rho-\lambda} L_i^{\theta+\lambda-1}.
\end{aligned}
$$

Then we have

$$
\log MP_i = (1-\alpha)\log K + (\alpha+\phi-\rho)\log L + (\eta+\rho-\lambda)\log L_n + (\theta+\lambda-1)\log L_i,
$$

and the change in the maginal product of life scientist i is

$$
\begin{aligned}
d\log MP_i &= (1-\alpha) d\log K + (\alpha+\phi-\rho) d\log L + (\eta+\rho-\lambda) d\log L_n + (\theta+\lambda-1)\log L_i \\
d\log MP_i &= s_k d\log K + (\phi - s_k + \frac{1}{\sigma_n}) l + (\eta - \frac{1}{\sigma_n} + \frac{1}{\sigma_c}) l_n + (\theta - \frac{1}{\sigma_c}) l_i,
\end{aligned}
$$

where $s_k = 1-\alpha$ is the capital's share of output, $\sigma_n = \frac{1}{1-\rho}$ is the elasticity of substitution between life scientists in different coauthorship networks, and $\sigma_c = \frac{1}{1-\lambda}$ is the elasticity of substitution between coauthors within each coauthorship network.